

Identifikation von Fahrzeugnutzungsprofilen und deren Einflussfaktoren

Zur Erlangung des akademischen Grades eines
DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)
von der KIT Fakultät für
Bauingenieur-, Geo- und Umweltwissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Ulrich Niklas, M. Sc. – Master of Science
aus Nördlingen

Tag der mündlichen Prüfung: 22.07.2021

Referent: Prof. Dr.-Ing. Peter Vortisch
Korreferent: Prof. Dr.-Ing. Tobias Kuhnimhof

Karlsruhe (2021)

Kurzfassung

Das Fahrzeug ist eines der wichtigsten Verkehrsmittel zur Realisierung von Mobilität. Die Nutzung eines Fahrzeugs setzt sich aus der individuellen Pkw-Mobilität mehrerer Personen zusammen und kann durch verschiedene Faktoren, wie den Kontext (z. B. Urbanität), das Fahrzeug selbst und die Eigenschaften der Nutzer, beeinflusst werden. Diese Arbeit liefert eine grundlegende Metrik zur Strukturierung der Fahrzeugnutzung in drei Dimensionen (Zeit, Distanz und Fahrtzweck) unter Berücksichtigung der genannten Einflussfaktoren. Durch die Identifikation von Fahrzeugnutzungsprofilen ist es nicht nur möglich, die durchschnittliche Nutzung, sondern ähnliche Nutzergruppen differenziert zu betrachten. Diese Informationen liefern Rückschlüsse über die Bedürfnisse der Nutzungsprofile sowie deren Anforderungen an ein Fahrzeug. Deshalb wird im ersten Ansatz die Nutzung von Premium- und Elektrofahrzeugen mit der von Fahrzeugen, welche repräsentativ für eine Bevölkerung stehen, verglichen, indem Premiumfahrzeuge in bestehende Fahrzeugnutzungsprofile auf der Basis von Zeit und Distanz eingeordnet werden. Da bei passiv generierten Sensordaten, welche hauptsächlich in dieser Arbeit verwendet werden, der Fahrtzweck nicht direkt ablesbar ist, wird im nächsten Abschnitt eine Methodik zur Approximation des Fahrtzwecks entwickelt. Darauf aufbauend werden Fahrzeugnutzungsprofile über alle drei Dimensionen gebildet und interpretiert. Damit der Einfluss des Kontextes quantifiziert werden kann, wird ein internationaler und einheitlicher Urbanitätsindex in Abhängigkeit der Erreichbarkeit (z. B. ÖPNV-Qualität), der Dichte (z. B. Bevölkerungsdichte) und der Variabilität (z. B. Anzahl Restaurants) entwickelt. Abschließend wird die Fahrzeugnutzung in Verbindung mit der Urbanität und den Fahrzeugeigenschaften betrachtet.

Abstract

The vehicle is one of the most important means of transport for realisation of mobility. Car usage consists of the individual usage of several people in one vehicle and can be influenced by various factors such as the context (e.g. urbanity), the vehicle itself and the attitudes of its users. This thesis provides a basic metric for structuring car usage into three dimensions (time, distance and trip purpose) taking into account factors influencing car usage. The identification of car usage profiles makes it possible to take a differentiated view on similar user groups and not only to focus on the average user. This information provides insights into the needs of usage profiles and their requirements for a vehicle. Therefore, the first approach compares the usage of premium and electric vehicles with vehicles, which are representative of a population, by assigning them into existing car usage profiles based on the dimensions time and distance. Since for passively generated sensor data, which are mainly used in this thesis, the trip purpose cannot be derived directly, a methodology for approximating the trip purpose is developed. Based on this, car usage profiles are created and interpreted across all three dimensions. In order to quantify the influence of the context on car usage, an international and uniform urbanity index is developed depending on accessibility (e.g. quality of public transport), density (e.g. population density) and variability (e.g. number of restaurants). Finally, car usage is considered in relation to urbanity and car characteristics.

Vorwort

Diese Doktorarbeit wäre ohne den Einfluss einiger Menschen nicht möglich gewesen, denen ich an dieser Stelle meinen besonderen Dank aussprechen möchte.

Als Erstes möchte ich Herrn Prof. Dr. Peter Vortisch für die Bereitschaft zur Betreuung dieser Industriepromotion danken. Auch für die persönlichen und virtuellen Rücksprachen und Anmerkungen zu meiner Arbeit möchte ich mein Lob aussprechen. Darüber hinaus gilt auch Herrn Prof. Dr. Tobias Kuhnimhof meine Wertschätzung, welcher bereitwillig die Zweitkorrektur übernommen hat und mir auch telefonisch stets beiseitestand.

Diese Arbeit wurde im Rahmen des ProMotion-Programms der BMW Group verfasst. Das Programm bietet den optimalen Rahmen für eine Industriepromotion und zeigt, dass Wissenschaft und Industrie gemeinsam wachsen können. Meine ganz besondere Anerkennung geht an meinen Betreuer Dr. Christian Gorges. Seine ehrlichen und konstruktiven Anmerkungen haben sich stets positiv auf die Qualität meiner Arbeit ausgewirkt. Zudem freue ich mich auch, dass über die Zusammenarbeit und Betreuung hinaus eine Freundschaft entstanden ist. Natürlich möchte ich mich auch ganz herzlich bei Dr. Johanna Kopp erkenntlich zeigen, welche mir die Tür zur Promotion geöffnet und den nötigen Freiraum zur Entfaltung meiner Arbeit geschaffen hat. Auch Dr. Peter Phleps möchte ich an dieser Stelle nennen, dessen herzliche und motivierende Art sich jeder in einem Team nur wünschen kann.

Mit Blick nach Karlsruhe möchte ich vor allem Sascha von Behren für sein unaufhaltsames Engagement meinen Dank aussprechen. Ohne Dich wären bestimmt einige wissenschaftliche Veröffentlichungen nicht entstanden. Unvergessen bleiben die vielen Abende, an denen sich alles um Mobilität, Fahrzeugnutzung und statistische Verfahren gedreht hat. Natürlich möchte ich auch allen anderen Mitarbeitern des IfV meine Wertschätzung zum Ausdruck bringen, deren Diskussionen und Gespräche meine Arbeit und mich persönlich bereichert haben.

Auf privater Seite möchte ich mich bei meinen Freunden bedanken, welche bereitwillig ein zweiwöchiges Wegetagebuch ausgefüllt haben. Zudem möchte ich auch die Unterstützung meiner Grundschullehrerinnen Fritzi und Nadja

anerkennen. Speziell meiner Freundin Nadja möchte ich für die letzten drei Jahre danken. Du warst mir immer eine Stütze und hast mir stets geholfen, den Kopf frei und einen frischen Blick auf die Arbeit zu bekommen. Zu guter Letzt möchte ich an dieser Stelle noch meine Eltern erwähnen. Sie haben nicht nur während des Verfassens meiner Doktorarbeit hinter mir gestanden, sondern haben mich auf meinem gesamten Lebensweg immer unterstützt und meine Entscheidungen nie infrage gestellt. Das werde ich Euch nie vergessen.

Inhaltsverzeichnis

Kurzfassung	II
Abstract	III
Vorwort	IV
Inhaltsverzeichnis	VI
Abbildungsverzeichnis	IX
Tabellenverzeichnis	XI
Abkürzungsverzeichnis	XII
1 Einleitung	1
1.1 Motivation	1
1.2 Ziel der Arbeit	2
2 Fahrzeugnutzung im Mobilitätskontext	5
2.1 Individuelle Pkw-Mobilität	5
2.2 Fahrzeugnutzung	7
2.3 Segmentierungsansätze	9
2.3.1 Fahrzeugsegmentierung	10
2.3.2 Kundensegmentierung	10
2.3.3 Fahrzeugnutzungssegmentierung	11
2.3.4 Formale Beschreibung der Segmentierungsansätze	12
3 Zeit-Distanz-Fahrzeugnutzungsprofile	13
3.1 Premiumfaktor	13
3.1.1 Motivation	13
3.1.2 Literaturübersicht	15
3.1.3 Daten	18
3.1.4 Methodik	22
3.1.5 Ergebnisse	26
3.1.6 Diskussion	37

3.2	Elektrofaktor	40
3.2.1	Motivation	40
3.2.2	Literaturübersicht	42
3.2.3	Methodik	44
3.2.4	Ergebnisse	47
3.2.5	Diskussion	51
3.3	Zwischenfazit	53
4	Zeit-Distanz-Fahrtzweck-Fahrzeugnutzungsprofile	54
4.1	Identifikation des Fahrtzwecks	54
4.1.1	Motivation	54
4.1.2	Literaturübersicht	55
4.1.3	Daten	58
4.1.4	Methodik	60
4.1.5	Limitationen	74
4.2	Bestimmung von Zeit-Distanz-Fahrtzweck-Fahrzeugnutzungsprofilen	76
4.2.1	Literaturübersicht	76
4.2.2	Daten	77
4.2.3	Methodik	82
4.2.4	Interpretation der Cluster	96
4.3	Zwischenfazit	99
5	Einflussfaktoren auf die Fahrzeugnutzung	100
5.1	Überblick über die Einflussfaktoren auf die Fahrzeugnutzung	100
5.2	Quantifizierung von Urbanität	101
5.2.1	Motivation	101
5.2.2	Literaturübersicht	106
5.2.3	Daten	109
5.2.4	Methodik	113
5.2.5	Ergebnisse	118
5.2.6	Diskussion	126
5.3	Analyse verschiedener Einflussfaktoren auf die Fahrzeugnutzung	127
6	Fazit und Ausblick	136
	Literaturverzeichnis	143
	Anhang	154
A	Hauptkomponentenanalyse	154

B	k-Means-Verfahren	155
C	Metriken zur Bewertung von Clusterlösungen	156
D	Satz von Bayes	157

Abbildungsverzeichnis

1.1	Zusammenhänge zwischen Marktakteuren und staatlichen Maßnahmen	2
2.1	Die drei Dimensionen der Fahrzeugnutzung und deren Einflussfaktoren	7
2.2	Analyse von Fahrzeugnutzung zur Fahrzeugentwicklung	9
3.1	Vorgehen für den Vergleich von Premiumfahrzeugen und gewöhnliche Fahrzeugen	15
3.2	Konfusionsmatrix	26
3.3	VKT zwischen Premium- und gewöhnlichen Fahrzeugen (monatlich)	28
3.4	Premiumfahrzeuge: Verteilung für zurückgelegte Fahrdistanzen	31
3.5	Premiumfahrzeuge: Verteilung der zurückgelegten Fahrten	32
3.6	Premiumfahrzeuge: Anzahl Passagiere nach Fahrzeugnutzungsprofil	34
3.7	Anteil der Zulassungen nach Antrieben in Norwegen	41
3.8	Wahrscheinlichkeit der Zuteilung zu den Fahrzeugnutzungsprofile	48
4.1	Schematische Darstellung der Fahrzeugnutzung in drei Dimensionen	55
4.2	Darstellung der Fahrtverflechtungen	60
4.3	Wichtige Standorte und approximierter Wohnort eines Fahrzeugs	64
4.4	Fahrtzwecke nach Fahrdistanz	68
4.5	Fahrtzwecke nach Standzeit	69
4.6	Fahrtzwecke nach Startzeit	70
4.7	Präzision entlang der Wahrscheinlichkeit der Zuordnung	73
4.8	Schematische Darstellung des Aktionsradius eines Fahrzeugs	79
4.9	Korrelationen der Inputvariablen	82
4.10	Schematische Netzarchitektur des verwendeten Autoencoders	85
4.11	Paarplot mit den fünf z -Variablen	88
4.12	t-SNE-KDE-Plot	89
4.13	Berechnete Gesamtwahrscheinlichkeit π_k der fünf Gauß-Verteilungen	94
4.14	t-SNE-KDE-Plot in Abhängigkeit von den fünf Clustern	94
4.15	Paarplot in Abhängigkeit von den fünf Clustern	95
5.1	Schematischer Zusammenhang zwischen den Einflussfaktoren und die Fahrzeugnutzung	101
5.2	Prozess der Entwicklung des Urbanitätsindex	105
5.3	Variablen mit dem größten Einfluss auf die Klassifizierung	116

5.4	Partielle Abhängigkeiten der Variablen entlang der Raumtypen	117
5.5	Vergleich der Raumtypisierung des BBSR mit der Raumtypisierung anhand des UI am Beispiel von Berlin	119
5.6	Vergleich von Deutschland (Berlin) und Frankreich (Paris)	119
5.7	Bevölkerung nach Raumtypen und Ländern	124
5.8	Darstellung von Seoul	124
5.9	Darstellung USA und New York	125
5.10	Darstellung Barcelona, Madrid und Bilbao	125

Tabellenverzeichnis

1.1	Eigenschaften nationaler Mobilitätserhebungen	3
2.1	Fahrtzwecke nach Mobilitätserhebungen	6
3.1	Sensordaten: Datenstruktur	20
3.2	Modellauswahl	24
3.3	Relativer Durchschnitt der fünf clusterbildenden Variablen	27
3.4	Premiumfahrzeuge: Eigenschaften der verschiedenen Fahrzeugnutzungsprofile	29
3.5	Vergleich von Premiumfahrzeugen und gewöhnliche Fahrzeugen	30
3.6	Premiumfahrzeuge: Fahrzeugnutzungsprofile über die vier Jahreszeiten	35
3.7	Premiumfahrzeuge: Fahrzeugnutzungsprofile über verschiedene Länder	37
3.8	Entscheidungsfunktionen der Fahrzeugnutzungsprofile	46
3.9	Vergleich von Elektrofahrzeugen und gewöhnliche Fahrzeugen	49
4.1	Datenstruktur: GPS-basierte Sensordaten	58
4.2	Hyperparameter zur Approximation des Fahrtzwecks	67
4.3	Validierung des Modells zur Fahrtzweckschätzung	72
4.4	Zusätzlich generierte Fahrtinformationen	78
4.5	Fahrzeugtabelle	80
4.6	Merkmalsausprägungen der Originalvariablen der DPGMM-Clusterlösung	96
5.1	Literaturübersicht über Definitionen von Urbanität	108
5.2	Variablen zur Definition von Urbanität	112
5.3	Erklärende Variablen nach Raumtyp und Land	121
5.4	Verteilung verschiedener Faktoren nach Raumtyp und Land	123
5.5	Quelle-Ziel-Raumtypmatrizen: Anteil der Fahrten nach Raumtyp des approximierten Wohnortes	129
5.6	Raumtyp des approximierten Wohnortes nach Fahrzeugnutzungsprofilen	130
5.7	Quelle-Ziel-Raumtypmatrizen: Anteil der Fahrten (mittlere Fahrdistanz) nach Fahrzeugnutzungsprofilen	132
5.8	Verteilung der Fahrzeuge nach Raumtyp des approximierten Wohnortes und Segment/Konzept	134

Abkürzungsverzeichnis

AIC	Akaike-Informationskriteriums
BBSR	Bundesinstitut für Bau-, Stadt- und Raumforschung
BEV	batterieelektrischen Fahrzeug
BIC	bayesianischen Informationskriteriums
CAN	Controller Area Network
CATI	Computer-Assisted-Telephone-Interviewing
CHTS	California Household Travel Survey
CLC	Corine-Land-Cover
CRP	Chinese Restaurant Process
CUMILE	Car Usage Model Integrating Long Distance Events
CVRP	Clean Vehicle Rebate Program
DBSCAN	Density-based Spatial Clustering of Applications with Noise
DPGMM	Dirichlet Process Gaussian Mixture Model
ECOC	Error-Correcting-Output-Code
EM	Erwartungs-Maximierungs
FAS	Fahrerassistenzsystem
FLE	Fahrleistungserhebung
FNP	Fahrzeugnutzungsprofil
GfK	Gesellschaft für Konsumforschung
GIS	Geoinformationssystem
GKL	Große Klasse
GMM	Gaussian Mixture Models
GMT	Greenwich Mean Time
GPS	Global-Positioning-System

HOV	High-Occupancy-Vehicle
HTS	Mobilitätserhebungen
IG	Informationsgewinn
INVERMO	Die intermodale Vernetzung von Personenverkehrsmitteln unter Berücksichtigung der Nutzerbedürfnisse
KDE	Kerndichteschätzer
KKL	Kleine Klasse
kNN	k-Nearest Neighbors
MiD	Mobilität in Deutschland
MIV	motorisierten Individualverkehr
MKL	Mittlere Klasse
MOP	Deutsche Mobilitätspanel
MPV	Multi Purpose Vehicle
NHTS	National Household Travel Survey
NUTS	Nomenclature des unités territoriales statistiques
ODB	On-Board-Diagnose
ODM	On-Demand-Mobility
OECD	Organisation für wirtschaftliche Zusammenarbeit und Entwicklung
OSM	Open-Street-Map
PCA	Hauptkomponentenanalyse
Pkw	Personenkraftwagen
PLZ	Postleitzahl
POI	Point-of-Interest
SAC	Sports Active Coupé
SAG	stochastisches durchschnittliches Gradientenabstiegsverfahren
SAT	Sports Active Tourer
SAV	Sports Activity Vehicle
SCMD	Swedish car movement data

SOP	Start der Produktion
SR	Softmax-Regression
SrV	Mobilität in Städten
SUV	Sports Utility Vehicle
SVM	Support Vector Machine
t-SNE	t-Distributed-Stochastic-Neighbor-Embedding
UI	Urbanitätsindex
UK	Vereinigten Königreich
UKL	Untere Klasse
USA	Vereinigten Staaten von Amerika
VAN	Anonymisierte Fahrgestellnummer
VKT	zurückgelegte Distanzen
WLTP	weltweit harmonisiertes Testverfahren für leichtgewichtige Nutzfahrzeuge
ÖPNV	öffentlicher Personennahverkehr

1 Einleitung

1.1 Motivation

Tully und Baier (2007) zeigen in ihrer sozialwissenschaftlichen Abhandlung zur nachhaltigen Gestaltung von Mobilität auf, dass moderne Gesellschaften sich durch einen hohen Grad an räumlicher Mobilität auszeichnen. Wohn- und Arbeitsplatz, Einkaufen sowie sonstige Freizeitaktivitäten sind oftmals räumlich voneinander getrennt. Darüber hinaus sind die Mobilität und die damit verbundene Vernetzung der Gesellschaft Treiber vieler wohlfahrtsfördernder Entwicklungen. Adam Smith zeigt bereits in seinem Buch zur klassischen Nationalökonomie „Der Wohlstand der Nationen“ aus dem Jahr 1776, dass internationaler Handel und die Spezialisierung der jeweiligen Länder auf bestimmte Produktionsgüter die Wohlfahrt aller beteiligten Länder verbessern können (Smith, 2010). Hierfür sind der Transport von Gütern und die Mobilität von Personen zwingend erforderlich. Die räumlichen Distanzen zwischen Standorten können mithilfe von unterschiedlichen Verkehrsmitteln zurückgelegt werden. Der Personenkraftwagen (Pkw) wird unter allen verfügbaren Verkehrsmitteln am häufigsten zur Fortbewegung genutzt. Knapp 81 % der Verkehrsleistung in Deutschland im Jahr 2018 ist auf den motorisierten Individualverkehr (MIV) zurückzuführen (BMVI, 2020). Dieser hohe Nutzungsgrad führt dazu, dass der Pkw als Fortbewegungsmittel stetig weiterentwickelt wird. Die technologischen Verbesserungen eines Automobils haben sich in der Vergangenheit primär auf die Optimierung von Leistung und Sicherheit fokussiert. Derzeit (2020) spielen aber die möglichst klimaneutrale Produktion und Nutzung von Fahrzeugen¹ eine immer größere Rolle. Diese umweltschonenden Prämissen werden noch zusätzlich vom Staat durch zielgerichtete Maßnahmen auf das Angebot und die Nachfrage (Abbildung 1.1) unterstützt. Angebotsorientierte Instrumente werden primär durch nationale CO₂-Flottenemissionsziele gesteuert. Hierbei wird ein fixer bzw. variabler CO₂-Zielwert vom Staat vorgegeben. Der Flottenemissionswert eines Automobilherstellers wird über ein gewichtetes, harmonisches Mittel berechnet. Dabei wird jedem zugelassenen Fahrzeug ein CO₂-Abgaswert zugewiesen, welcher über bestimmte Testverfahren (z. B. weltweit harmonisiertes Testverfahren für leichtgewichtige Nutzfahrzeuge (WLTP)) berechnet wird. Wird der CO₂-Zielwert von einem Automobilhersteller nicht erreicht, werden Strafzahlungen, abhängig von der Höhe der Abweichung, eingefordert. Infolge von angebotsorientierten Maßnahmen hat der Automobilhersteller einen Anreiz, neue Antriebstechnologien bzw. neue Geschäftsmodelle in den Markt zu bringen, um die

¹Der Begriff Fahrzeug wird in dieser Arbeit als Synonym zum Pkw verwendet.

Flottenziele zu erreichen. Über nachfrageorientierte Maßnahmen versucht der Staat, Anreize für den Käufer zu schaffen, schadstoffarme Fahrzeuge zu erwerben bzw. die Nutzung von Fahrzeugen mit Verbrennungsmotor zu reduzieren. Auf der einen Seite wird durch lokale Regulierungen (z. B. Dieserverbote in Stadtzentren in Rom, High-Occupancy-Vehicle (HOV)-Lanes für Elektrofahrzeuge in Kalifornien) und eine Benzinsteuer die Nutzung von Fahrzeugen mit Verbrennungsmotor erschwert und verteuert, während auf der anderen Seite durch den Kauf von emissionsarmen Fahrzeugen (z. B. Elektrofahrzeugen) die Besteuerung von Emissionswerten und sogenannten Feebates gefördert wird (Adamou, Clerides & Zachariadis, 2014; Brand, Anable & Tran, 2013). Die angebots- und nachfrageorientierten Maßnahmen können die Kaufentscheidungen beeinflussen oder führen zu Veränderungen und Limitationen in der Fahrzeugnutzung. Beispielsweise erreichen Elektrofahrzeuge nicht die Reichweiten wie Fahrzeuge mit Verbrennungsmotor. Zudem dauert das Laden von Elektrofahrzeugen länger und ihre Verfügbarkeit ist vor allem in ländlichen Gebieten limitiert. Neue Geschäftsmodelle (z. B. On-Demand-Mobility (ODM)) sind nicht flächendeckend verfügbar, was die Nutzung auf urbane Gebiete beschränkt. Des Weiteren wirken sich eine antriebspezifische Beschränkung der Fahrzeugnutzung und die Besteuerung fossiler Kraftstoffe nicht nur auf die Nutzung, sondern auch auf die Preisstabilität der Fahrzeuge im Gebrauchtwagenmarkt aus.

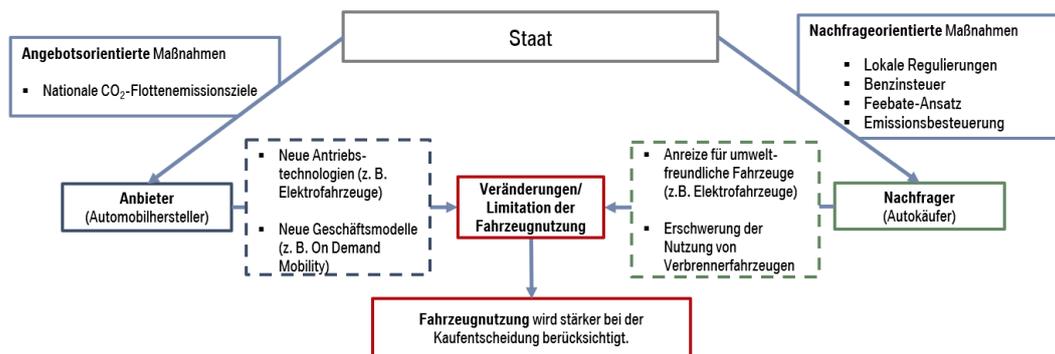


Abbildung 1.1: Zusammenhänge zwischen Marktakteuren und staatlichen Maßnahmen

Die Wirkungskette von angebots- und nachfrageorientierten staatlichen Maßnahmen und die damit einhergehenden Konsequenzen führen zu der Hypothese, dass die Fahrzeugnutzung im Kaufentscheidungsprozess in der Zukunft eine immer stärker werdende Rolle spielen wird. Deshalb wird in dieser Arbeit die Fahrzeugnutzung umfassend diskutiert.

1.2 Ziel der Arbeit

Die Untersuchung der Fahrzeugnutzung ist in der Wissenschaft keine Neuheit. Nationale Erhebungen, wie beispielsweise das Deutsche Mobilitätspanel (MOP), die Mobilität in Deutschland (MiD) oder der U.S. National Household Travel Survey (NHTS) (Tabelle 1.1),

befragen bereits seit vielen Jahren Menschen zu ihrer (Pkw-)Mobilität. Umfragen sind eine verbreitete Methode, um Wissen über die Mobilität der Menschen und deren Hintergründe und Soziodemografie in einem repräsentativen Querschnitt zu generieren. Viele nationale Erhebungen (MiD, U.S. NHTS) fragen das Verhalten innerhalb eines Tages mittels Fragebogen (telefonisch, schriftlich, Onlinebefragung) ab. Durch die Stichtagserhebung hält sich der Aufwand für die Teilnehmer in Grenzen, was die Bereitschaft zur Teilnahme steigert. Bei der MiD bzw. dem U.S. NHTS nahmen bei der letzten Umfrage im Jahr 2017 316 000 Personen (156 000 Haushalte) bzw. 321 000 Personen (118 000 Haushalte) an der Befragung teil. Diese Stichtagserhebungen finden in einem regelmäßigen Zyklus von fünf bzw. acht Jahren statt. Die U.S. NHTS ist eine der ältesten Mobilitätsbefragungen, welche 1969 erstmalig erhoben wurde. Stichtagserhebungen fragen zwar sehr viele Menschen bzw. Haushalte ab, sind aber durch die Querschnittsbetrachtung beschränkt. Dadurch können beispielsweise Regelmäßigkeiten im Verhalten nur schwer identifiziert werden. Das MOP bildet die (Pkw-)Mobilität im Längsschnitt über einen Zeitraum von einer Woche ab. Hierbei wird das alltägliche Verhalten innerhalb einer Woche besser erfasst. Darüber hinaus werden die Befragten in insgesamt drei aufeinanderfolgenden Jahren beobachtet. Dies führt dazu, dass auch Veränderungen über die Zeit bestimmt werden können. Hierfür werden die Teilnehmer mittels Wegetagebuch im Frühjahr und über ein Tankbuch im Herbst zu ihrer (Pkw-)Mobilität befragt. Aufgrund des höheren Aufwandes und der damit verbundenen Kosten betrug die Nettostichprobengröße 2019 3074 Personen bzw. 1850 Haushalte. Beim MOP können durch den längeren Beobachtungszeitraum zwar Regelmäßigkeiten innerhalb einer Woche identifiziert werden, allerdings ist die Identifikation von unregelmäßigen Fahrten dennoch schwierig.

Tabelle 1.1: *Eigenschaften nationaler Mobilitätsbefragungen*

	Erhebungen		
	MOP	MiD	U.S. NHTS
Stichprobe	3074 Personen 1850 Haushalte	316 000 Personen 156 000 Haushalte	321 419 Personen 118 208 Haushalte
Beobachtungszeitraum	1 Woche	1 Tag	1 Tag
Erhebungsmethode	Längsschnitterhebung	Stichtagserhebung	Stichtagserhebung
Frequenz	jährlich	ca. alle 5 Jahre	ca. alle 8 Jahre
Erstmalige Erhebung	1994	2002	1969
Letzte Erhebung	2019	2017	2017

Zudem bilden Umfragen immer das Verhalten von Individuen ab. Da beim Kauf eines Fahrzeugs nicht nur die individuelle Pkw-Mobilität des Käufers, sondern die Pkw-Mobilität aller Nutzer berücksichtigt wird, muss das Fahrzeug, sprich die Fahrzeugnutzung, ganzheitlich über einen längeren Zeitraum beobachtet werden.

Sensordaten erfüllen genau dieses Kriterium. Sie werden automatisch vom Fahrzeug generiert und bilden die Fahrzeugnutzung, unabhängig vom Nutzer, ab. Fahrzeuge müssen lediglich die technischen Voraussetzungen (z. B. Software, Ausstattung) erfüllen und der Fahrzeugnutzer muss der Datenaufzeichnung zustimmen. Hierbei ist die Berücksichtigung datenschutzrechtlicher Regelungen (Datenschutz-Grundverordnung (DSGVO)) von zentraler Bedeutung. In der DSGVO werden die Nutzung und Verarbeitung personenbezogener Daten geregelt und vereinheitlicht. Ist das Skript einmal im System implementiert, kann dieses auf beliebig viele Fahrzeuge ausgerollt werden. Dies lässt die Stichprobe und den Beobachtungszeitraum beliebig groß werden. Durch die rasante Entwicklung von Big-Data-Technologien (Parallelisierung, Partitionierung von Daten, Lambda-Architekturen) ist es nun auch möglich die Skalierbarkeit bei exponentiell ansteigenden Datenmengen zu gewährleisten. Unter Skalierbarkeit wird nach Marz und Warren (2015) „[...] die Fähigkeit, die Geschwindigkeit bei wachsenden Datenmengen oder steigender Arbeitslast [...] aufrechtzuerhalten“, verstanden. Diese Big-Data-Technologien ermöglichen es erst, große Mengen an Fahrzeugnutzungsdaten zu verarbeiten und zu analysieren.

Aus den genannten Gründen werden in dieser Arbeit Sensordaten zur Abbildung der Fahrzeugnutzung im Mobilitätskontext sowie Umfragedaten aus nationalen Mobilitätsbefragungen verwendet. Folgende Fragen werden im Zuge dieser Arbeit beantwortet:

- In welchen Fachabteilungen eines Automobilherstellers und auf welche Art und Weise gibt es einen Bedarf zur Untersuchung von Fahrzeugnutzung?
- Welche Dimensionen und Einflussfaktoren der Fahrzeugnutzung sind für den zugrunde liegenden Anwendungsbereich relevant?
- Wie und mit welchen Daten lassen sich Nutzungsmuster identifizieren?
- Welche Eigenschaften hat ein Fahrzeugnutzungsprofil (FNP)?
- Wie unterscheiden sich FNPs nach differenzierenden Faktoren?

2 Fahrzeugnutzung im Mobilitätskontext

2.1 Individuelle Pkw-Mobilität

Traditionell wird die Fahrzeugnutzung aus Sicht von Individuen betrachtet, was als Pkw-Mobilität definiert wird. Dabei kann die Pkw-Mobilität und die damit verbundenen Verkehrsströme für die Entwicklung oder zum Ausbau von Verkehrsinfrastrukturen genutzt werden. Häufig können auch Veränderungen über die Zeit (verursacht durch die in Unterkapitel 1.1 beschriebenen Limitationen) zur Planung von Städten einen Mehrwert liefern. Viel diskutiert wird in diesem Zusammenhang, wie urbane Strukturen im Allgemeinen die Nutzung von Verkehrsmitteln beeinflussen können (Dieleman, Dijst & Burghouwt, 2002; Pan, Shen & Zhang, 2009). Gegenwärtig (2020) stehen der Ausbau von Ladestationen und die Herausforderungen mit dem immer größer werdenden Anteil an Elektrofahrzeugen am Markt im Mittelpunkt politischer Diskussionen (Hardman et al., 2018; Pagany, Ramirez Camargo & Dorner, 2019). Des Weiteren existieren auch Studien, welche den Zusammenhang zwischen der Abhängigkeit vom Pkw und der Qualität des öffentlichen Personennahverkehrs (ÖPNV) in einem Gebiet untersuchen (Buehler, Pucher, Gerike & Götschi, 2017; von Behren et al., 2018; Wiersma, Bertolini & Straatemeier, 2017). Aus diesen Beispielen geht hervor, dass der **Kontext**, in dem sich ein Fahrzeug bewegt, einen Einfluss auf die Abhängigkeit vom Pkw und damit auch auf die Nutzung hat. Hardman et al. (2018) zeigen in einer Literaturübersicht auf, dass sich die Kundenpräferenzen hinsichtlich der Ladeinfrastruktur unterscheiden lassen. Der **Kunde** selbst, seine Einstellung zu Verkehrsmitteln (Ramos, Bergstad & Nässén, 2020; von Behren, Bönisch, Niklas & Chlond, 2020) sowie die individuellen Lebensumstände (Cheng et al., 2019; Ehteshamrad, Saffarzadeh & Mamdoohi, 2017; Mackett, 2002) können ebenfalls als Einflussfaktoren auf die Pkw-Mobilität wahrgenommen werden. Aber auch das **Fahrzeug** an sich (Antrieb, Fahrzeuggröße, Fahrzeugalter) kann die Fahrzeugnutzung beeinflussen. Beispielsweise zeigen Niklas, von Behren, Eisenmann, Chlond und Vortisch (2020) und Niklas, von Behren, Chlond und Vortisch (2020) auf, dass sich die Nutzung von Premium- bzw. Elektrofahrzeugen von der Nutzung von gewöhnlichen Fahrzeugen unterscheidet. Wird in der Literatur von Pkw-Mobilität gesprochen, werden hauptsächlich drei Dimensionen mit unterschiedlicher Intensität betrachtet. Die **Distanz** spielt bei der Eignung von Elektrofahrzeugen eine wichtige Rolle. In diesem Zusammenhang wird sie im Wesentlichen auf Tagesebene

betrachtet (Jakobsson, Gnann, Plötz, Sprei & Karlsson, 2016; Khan & Kockelman, 2012). Der **Fahrtzweck** ist ebenfalls ein wichtiger Bestandteil bei der Betrachtung von Pkw-Mobilität (Ramos et al., 2020). Das MOP, die MiD sowie der U.S. NHTS unterscheiden Fahrten in neun Fahrtzwecke (Administration, 2017; Ecke et al., 2019; Nobis & Kuhnimhof, 2018). Die in Tabelle 2.1 aufgeführten Begriffe stammen aus den Fragebögen der jeweiligen Erhebungen. Durch einen Vergleich der abgefragten Fahrtzwecke werden die unterschiedlichen Definitionen und Abgrenzungen zwischen den Erhebungen deutlich. Zur Vergleichbarkeit und im Wissen, dass die Fahrtzwecke aus den einzelnen Erhebungen nicht perfekt übereinstimmen, wurden die unterschiedlichen Benennungen zu einem Begriff zusammengefasst.

Beispielsweise weisen das MOP und die MiD „Ausbildung“ als separaten Zweck aus, während dieser beim U.S. NHTS zusammengefasst unter mehreren Auswahlmöglichkeiten („School/Daycare/Religious“) wiederzufinden ist. Gleichzeitig werden im MOP und dem U.S. NHTS die Zwecke „Einkauf“ und „Besorgung“ zusammen abgefragt, die MiD definiert diese wiederum nur als „Einkauf“. Die U.S. NHTS nähert sich dem Zweck „Freizeit“ über „Social/Recreational“. Frisör- bzw. Arztbesuche werden im MOP und der MiD etwas breiter als „private Erledigung“ zusammengefasst, während die U.S. NHTS diesen Zweck direkt als „Medical/Dental services“ ausweist. Der Fahrtzweck „Meal“ wird bei U.S. NHTS als alleinstehender Fahrtzweck abgefragt, beim MOP und der MiD fällt er unter „Freizeit“.

Tabelle 2.1: Fahrtzwecke nach Mobilitätserhebungen

Fahrtzweck	Erhebungen		
	MOP	MiD	U.S. NHTS
Nach Hause	Nach Hause	nach Hause	Home
Arbeit	Zur Arbeit	Zur Arbeit	Work
Dienstlich	Dienstlich/geschäftlich	Beruflich	–
Ausbildung	Zur Ausbildung	Zur Ausbildung	School/Daycare/Religious
Einkauf	Einkauf/Besorgung	Einkauf	Shopping/Errands
Freizeit	Freizeit	Freizeitaktivität	Social/Recreational
Holen/Bringen	Holen/Bringen	Holen/Bringen	Transport someone
Erledigungen	Private Erledigung	Private Erledigung	Medical/Dental services
Essen	–	–	Meal
Sonstiges	Anderes, und zwar:	Anderer Zweck	Something else

Eine weitere Dimension ist die **Zeit**. Ähnlich wie bei der Distanz kann diese aus unterschiedlichen Perspektiven betrachtet werden. Eisenmann und Buehler (2018) unterscheiden beispielsweise Fahrten, die am Wochenende oder werktags stattfinden. Verreault und Morency (2011) betrachten jeden Wochentag für sich und Sodenkamp, Wenig, Thiesse und Staake (2019) untersuchen die Pkw-Mobilität unter anderem mit Hilfe von Tagesganglinien.

Abbildung 2.1 fasst die Erkenntnisse und Zusammenhänge aus der Literatur zusammen. Die Fahrzeugnutzung (im Mobilitätskontext) wird durch die Dimensionen Distanz, Fahrtzweck und Zeit beschrieben. Die drei Dimensionen können wiederum durch den Kontext, den Kunden und das Fahrzeug beeinflusst werden.



Abbildung 2.1: Die drei Dimensionen der Fahrzeugnutzung und deren Einflussfaktoren

2.2 Fahrzeugnutzung

Da das Fahrzeug das zentrale Element eines Automobilherstellers ist, wird in vielen Bereichen des Unternehmens die Fahrzeugnutzung untersucht. Im Gegensatz zur individuellen Pkw-Mobilität wird diese aus Sicht eines Fahrzeugs betrachtet. Da sich die Pkw-Mobilität mehrerer Individuen in einem Fahrzeug bündeln kann, wird im Folgenden auch von Fahrzeugnutzung (im Mobilitätskontext) und nicht von der Pkw-Mobilität gesprochen. Um zu identifizieren, zu welchem Zweck Nutzungsdaten eines Fahrzeugs im Industriekontext beobachtet werden, wurden insgesamt 22 Interviews geführt. Aus datenschutzrechtlichen Gründen existieren keine Aufzeichnungen oder Transkripte der Interviews. Zur Dokumentation wurden ausführliche Notizen während der Gespräche gemacht, deren Kernpunkte im Nachfolgenden erläutert werden. Die Interviews hatten eine Dauer von durchschnittlich 30 Minuten. Der Gesprächsleitfaden der unstrukturierten Interviews setzt sich aus drei offenen Stufen zusammen. Die erste Stufe beschäftigt sich mit dem Tätigkeitsbereich des Befragten. Im nächsten Schritt werden die Ver-

wendung von Fahrzeugnutzungsdaten und deren Charakteristika im Fachbereich thematisiert. Abschließend werden mögliche Verbesserungen zur Verwendung von Fahrzeugnutzungsdaten diskutiert. Lamnek (2005) sowie Katenkamp und Schröder (2003) zeigen in ihren Grundlagenarbeiten zur qualitativen Sozialforschung, dass durch ein offenes Konzept viele Informationen gesammelt werden, welche die Perspektive des Befragten (und des Fachbereichs) widerspiegeln. Alle Teilnehmer stammen aus dem Umfeld eines Automobilherstellers. Die Arbeitsinhalte der Befragten erstrecken sich von strategischen Themen (Regulatorik, Mobilität im Allgemeinen, Segmentierung) über sicherheitsrelevante Themen (Absicherung von Bauteilen/Funktionen, Gewährleistungsfälle, Unfallanalyse, Qualitätssicherung) bis hin zum Aftersales (Vertrieb, Funktionsnutzung).

Die Entwicklung eines Fahrzeugs kann zeitlich in drei Phasen (Vorentwicklungs-, Entwicklungs- und Beobachtungsphase im Feld) unterteilt werden (Abbildung 2.2). Die Vorentwicklungs- und Entwicklungsphase befinden sich vor dem sogenannten Start der Produktion (SOP). Der SOP beschreibt den Beginn der Serienfertigung eines Fahrzeugs. Thematisch können die generelle Fahrzeugnutzung in Fahrzeugnutzung im Mobilitätskontext, Beobachtung von Funktionen und Fahrdynamik voneinander abgegrenzt werden.

Die Analyse der **Funktionsnutzung** und **Fahrdynamik** verläuft über die drei Phasen ähnlich und ist vor allem für sicherheitsrelevante Themen von Bedeutung, da gemäß internationalen Standards (ISO 26262) die Nutzung ganzheitlich zur Qualitätssicherung beobachtet werden muss. Bei der Funktionsnutzung und Fahrdynamik ist die Verflechtung zwischen den einzelnen Phasen deutlich erkennbar und prozessual verankert. Die **Fahrzeugnutzung** (im Mobilitätskontext) wird primär in der Vorentwicklungsphase zur Untersuchung strategischer Themen auf der Basis offizieller Mobilitätsbefragungen (z. B. MiD, MOP) betrachtet. Hierbei stehen die Identifikation von Trends (neue Mobilitätskonzepte und Antriebstechnologien) und die Veränderung von Rahmenbedingungen (Regulatorik, Ladeinfrastruktur) im Vordergrund. Die Untersuchung interner Fahrzeugnutzung (im Mobilitätskontext) in der Beobachtungsphase im Feld zur Identifikation von Potenzialen und Risiken in der Vorentwicklungsphase findet bisher nur sehr selektiv und ohne prozessuale Verankerung statt.

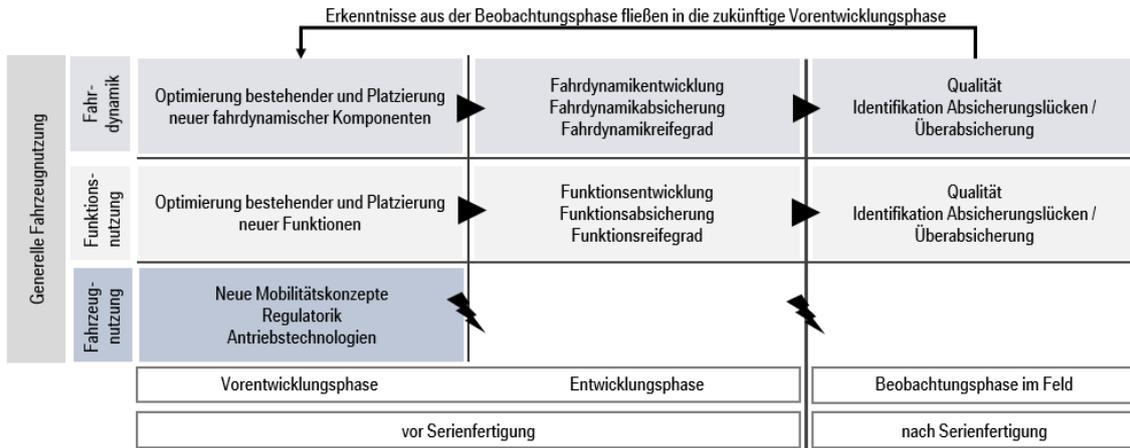


Abbildung 2.2: Analyse von Fahrzeugnutzung zur Fahrzeugentwicklung

Zusammenfassend kann die generelle Fahrzeugnutzung in die beschriebenen drei Bereiche Fahrzeugnutzung (im Mobilitätskontext), Funktionsnutzung und Fahrdynamik aufgeteilt werden. Die in Unterkapitel 2.1 erörterten Dimensionen der Pkw-Mobilität beschreiben Informationen zu Distanz, Zeit und dem Fahrtzweck und werden zur Beschreibung der Fahrzeugnutzung (im Mobilitätskontext) in dieser Arbeit verwendet. Bei fahrdynamischen Bauteilen und Funktionen sind die drei Phasen miteinander verbunden, da die Erkenntnisse aus den einzelnen Phasen in die jeweils nächste einfließen. Die Fahrzeugnutzung (im Mobilitätskontext) wird bisher nicht durchgängig in allen Bereichen beobachtet. Inwieweit sich beispielsweise lokale Regulierungen (z. B. Einfahrverbote) mit der gegenwärtigen Fahrzeugnutzung decken, wird bisher nur selektiv untersucht. Aus diesem Grund werden in dieser Arbeit Fahrzeugnutzungsdaten verwendet, um diese im Mobilitätskontext zu analysieren. Ziel dabei ist es, die Fahrzeugnutzung im Feld zu beobachten und mit strategischen Themen abzugleichen und darauf die Vorentwicklung von Fahrzeugen auszurichten.

2.3 Segmentierungsansätze

Ein etablierter Ansatz sowohl für die langfristige Unternehmensplanung in der Vorentwicklungsphase als auch für die Wissenschaft ist es, einen heterogenen Markt (z. B. Automobilmarkt oder Mobilitätsmarkt) in homogene Unterbereiche (z. B. Fahrzeugsegmentierung oder Mobilitätstypen) einzuteilen. Durch die Einteilung eines Marktes in Subgruppen ist es zum einen möglich, auf der Angebotsseite das Wettbewerbsumfeld besser zu beobachten und mit den eigenen Produkten zu vergleichen (Fahrzeugsegmentierung), und zum anderen kann auf der Nachfrageseite ein Verständnis der Bedürfnisse und Wünsche der Kunden entwickelt werden (Kundensegmentierung) (Esch & Manger, 2019). Die Segmentierung von Fahrzeugen und Kunden dient dabei als strukturgebendes Element, wodurch die gewonnenen Erkenntnisse gezielt in die Entwicklungsphase einfließen können. In diesem Unterkapitel werden zwei

etablierte Segmentierungsansätze vorgestellt, um daraus die thematische und methodische Herangehensweise zur weiteren Untersuchung von Fahrzeugnutzung (im Mobilitätskontext) zu bestimmen.

2.3.1 Fahrzeugsegmentierung

Ziel einer Fahrzeugsegmentierung ist es, die Produkte der Wettbewerber mit den eigenen vergleichen zu können. Im Automobilmarkt werden Fahrzeuge auf der Basis von marktbezogenen Kriterien in sogenannte „Segmente“ und „Konzepte“ unterteilt (BMW, 2020b).

Aus Sicht eines Automobilherstellers gibt es drei Klassen (Untere Klasse (UKL), Mittlere Klasse (MKL), Große Klasse (GKL)), welche sich in acht **Fahrzeugsegmente** (UKL0, UKL1, UKL2, Kleine Klasse (KKL)1, KKL2, MKL, GKL, GKL+) unterteilen. Segmente definieren sich in erster Linie über die Fahrzeuglänge. Darüber hinaus hat auch die Preispositionierung bei der Einteilung der Fahrzeuge einen Einfluss. Beispielsweise kann ein Fahrzeug in einem anderen Segment geführt werden als andere Modelle mit gleicher Länge, wenn es aufgrund der Preispositionierung vom Kunden eher als Alternative in einem anderen Segment wahrgenommen wird. GKL+ definiert sich ausschließlich über den Preis (ab ca. €160 000 (brutto)). Generell gilt: je größer das Segment, desto größer die Fahrzeuglänge.

Die Zuordnung eines Fahrzeugs zu einem **Konzept** wird über die Karosserieform bestimmt. Dabei wird zwischen Motorraum, Fahrgastraum und Kofferraum unterschieden. Je nach Verhältnis der Räume zueinander, fallen sie in eines der zwölf Konzepte (Sedan „Classic“, Sedan „Coupé“, Estate, Hatch, Coupé, Convertible, Roadster, Multi Purpose Vehicle (MPV), Sports Active Tourer (SAT), Sports Active Coupé (SAC), Sports Activity Vehicle (SAV), Sports Utility Vehicle (SUV)). Bei einem Sedan sind der Motor-, Fahrgast- und Kofferraum jeweils getrennt voneinander, hier wird von einem sogenannten Drei-Box-Design gesprochen. Bei einem Estate wird die Trennung des Fahrgast- und Kofferraums aufgehoben, wodurch das Ladevolumen steigt und der Transport von sperrigen Gegenständen möglich ist. Das MPV besteht aus einem Ein-Box-Design, das sich durch viel Laderaum, Variabilität und Komfort auszeichnet. SUVs sind Geländewagenkonzepte mit hoher Bodenfreiheit und Offroadorientierung. Ein SUV verfügt generell über einen Leiterraum, Differenzialsperren und einen Unterbodenschutz. Das SAV-Konzept kann als sportlich-urbane Weiterentwicklung des SUV definiert werden und zeichnet sich durch limousinenähnliche Fahreigenschaften aus (BMW, 2020b).

2.3.2 Kundensegmentierung

Ziel der Kundensegmentierung ist es, Menschen mit ähnlicher Soziodemografie und Motivation sowie gleichartigen Einstellungen und Bedürfnissen mit differenzierten Produkten anzusprechen. Dadurch können die Wünsche der Kunden zielgenauer befriedigt werden (Eurosensors,

2018). Eine etablierte Herangehensweise mit großer Stichprobe ist der Ansatz der sozialen Milieus. Probanden werden mittels standardisierter Fragebögen in ein Milieu eingeteilt. Jedes Milieu repräsentiert dabei unterschiedliche Einstellungen, Motivationen und Bedürfnisse. Die Teilnehmer werden gewichtet, um einen repräsentativen Schnitt der Bevölkerung darzustellen. Soziale Milieus betrachten zwei Dimensionen: den sozialen Status und die Werteorientierung. Der soziale Status beschreibt die sozioökonomischen Charakteristika (Einkommen und Vermögen, Bildung und Beruf). Je höher der soziale Status, desto höher sind die Kaufkraft, das Einkommen und die Berufsgruppe und damit auch die Wahrscheinlichkeit Premiumprodukte zu kaufen. Die Werteorientierung beschreibt die Einstellung zu bestimmten Werten (Familie, Lebensziele, Lebensstil, Einstellung zur Arbeit und Freizeit, Gesundheit etc.). Entlang der Werteorientierung wird zwischen drei Anschauungen unterschieden. Die traditionelle Anschauung berücksichtigt die Präferenz für gewöhnliche und traditionelle Produkte. In der modernen Weltanschauung wollen Zielgruppen ihre soziokulturelle Identität durch die Wahl ihrer Produkte und Marken zum Ausdruck bringen. In der postmodernen Weltanschauung handeln Menschen nach ihrem individuellen Geschmack, unabhängig von gegenwärtigen Trendbewegungen. Je höher die Werteorientierung, desto moderner ist die Grundorientierung des jeweiligen Milieus im soziokulturellen Sinn (Esch & Manger, 2019; Eurosenors, 2018).

Auf der Basis der Dimensionen können die Probanden in elf soziale Milieus (etabliertes Milieu, traditionelles bürgerliches Milieu, traditionelles Arbeitermilieu, konsum-materialistisches Milieu, modernes bürgerliches Milieu, aufstiegsorientiertes Milieu, liberal-intellektuelles Milieu, sozial-kritisches Milieu, modernes Arbeitnehmersmilieu, hedonistisches Milieu und postmodernes Milieu) eingeteilt werden. Mit Hilfe der Eigenschaften der jeweiligen sozialen Milieus können die Marken der Automobilhersteller in einem zweidimensionalen Koordinatensystem verortet werden. Beispielsweise spricht BMW primär Individuen des aufstiegsorientierten Milieus an, während MINI sich eher auf die Zielgruppe des postmodernen Milieus fokussiert. Diese Logik lässt sich nicht nur auf Markenebene, sondern auch auf Fahrzeugebene adaptieren. Die Erkenntnisse aus den sozialen Milieus fließen dann in die Entwicklungsphase ein.

2.3.3 Fahrzeugnutzungssegmentierung

Durch eine herstellerorientierte Fahrzeugsegmentierung können markenunabhängig Fahrzeuge in Relation zueinander gestellt werden. Über eine soziodemografisch und einstellungsorientierte Kundensegmentierung werden Individuen auf der Basis ihres sozialen Status und ihrer Werteorientierung in Milieus eingeteilt. Die Grundannahme ist, dass Individuen im gleichen Milieu ähnliche Bedürfnisse und Wünsche in Bezug auf ein Fahrzeug oder eine Marke haben. Durch die Abbildung eines repräsentativen Querschnitts kann zudem abgeleitet werden,

wie hoch der Bedarf bestimmter Milieus an Fahrzeugen oder Marken ist. Zudem können Verschiebungen der sozialen Milieus über die Zeit beobachtet und die Ausrichtung eines Fahrzeugs oder einer Marke dementsprechend angepasst werden.

Beide Segmentierungsansätze berücksichtigen allerdings nicht die Nutzung der Fahrzeuge an sich. Informationen zur Fahrzeugnutzung fließen somit nur bedingt in die strategische Ausrichtung eines Fahrzeugs ein. Mit Hilfe eines nutzungsorientierten Segmentierungsansatzes würde die tatsächliche Nutzung der Fahrzeuge (in der Beobachtungsphase im Feld) stärker in den Mittelpunkt des Fahrzeugentwicklungsprozesses (in der Vorentwicklungsphase) rücken (Abbildung 2.2). Dies hätte den Vorteil, dass Veränderungen oder Limitationen der Fahrzeugnutzung (Abbildung 1.1) in der Vorentwicklungsphase beobachtet und damit in der Entwicklungsphase berücksichtigt werden könnten.

2.3.4 Formale Beschreibung der Segmentierungsansätze

Formal haben Segmentierungsansätze immer das Ziel, Objekte i (z. B. Fahrzeuge) mit bestimmten Eigenschaften x_i (z. B. Tagesdistanz) auf Ähnlichkeiten zu prüfen und diese in Segmente k zusammenzufassen. Methodisch entspricht dies einer Clusteranalyse bzw. einer Klassifikation. Bei einer **Clusteranalyse** werden mit Hilfe von statistischen Verfahren Ähnlichkeiten durch bestimmte Metriken (z. B. euklidische Distanz) identifiziert und Objekte in Segmente (statistisch: Cluster) eingeteilt. Da diese mittels statistischer Verfahren identifiziert werden, sind die Besonderheiten nicht bekannt. Die Anzahl, die Eigenschaften und die Einteilung der Objekte in Segmente hängen bei diesem „A-posteriori-Ansatz“ stark von der zugrundeliegenden Datenverteilung ab. Nachdem die Zugehörigkeit von Objekt i in Segment k bestimmt worden ist, müssen die einzelnen Segmente inhaltlich interpretiert werden. Ein weiterer Ansatz des Clusterings ist die Angabe von vordefinierten Kriterien (Beispiel: Fahrzeug mit einem Kaufpreis größer z entspricht dem Segment k). Dieser lineare Ansatz ist unabhängig von der Verteilung der Daten und bildet „A-priori-Segmente“ ab. Durch die vordefinierten Segmente sind diese bereits inhaltlich beschrieben, wodurch keine nachträgliche Interpretation notwendig ist. Bei der **Klassifikation** wird der Zusammenhang von Objekten i mit Eigenschaften x_i und einer Zielvariable y_i (z. B. eine existierende Fahrzeugnutzungssegmentierung) erlernt. Hierbei kann die Präzision der Schätzung durch die vorhandene Zielvariable genau bestimmt werden (James, Witten, Hastie & Tibshirani, 2013). Auch hier ist durch die bereits vorhandene Zielvariable eine Interpretation der Segmente nicht notwendig.

Um einen nutzungsorientierten Segmentierungsansatz zu entwickeln, werden in dieser Arbeit verschiedene Ansätze zur Identifikation von Gruppen in der Fahrzeugnutzung analysiert und sogenannte Fahrzeugnutzungsprofile (FNPs) auf der Basis der in Unterkapitel 2.1 beschriebenen Dimensionen abgeleitet.

3 Zeit-Distanz-Fahrzeugnutzungsprofile

Wie bereits in Unterkapitel 2.1 erläutert, wird die Pkw-Mobilität mit den unterschiedlichen Dimensionen in der Forschung intensiv untersucht. In diesem Zusammenhang gibt es einige Arbeiten, welche sich mit der Identifikation von FNPs/Pkw-Mobilitätstypen auf der Basis repräsentativer Daten beschäftigen, wie beispielsweise Eisenmann und Buehler (2018) und Haustein und Nielsen (2016). Deshalb ist es in einem ersten Schritt wichtig zu prüfen, inwieweit sich die Verwendung von unternehmensinternen Fahrzeugnutzungsdaten mit der von repräsentativen Mobilitätsdaten unterscheidet. Stimmt die Nutzung nicht überein, müssen FNPs auf der Basis interner Sensordaten entwickelt werden, um unternehmensinterne Entscheidungen treffen zu können. Unterkapitel 3.1 vergleicht bezüglich Zeit und Distanz die Fahrzeugnutzung von Premiumfahrzeugen mit der von gewöhnlichen Fahrzeugen². In Unterkapitel 3.2 wird der Ansatz, basierend auf den Limitationen, methodisch weiterentwickelt, um die Nutzung von Elektrofahrzeugen und gewöhnlichen Fahrzeugen gegenüberzustellen.

3.1 Premiumfaktor

*Dieses Unterkapitel basiert großteils auf der Veröffentlichung „Premium factor – Analysing usage of premium cars compared to conventional cars“
(Niklas, von Behren, Eisenmann et al., 2020).*

3.1.1 Motivation

Der Pkw ist ein flexibles Mobilitätswerkzeug im Vergleich zu anderen Verkehrsmitteln (z. B. öffentlicher Personennahverkehr (ÖPNV)). Das Fahrzeug kann für verschiedenste Zwecke, wie z. B. für Pendeln, Holen und Bringen von Personen, Transport von Gegenständen und Freizeitaktivitäten, genutzt werden. Um die Pkw-Mobilitätsbedürfnisse zu befriedigen, kann zwischen einer Reihe von Fahrzeugen gewählt werden, welche sich in Preis, Größe und Komfort unterscheiden. In diesem Zusammenhang ergibt sich folgende Frage: Unterscheidet sich die Nutzung anhand des Fahrzeugtyps (z. B. Premiumfahrzeug vs. kein Premiumfahrzeug)? Premiumfahrzeuge definieren sich hauptsächlich über den Preis. Aufgrund des höheren Preises, bieten Premiumfahrzeuge einen höheren Komfort und eine größere Ausstattung. Typische

²Die Nutzung von gewöhnlichen Fahrzeugen spiegelt die repräsentative Fahrzeugnutzung einer Bevölkerung (z. B. Deutschland oder Kalifornien) wider.

deutsche Hersteller von Premiumfahrzeugen sind Audi, BMW oder Mercedes. Gewöhnliche Fahrzeuge berücksichtigen alle Arten von Fahrzeugen, unabhängig von der Marke, für eine repräsentative Bevölkerung. Deshalb eignen sich repräsentative Mobilitätshebungen (HTS) gut, um die gewöhnliche Fahrzeugnutzung abzubilden. Insbesondere Premiumfahrzeuge erfordern ein großes finanzielles Investment im Vergleich zu gewöhnlichen Fahrzeugen. Bedeutet das aber gleichzeitig auch, dass sich die Nutzung zwischen den Fahrzeugtypen unterscheidet? Wie bereits eingangs erwähnt, soll dadurch beantwortet werden, ob sich die Ausgestaltung von Premiumfahrzeugen aufgrund der Fahrzeugnutzung unterscheiden muss. Beispielsweise liefern die Resultate Erkenntnisse darüber, ob sich die Elektroreichweite für eine Premiumflotte differenzieren muss. Um dies zu beantworten, werden Premiumfahrzeuge mit gewöhnlichen Fahrzeugen verglichen. Zur Beantwortung solcher Fragestellungen haben sich im Bereich der Mobilitätsforschung zwei datengetriebene Richtungen etabliert: Mobilitätsdaten von Personen (eine HTS) und Sensordaten von Fahrzeugen.

Bei HTS wird das individuelle Mobilitätsverhalten im Hinblick auf die Pkw-Mobilität auf Personenebene beobachtet. Aus diesen Informationen werden die Nutzungshäufigkeit und die zurückgelegten Distanzen eines Fahrzeugs abgeleitet. Das Hauptproblem bei diesen Studien ist, dass die Fahrzeugperspektive vernachlässigt wird. Fahrzeuge in einem Haushalt werden oft von verschiedenen Individuen genutzt, vor allem in einem Mehrpersonenhaushalt mit einem Fahrzeug. Deshalb ist die Fahrzeugnutzung eine Kombination aus der Nutzung von mehreren Personen. Bei bestimmten Fragestellungen ist genau diese Betrachtungsweise von entscheidender Relevanz. Forschungsarbeiten zu neuen Antriebstechnologien (z. B. Elektromobilität) und die Ersetzung von herkömmlichen Fahrzeugen mit Verbrennungsmotor mit Elektrofahrzeugen setzen die Betrachtung der Fahrzeugperspektive zu fundierten Ableitungen voraus. Das hat den Vorteil, dass die Pkw-Mobilität aller Nutzer eines Fahrzeugs berücksichtigt wird. Durch die Fahrzeugperspektive ist es nun möglich zu bestimmen, ob alle Nutzer eines Fahrzeugs für beispielsweise Elektromobilität geeignet sind oder nicht. Der Ansatz, die Pkw-Mobilität aus einer Fahrzeugperspektive zu betrachten, wurde von Eisenmann (2019) realisiert. Dabei modellierte sie, basierend auf HTS-Daten, die Fahrzeugnutzung für gewöhnliche Fahrzeuge aus einer Fahrzeugperspektive für Deutschland und Kalifornien über einen Beobachtungszeitraum von einem Jahr. Darauf aufbauend teilten Eisenmann und Buehler (2018) die Fahrzeuge und deren Nutzung in acht eindeutige FNPs ein. In den Daten sind zwar Premiumfahrzeuge enthalten, allerdings ist die Stichprobe von Premiumfahrzeugen so klein, dass eine Auswertung nicht valide ist.

Ein anderer Ansatz, die Nutzung von Premiumfahrzeugen zu betrachten, basiert auf Sensordaten. Allerdings sind der Zugang zu solchen Daten und deren Vergleichbarkeit schwierig. Darüber hinaus ist die Stichprobe von Sensordaten in der Literatur meistens klein, wodurch keine robusten Aussagen ableitbar sind.

In diesem Kapitel werden beide Ansätze kombiniert mit dem Ziel, Premiumfahrzeuge (aus Sensordaten) mit gewöhnlichen Fahrzeugen (von HTS) in Deutschland und Kalifornien zu vergleichen. Aufgrund der unterschiedlichen Datenerfassung wird ein Gerüst benötigt, um beide Datenquellen vergleichbar zu machen. Die acht FNP's von Eisenmann und Buehler (2018) erfüllen diese Anforderungen. Sie sind abstrakt genug, um einen Vergleich zwischen beiden Datenquellen zu erlauben. Basierend auf diesem Gerüst ist das Ziel, die relativen Größen der FNP's zwischen Premiumfahrzeugen und gewöhnlichen Fahrzeugen gegenüberzustellen, um folgende Forschungsfrage zu beantworten:

- Werden Premiumfahrzeuge anders genutzt als gewöhnliche Fahrzeuge?

Um diese Frage zu beantworten, wird wie folgt vorgegangen (Abbildung 3.1): Die Literaturanalyse betrachtet die unterschiedlichen Datenquellen zur Untersuchung von Fahrzeugnutzung sowie deren Vor- und Nachteile. Danach findet die Beschreibung der verwendeten Datenquellen statt. Das implementierte Gerüst (acht FNP's) und das verwendete maschinelle Lernverfahren zur Einteilung der Premiumfahrzeuge in vorhandene FNP's wird als Nächstes vorgestellt. Infolgedessen werden die zugewiesenen Premiumfahrzeuge mit den bereits eingeteilten gewöhnlichen Fahrzeugen in Deutschland und Kalifornien verglichen. Der letzte Teil schließt das Unterkapitel mit einer Diskussion der Limitationen und einem Ausblick für weiterführende Forschungsarbeiten ab.

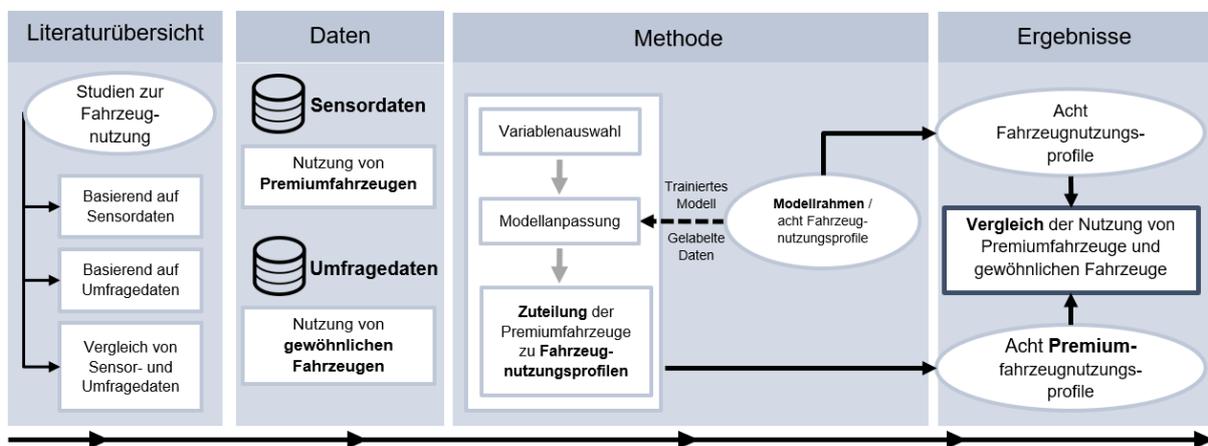


Abbildung 3.1: Vorgehen für den Vergleich von Premiumfahrzeugen und gewöhnliche Fahrzeugen

3.1.2 Literaturübersicht

Es gibt eine Vielzahl von Studien, welche sich mit der Pkw-Mobilität im Alltag beschäftigen (Bäumer, Hautzinger, Kuhnimhof & Pfeiffer, 2018; Weiß, Chlond, von Behren, Hilgert & Vortisch, 2016). Jedoch basieren diese Analysen häufig auf HTS, wodurch die Pkw-Mobilität aus der Perspektive von Individuen und nicht aus der Fahrzeugperspektive betrachtet

wird. Haustein und Siren (2015) nutzen die Eurobarometer-Umfrage mit 27 000 Probanden zur Identifikation von Unterschieden im Mobilitätsverhalten zwischen den verschiedenen europäischen Ländern. Hierfür verwenden sie einen k-Means-Algorithmus, um Individuen auf der Basis ihres Mobilitätsverhaltens in acht interpretierbare Cluster einzuteilen. 61,5% der Haushalte in Westdeutschland werden den „busy green drivers“ und „convenience drivers“ zugeordnet. Ein weiterer Ansatz ist die Nutzung von Umfragedaten im Längsschnitt, welche das Mobilitätsverhalten über mehrere Wochen beinhalten. Beispielsweise betrachten Axhausen, Zimmermann, Schönfelder, Rindsfuser und Haupt (2002) sowie Löchl (2005) in Deutschland und der Schweiz das Mobilitätsverhalten über sechs Wochen in den Jahren 1999 bzw. 2003. Die Studien ergeben, dass das Mobilitätsverhalten für bestimmte Wegezwecke (z. B. Einkaufen) robust ist. Eine starke Stabilität ergibt sich auch, wenn regelmäßig die gleichen Standorte (z. B. Supermarkt) angefahren werden (Schönfelder, 2006; Susilo & Axhausen, 2014). Bäumer et al. (2018) vergleichen die Jahresfahrleistung zwischen verschiedenen deutschen HTS (MOP, MiD und Fahrleistungserhebung (FLE)) ab 2000. Die Studie zeigt, dass die Jahresfahrleistung über die letzten zwei Jahrzehnte zwischen 13 000 km und 15 000 km beträgt und damit sehr stabil über die Zeit ist. Eisenmann (2019) analysiert die Fahrzeugnutzung und unterscheidet die Fahrzeuge je nach Stellung im Haushalt (Erstfahrzeug in einem Haushalt mit mehreren Autos, Zweitfahrzeug in einem Haushalt mit mehreren Autos etc.) auf der Basis des MOP. Fahrzeuge in einem Haushalt mit mehreren Autos haben eine Monatsfahrleistung von ca. 1500 km. Pasaoglu et al. (2012) analysieren die Pkw-Mobilität von verschiedenen Umfragen in sechs europäischen Ländern. Die durchschnittliche Tagesfahrleistung beträgt zwischen 50 km und 60 km, abhängig vom Wochentag. Darüber hinaus ergibt sich eine durchschnittliche Anzahl an Fahrten pro Tag von ungefähr 2,0 in Deutschland.

In den letzten Jahren werden Sensordaten vermehrt in der Mobilitätsforschung eingesetzt. Dabei ist es möglich die Nutzung direkt und ohne Transformationen aus einer Fahrzeugperspektive zu beobachten. Basierend auf Global-Positioning-System (GPS)-Daten, welche 2004 über 52 Wochen in der Metropolregion Atlanta gesammelt wurden, analysieren Elango, Guensler und Ogle (2007) und Pearre, Kempton, Guensler und Elango (2011) die Fahrzeugnutzung mit 500 Fahrzeugen. Pearre et al. (2011) zeigen eine durchschnittliche Tagesfahrleistung von 32,6 Meilen (inklusive Tagen, an denen nicht gefahren wird) auf. Elango et al. (2007) betrachten die Fahrzeugnutzung differenziert nach der Zusammensetzung des Haushalts. Hierbei stellt sich heraus, dass Fahrzeuge in bestimmten Haushaltskonstellationen (Mehrpersonenhaushalte, Haushalte mit mehreren Fahrzeugen und Haushalte mit hohem Einkommen) tendenziell variabler eingesetzt werden. Khan und Kockelman (2012) sammeln Daten von 419 Fahrzeugen aus 255 Haushalten von November 2004 bis April 2006. Basierend auf 269 357 Fahrten an 143 004 Fahrzeugtagen, beträgt die durchschnittliche Tagesfahrleistung 25,4 Meilen. Schuessler und Axhausen (2009) reichern GPS-Daten von 4882 Teilnehmern aus der Schweiz

mit weiteren Informationen an. Anhand von Geschwindigkeit und Beschleunigung konnten das Verkehrsmittel und der Wegezweck bei einer aufgezeichneten Fahrt ermittelt werden. Kullingsjö und Karlsson (2012) liefern eine Übersicht über existierende GPS-basierte Mobilitätsstudien. Dieser Überblick zeigt eindeutig, dass die Stichprobe oftmals sehr klein ist und sich die Analysen auf spezifische Städte oder Regionen beschränken.

Einige Mobilitätsstudien beschäftigen sich mit einem Vergleich von Sensor- und Umfragedaten. Stopher, FitzGerald und Xu (2007) untersuchen die Genauigkeit von HTS aus Sydney durch GPS-Vermessungen von 118 Fahrzeugen in 70 Haushalten. Die 70 Haushalte nahmen sowohl bei der HTS aus Sydney als auch bei der GPS-Vermessung teil. Die parallele Teilnahme ermöglicht es, die Genauigkeit zwischen beiden Erhebungsinstrumenten zu bestimmen. Die durchschnittlich zurückgelegte Distanz liegt unter HTS bei 8,85 km pro Fahrt, während die GPS-Vermessung einen Wert von 8,07 km ergibt. Bei HTS werden somit 0,78 km höhere Distanzen angegeben als tatsächlich gemessen wurden. Darüber hinaus wurden 7,4 % der Fahrten in der HTS nicht berichtet. Das könnte vor allem daran liegen, dass die HTS-Teilnehmer Fahrten mit geringen Standzeiten zwischen den Fahrten zusammenlegten. Auch Wolf, Oliveira und Thompson (2003) untersuchen die Verzerrung durch Über- oder Unterberichterstattung. Dabei vergleichen sie, ähnlich wie Stopher et al. (2007), die zurückgelegten Distanzen aus einer HTS von drei Regionen in Kalifornien mit GPS-Daten der selben Haushalte. Fahrten außerhalb der Hauptverkehrszeiten werden eher nicht berichtet. Jedoch gibt es keine signifikante Richtung, ob Umfragen eine Über- oder Unterberichterstattung widerspiegeln. In San Diego beträgt die durchschnittliche Fahrdistanz, basierend auf Computer-Assisted-Telephone-Interviewing (CATI), 6,9 Meilen. Im Vergleich wird über GPS-Daten 5,8 Meilen gemessen. In Alameda ist die Beziehung genau umgekehrt, während in Sacramento die durchschnittliche Fahrdistanz zwischen CATI- und GPS-Daten gleich ist. Auer, Bogenberger, Rehborn, Koller und Palmer (2017) analysieren die Charakteristika des individuellen Mobilitätsverhaltens aus der Fahrzeugperspektive und vergleichen die Ergebnisse mit verschiedenen HTS. Aufgrund von Datenbeschränkungen werden nur Eckdaten, wie z. B. die durchschnittliche Fahrtzeit, Distanz oder Geschwindigkeit der Fahrzeugflotte, verglichen. Das individuelle Mobilitätsverhalten über einen längeren Zeitraum und spezifische FNP's werden in dieser Arbeit nicht untersucht. Kelly, Krenn, Titze, Stopher und Foster (2013) liefern eine Übersicht über existierende Studien, welche die Ergebnisse von HTS und GPS-Studien vergleichen. Alle betrachteten Studien zeigen auf, dass bei Umfragen tendenziell eine Überschätzung der Fahrtzeiten stattfindet und der Beobachtungszeitraum meistens ein Tag entspricht. Manche Studien arbeiten auch mit Mobilitätsdaten, welche das Mobilitätsverhalten über mehrere Wochen aufzeichnen. Eine Limitation der Literaturübersicht ist, dass lediglich eine Differenzierung der Fahrtweiten zwischen HTS und GPS-Studien stattfindet. Unterschiede bezüglich Fahrdistanzen werden nicht berücksichtigt. Nachteil bei der Verwendung von Sensordaten zur Analyse der Fahrzeugnutzung

ist die generelle Verfügbarkeit. Die meisten Analysen sind sehr spezifisch und reflektieren nicht die gesamte Fahrzeugflotte. Des Weiteren ist die Stichprobe oftmals sehr klein und abhängig von spezifischen Städten oder Fahrzeugtypen. Aufgrund dessen sind die Daten meistens auch nicht repräsentativ für eine Bevölkerung. Die Gründe für diese Limitationen können auch auf die hohen Kosten der Erhebungsmethode zurückgeführt werden. Eine weitere Einschränkung ist, dass keine Informationen zu Fahrtzwecken direkt aus den Sensordaten abgelesen werden kann, was bereits in vielen Studien diskutiert wird (Bricka, Sen, Paleti & Bhat, 2012; Eisenmann & Buehler, 2018; Elango et al., 2007; Kelly et al., 2013; Wolf et al., 2003). Auf der anderen Seite kann es bei HTS zu Ungenauigkeiten in der Berichterstattung kommen. Durch die Eigenangaben der Teilnehmer werden manche Fahrten nicht zurückgemeldet. Dies kann daran liegen, dass die Probanden die Fahrt vergessen oder sie als nicht wichtig erachten (Bricka et al., 2012). In der Realität kann ein Fahrzeug aber auch von Individuen aus unterschiedlichen Haushalten genutzt werden. Darüber hinaus ist der Beobachtungszeitraum oftmals auf wenige Tage oder Wochen beschränkt (Kullingsjö & Karlsson, 2012). Das Modell von Chlond, Weiss, Heilig und Vortisch (2014) und Eisenmann (2019) berücksichtigt diese Limitationen. Dabei identifizieren Eisenmann und Buehler (2018) mit der modellierten Fahrzeugnutzung acht FNPs. Aufgrund der kleinen Stichprobe von Premiumfahrzeugen ist es aber schwierig, die Unterschiede der Nutzung von Premiumfahrzeugen und gewöhnlichen Fahrzeugen allein mit dieser Datenquelle valide abzubilden. Trotzdem sind die modellierten Fahrzeugnutzungsdaten von Chlond et al. (2014) und Eisenmann (2019) eine gute Grundlage, um Sensordaten mit Umfragedaten zu vergleichen, weil die Nutzung aus einer Fahrzeugperspektive betrachtet wird und es möglich ist, die verwendeten Attribute für die Erstellung von FNPs mit Sensordaten zu reproduzieren. Im Folgenden soll das Gerüst von Eisenmann und Buehler (2018) verwendet werden, um die modellierten Fahrzeugnutzungsdaten mit Sensordaten anzureichern. Das Ziel ist es, die Unterschiede und Ähnlichkeiten der Nutzung von Premiumfahrzeugen und gewöhnlichen Fahrzeugen zu untersuchen.

3.1.3 Daten

Zur Beantwortung der Forschungsfrage werden zwei Datenquellen verwendet: Bei „Car Usage Model Integrating Long Distance Events (CUMILE)“ handelt es sich um modellierte Fahrzeugnutzungsdaten auf der Basis von Umfragedaten, welche die Nutzung von einer repräsentativen Gesamtflotte über einen Zeitraum von einem Jahr in Deutschland und Kalifornien betrachten. Die verwendeten Sensordaten eines deutschen Premiumautomobilherstellers bilden die Nutzung von Premiumfahrzeugen für einen Monat in Deutschland und Kalifornien ab. Beide Datensätze berücksichtigen sowohl private als auch gewerbliche Fahrzeuge.

Umfragedaten

Die Umfragedaten beinhalten 1659 gewöhnliche Fahrzeuge mit 1 207 115 Fahrten in Deutschland und 1681 gewöhnliche Fahrzeuge mit 2 084 508 Fahrten in Kalifornien. CUMILE wurde ursprünglich für Deutschland (CUMILE-GER), basierend auf dem MOP und einer Studie zu Fernverkehrsfahrten (Die intermodale Vernetzung von Personenverkehrsmitteln unter Berücksichtigung der Nutzerbedürfnisse (INVERMO)), entwickelt. Das MOP ist eine deutsche HTS, welche aus Informationen zur Alltagsmobilität (MOP-EM), der Fahrleistung und dem Kraftstoffverbrauch (MOP-FCOR) besteht:

- Beim MOP-EM berichten 1000–1500 Haushalte zu ihrer Alltagsmobilität über einen Zeitraum von einer Woche. Hierbei werden soziodemografische Informationen des Haushalts (z. B. Anzahl der Fahrzeuge im Haushalt, Haushaltsnettoeinkommen) und Informationen der Haushaltsmitglieder (z. B. Geschlecht, Alter, Führerscheinbesitz) gesammelt (Ecke et al., 2019; Zumkeller, 2009).
- Beim MOP-FCOR machen Haushalte mit Fahrzeugen aus dem MOP-EM acht Wochen lang Angaben zur Gesamtkilometeranzahl und dem Kraftstoffverbrauch. Fahrzeuginformationen wie Antriebsart und Motorleistung sind ebenfalls im Datensatz enthalten.

Bei INVERMO werden ungefähr 10 000 Teilnehmer in Deutschland zu ihren letzten drei Fernverkehrsereignissen (Distanz, Verkehrsmittelwahl, Tag der Abfahrt, Tag der Ankunft und Wegezweck) befragt (Zumkeller, Chlond, Last & Manz, 2006). Die Erhebung wird von 2000 bis 2002 durchgeführt. Der CUMILE-GER-Algorithmus modelliert mit den beschriebenen Daten die Fahrzeugnutzung über ein Jahr, indem vier Faktoren bestimmt werden:

1. Pkw-Mobilitätsdaten der Umfrageteilnehmer während der MOP-EM-Woche
2. Fahrzeugnutzung für klassische Tage in einem Jahr
3. Fahrzeugnutzung am Tag während des Zeitraums der MOP-FCOR-Umfrage
4. Fahrzeugnutzung am Tag für die übrigen Tage in einem Jahr

Für den Vergleich der Fahrzeugnutzung zwischen Deutschland und Kalifornien modelliert Eisenmann und Buehler (2018) die Fahrzeugnutzung in Kalifornien in CUMILE-CAL. Das Konzept von CUMILE-CAL ist CUMILE-GER sehr ähnlich. Eisenmann und Buehler (2018) verwenden den California Household Travel Survey (CHTS) von 2010 bis 2012. Der CHTS beinhaltet GPS- und On-Board-Diagnose (ODB)-Informationen zum Mobilitätsverhalten über einen Zeitraum von einer Woche sowie eine Umfrage zum Fernverkehr.

Sensordaten

Sensordaten, welche in dieser Studie für Premiumfahrzeuge verfügbar sind, werden automatisch ohne Interaktion mit dem Fahrzeugnutzer versendet. Sobald das Fahrzeug 150 km zurückgelegt hat, kann der Autonutzer die Datensammlung bestätigen. Die Zustimmung beinhaltet keinerlei monetäre Vorteile. Die Darstellung der zugrundeliegenden Sensordaten erfolgt über verschiedene Applikationen, weshalb die Daten in Hintergrundsystemen verarbeitet, analysiert und aufbereitet werden. Die Interaktion zwischen den Sensordaten und dem Hintergrundsystem basiert auf Events. Die Telematik-Kontrolleinheit sammelt alle relevanten Daten und sendet diese mit einem Übertragungsverfahren (z. B. Controller Area Network (CAN) und Diagnosejobs) an das Hintergrundsystem. Sobald 500 Events gesammelt wurden oder das System seit fünf Minuten aktiv ist, wird automatisch ein Datenpaket an das Hintergrundsystem geschickt. Die Daten werden nach den relevanten Fahrtinformationen (z. B. Zeitstempel, Kilometerstand) gefiltert und in die dargestellte Datenstruktur (Tabelle 3.1) transformiert. Mit Hilfe der Variablen „Anonymisierte Fahrgestellnummer (VAN)“ und „Session-ID“ ist es möglich, Fahrten mit einem Fahrzeug zu verknüpfen. Dadurch kann die Fahrzeugnutzung über einen bestimmten Zeitraum beobachtet werden. Sensordaten werden durch eine Liste von vordefinierten Events getriggert („Trigger“), wodurch der Anfang und das Ende einer Fahrt bestimmt werden kann. Die Fahrdistanz definiert sich über den „Kilometerstand“ des gegenwärtigen und vorherigen Triggers. Ein Vorteil der Sensordaten ist, dass eine Vielzahl an Informationen, wie z. B. Funktionsnutzungsstatistiken, gespeichert werden. Beispielsweise ist es möglich, über die „Sitzplatzbelegung“ die Anzahl der Passagiere (Besetzungsgrad) für eine Fahrt zu bestimmen.

Tabelle 3.1: *Sensordaten: Datenstruktur*

Variable	Beschreibung	Datenformat
Anonymisierte Fahrgestellnummer	Eindeutige Fahrgestellnummer	String
Fahrt-ID	Eindeutige Fahrtnummer	String
Tagesstempel	Tag der Fahrt	yyyy-MM-dd
Trigger	Vordefinierte Liste von Events	Kategorisch
Kilometerstand	Gefahrene Kilometer des Fahrzeugs	Integer
Sitzplatzbelegung	Anzahl der Passagiere	Kategorisch

Nachdem alle relevanten Informationen aus den Sensordaten abstrahiert worden sind, finden im nächsten Schritt die Bereinigung und Bearbeitung der Daten statt, um falsche Datenpunkte auszuschließen. Folgende Kriterien werden hierfür implementiert, um die Verlässlichkeit der Daten zu garantieren: Um auszuschließen, dass keine neuen Fahrzeuge während des Beobachtungszeitraums hinzugefügt wurden, müssen vom Fahrzeug Bewegungsdaten vor dem

tatsächlichen Beobachtungszeitraum gesendet worden sein. Als Nächstes werden alle Fahrzeuge mit unplausiblen Bewegungsdaten, aufgrund der großen Stichprobe, aus dem Datensatz ausgeschlossen. Hierfür wurden Fahrzeuge mit extremen Fahrdistanzen (durchschnittliche Fahrdistanz von über 200 km) aus dem Datensatz entfernt. Die Gründe für solche Ausreißer können zum einen ein Fehler in der Datenübertragung sein, zum anderen kann es sich um Fahrzeuge zur Personenbeförderung handeln. Solche Fahrzeuge und ihre Nutzung sind nicht Teil der Forschungsfrage, weshalb sie aus dem Datensatz ausgegrenzt wurden. Insgesamt erfüllen 0,17 % der beobachteten Fahrzeuge mit 4394 Fahrten die beschriebenen Ausschlusskriterien. Mit dem Ziel, die Machbarkeit der Einordnung von Premiumfahrzeugen in die FNP von Eisenmann und Buehler (2018) zu demonstrieren, wurden Fahrzeugnutzungsdaten über einen Beobachtungszeitraum von einem Monat (Mai 2018) gesammelt. In Summe entspricht dies 74 568 Premiumfahrzeuge mit 5 380 763 individuellen Fahrten mit validem Kilometerstand. In Deutschland sind 45 590 Fahrzeuge mit 2 950 967 Fahrten und in Kalifornien sind 28 978 Fahrzeuge mit 2 429 796 Fahrten beobachtet worden.

Datenvergleich

Um die Daten besser zu verstehen, wurden wichtige Kennzahlen, wie die zurückgelegten Distanzen und die Anzahl der Fahrten, mit den beschriebenen Studien aus Abschnitt 3.1.2 verglichen. Bäumer et al. (2018) stellen eine Jahresfahrleistung zwischen 13 000 km und 15 000 km fest. Wird eine lineare Beziehung zwischen Monats- und Jahresfahrleistung unterstellt, beträgt die Jahresfahrleistung von Premiumfahrzeugen 17 676 km (1473 km im Monat). Der Unterschied kann dadurch erklärt werden, dass es sich bei den Premiumfahrzeugen um relativ junge Fahrzeuge handelt. Eisenmann (2019) beobachtet eine durchschnittliche Monatsfahrleistung von ungefähr 1500 km. Dieser Durchschnitt stimmt mit der Monatsfahrleistung (1473 km) von Premiumfahrzeugen fast überein. Pasaoglu et al. (2012) berechnen eine durchschnittliche Tagesfahrleistung, abhängig vom Wochentag, zwischen 50 km und 60 km in Deutschland. Die durchschnittliche Tagesfahrleistung von Premiumfahrzeugen ist mit 61,3 km leicht oberhalb angesiedelt. Der „Premiumfaktor“ könnte diesen größeren Wert erklären. Des Weiteren werden durchschnittlich 2,0 Fahrten pro Tag in Deutschland durchgeführt. Bei den Premiumfahrzeugen hingegen sind es durchschnittlich 2,2 Fahrten pro Tag.

Alles in allem sind die verfügbaren Sensordaten und ihre Kennzahlen konsistent mit den Ergebnissen bestehender Studien, da keine starken Abweichungen existieren.

3.1.4 Methodik

In diesem Abschnitt wird das verwendete Gerüst, bestehend aus den acht FNPs und deren Eigenschaften, vorgestellt. Danach wird der Selektionsprozess der angewandten Algorithmen präsentiert und die Implementierung sowie Validierung des verwendeten Algorithmus beschrieben.

Acht Fahrzeugnutzungsprofile

Basierend auf den modellierten Fahrzeugnutzungsdaten von CUMILE-GER und CUMILE-CAL, wenden Eisenmann und Buehler (2018) ein hierarchisches Clustering an, um acht FNPs zu identifizieren und zu interpretieren. Die Eigenschaften der FNPs berücksichtigen die Intensität und Variabilität der Dimensionen Zeit und Distanz. Die Autoren nutzen vier Faktoren um die Fahrzeuge in Gruppen einzuordnen:

1. Anzahl der Tage, an denen nicht gefahren wurde (Tage0)
2. Anzahl der Tage, an denen zwischen 1 Meile und 10 Meilen gefahren wurde (Tage10)
3. Anzahl der Tage, an denen über 100 Meilen gefahren wurde (Tage100)
4. Verhältnis der Fahrzeugnutzung an Werktagen und Wochenenden (AntWerktag)

Aus den vier Inputvariablen ergeben sich acht unterschiedliche FNPs:

1. Stillstandfahrzeug: keine Nutzung an durchschnittlich 256 Tagen im Jahr
2. Moderat genutztes Fahrzeug: regelmäßige Nutzung an Werktagen sowie am Wochenende
3. Täglich genutztes Fahrzeug: nahezu tägliche Nutzung (Tage0 (\emptyset): 36 Tage)
4. Werktagsfahrzeug: hauptsächliche Nutzung an Werktagen (AntWerktag (\emptyset): 96%)
5. Wochenendfahrzeug: überdurchschnittliche Nutzung am Wochenende
6. Langstreckenfahrzeug: primäre Nutzung für längere Tagesdistanzen
7. Kurzstreckenfahrzeug: primäre Nutzung für kürzere Tagesdistanzen
8. Allrounder: durchschnittliche Nutzung

Eisenmann und Buehler (2018) liefern eine detaillierte Beschreibung der einzelnen FNPs. Die präsentierten FNPs werden im nächsten Schritt als vergleichendes Instrument verwendet. In anderen Worten: Die Nutzung der Premiumfahrzeuge (aus Sensordaten) wird betrachtet, um diese in die entsprechenden FNPs einzuteilen. In Konsequenz ist es möglich diese mit gewöhnlichen Fahrzeugen (aus Umfragedaten) zu vergleichen. Die FNPs werden als sogenannte Labels genutzt, um die Premiumfahrzeuge mit Hilfe eines Klassifikationsalgorithmus zuzuordnen.

Zuteilung der Premiumfahrzeuge zu bestehenden Fahrzeugnutzungsprofilen

Für diesen Ansatz werden die Sensordaten in das Format der vier Inputvariablen transformiert. Um die identifizierten längeren Fahrdistanzen bei Premiumfahrzeugen zu berücksichtigen, werden die zurückgelegten Distanzen über den gesamten Beobachtungszeitraum (VKT) im Klassifikationsprozess miteinbezogen. Ein weiterer Vorteil der Einbindung von VKT ist, dass es sich hierbei nicht um eine modellierte, sondern um eine tatsächliche Aussage der Teilnehmer handelt. Somit wird der Einfluss von VKT auf die Einteilung der FNPs von Eisenmann und Buehler (2018) in das Klassifikationsmodell integriert. In Konsequenz können Fahrzeuge mit einer größeren Fahrleistung in das entsprechende FNP eingeordnet werden. Diese Maßnahme zieht die spezielle Nutzung von Premiumfahrzeugen mit in Betracht und liefert eine weitere erklärende Variable zur Einordnung in FNPs.

Das Modell wird somit auf der Basis von fünf Inputvariablen (Tage0, Tage10, Tage100, AntWerktag und VKT) trainiert:

1. Anzahl der Tage, an denen nicht gefahren wurde (Tage0)
2. Anzahl der Tage, an denen zwischen 1 km und 10 km gefahren wurde (Tage10)
3. Anzahl der Tage, an denen über 100 km gefahren wurde (Tage100)
4. Verhältnis der Fahrzeugnutzung an Werktagen und Wochenenden (AntWerktag)
5. Zurückgelegte Distanz über den Beobachtungszeitraum (VKT)

Da nun fünf statt vier Variablen für die Modellierung verwendet werden, ist es nicht mehr möglich, das Clusterverfahren (hierarchisches Clustering und k-Means) von Eisenmann und Buehler (2018) mit den entsprechenden Zentroiden nachzubilden.

Zur Einteilung der Premiumfahrzeuge in FNPs und zum Vergleich mit gewöhnlichen Fahrzeugen wird deshalb eine Methode des maschinellen Lernverfahrens verwendet. Um die am besten geeignete Methode für den vorliegenden Datensatz auszuwählen, wurden gängige lineare und nicht lineare Klassifikationsverfahren trainiert und die jeweilige Präzision bestimmt. Die Präzision der Schätzung erklärt sich über den Anteil der Vorhersagen, welche das Modell richtig prognostiziert haben. Die Ergebnisse sind in Tabelle 3.2 dargestellt. Alle linearen Modelle haben eine geringere Präzision im Vergleich zur quadratischen Support Vector Machine (SVM). Lediglich die kubische SVM weist eine höhere Präzision mit 96,3 % auf. Oftmals ist es sinnvoller, ein Verfahren zu wählen, welches nicht perfekt alle Daten in die richtigen Cluster (hier FNPs) einteilt, um eine robuste Generalisierung des Modells zu garantieren (James et al., 2013). Durch die allgemein hohe Präzision der Schätzung bringt das von Eisenmann und Buehler (2018) gewählte Klassifikationsverfahren nur einen geringen Vorteil. Bei der Klassifikation der gewöhnlichen Fahrzeuge auf der Basis von fünf clusterbildenden Variablen erreicht das quadratische SVM-Verfahren eine Präzision größer 95 %. Das bestätigt die Entscheidung VKT als zusätzliche Variable zur besseren Einteilung der Premiumfahrzeuge in

das Klassifikationsmodell mit aufzunehmen. Aufgrund der geringen Abweichung der Präzision zwischen quadratischer und kubischer SVM und um eine Überanpassung des Modells an die Trainingsdaten zu vermeiden, wird die quadratische SVM verwendet.

Tabelle 3.2: Modellauswahl

Methode	Präzision
Lineare Diskriminanzanalyse	89,1 %
Quadratische Diskriminanzanalyse	90,5 %
Lineare Support Vector Machine	91,5 %
Quadratische Support Vector Machine	95,5 %
Kubische Support Vector Machine	96,3 %
Fine Tree	92,6 %
Medium Tree	94,0 %
Coarse Tree	61,2 %

Die grundlegende Theorie von SVM besteht darin, optimale Hyperebenen zu finden, um die Fahrzeuge in das zugehörige Cluster (FNP) einzuteilen. Hyperebenen trennen die Cluster im Datenraum voneinander. Die Hyperebenen werden bestimmt, indem die Spanne zwischen den Clustern maximiert wird. Ein Vorteil von SVM im Vergleich zu anderen Klassifikationsverfahren ist, dass die Bestimmung der Hyperebenen auf einer relativ kleinen Anzahl von Datenpunkten nahe der Hyperebenen basiert. In anderen Worten: SVM ist robust gegenüber Ausreißern und effizient zu berechnen, da nicht alle Datenpunkte für den Klassifikationsprozess berücksichtigt werden müssen (Ding, Zhu & Zhang, 2017; James et al., 2013).

Als Erstes wird ein Modell trainiert, um den Zusammenhang zwischen den gewöhnlichen Fahrzeugen (mit einem Beobachtungszeitraum von einem Jahr) und den FNPs zu erlernen. Danach wird mit Hilfe des erlernten Modells jedes Premiumfahrzeug (mit einem Beobachtungszeitraum von einem Monat) einem FNP zugewiesen. Um die unterschiedlichen Beobachtungszeiträume zu berücksichtigen, werden die clusterbildenden Variablen durch eine z -Transformation standardisiert. Damit ist es möglich, die Premiumfahrzeuge mit den z -Werten zuzuteilen. Im Allgemeinen misst der z -Score ($z = (x - \bar{X})/S$) die Abweichung eines Datenpunktes x von seinem Mittelwert \bar{X} in Relation zur Standardabweichung S . Der standardisierte Datensatz hat einen Mittelwert von 0 und eine Standardabweichung von 1. Die z -transformierten Variablen weisen die gleiche Verteilungsfunktion auf und besitzen die gleiche Steigung und Krümmung wie die nicht transformierten Variablen.

Die SVM wurde ursprünglich zur Lösung von binären Klassifikationsproblemen ($k=2$) verwendet. Um multiple Klassifikationsprobleme ($k > 2$) zu lösen und nicht klassifizierbare Regionen zu vermeiden, wird ein sogenanntes Error-Correcting-Output-Code (ECOC) implementiert. Dadurch reduziert sich das multiple Klassifikationsproblem auf eine Serie von binären

Problemen. Dabei repräsentiert jede Zeile der Coding-Matrix $M \in \{1, -1, 0\}$ ein FNP und eine Spalte einen binären „Lerner“. Die Anzahl der Hyperebenen ist abhängig von k und wird durch $k(k-1)/2$ bestimmt. In diesem Fall entspricht die Größe der Coding-Matrix $M = 8 \times 28$. Somit müssen 28 binäre anstatt acht multiple Klassifikationsprobleme gelöst werden. Für mehr Details zum ECOC-Ansatz empfiehlt sich die Arbeit von Dietterich und Bakiri (1994). Nachdem die ECOC-Matrix für jede Klasse bestimmt worden ist, wird ein One-vs-one-Ansatz genutzt. Bei diesem Ansatz werden alle Klassen individuell miteinander verglichen. Das resultierende Label des Testdatensatzes ist das Label, welches am häufigsten dem Fahrzeug zugeordnet wird (Abe, 2005; James et al., 2013).

Der Modellinput entspricht dem gelabelten Datensatz der gewöhnlichen Fahrzeuge $C^{(l)} = \{(\mathbf{x}_i^{(l)}, y_i^{(l)})\}_{i=1}^{n_l}$ mit $\mathbf{x}_i^{(l)} = (x_{1;i}^{(l)}, \dots, x_{5;i}^{(l)}) \in \mathbb{R}^5$ und $y_i^{(l)} = (y_{1;i}^{(l)}, \dots, y_{s;i}^{(l)}) \in \{1, 2, \dots, 8\}$. Das Ziel des Verfahrens ist es, eine Regel zu konstruieren $(y_i^{(l)} = f(\mathbf{x}_i^{(l)}) \in \mathbb{R}^S \rightarrow \{1, 2, \dots, 8\})$, sodass für neue Datenpunkte $\mathbf{x}_j^{(u)} \in \mathbb{R}^d$ ein Label $y_j^{(u)} = f(\mathbf{x}_j^{(u)}) \in \{1, 2, \dots, 8\}$ bestimmt und ein gelabelter Datensatz mit Premiumfahrzeugen $P^{(u)} = \{(\mathbf{x}_j^{(u)}, y_j^{(u)})\}_{j=1}^{n_u}$ generiert werden kann.

- $\mathbf{x}_i^{(l)}$ ist der Inputvektor bestehend aus den fünf beschriebenen clusterbildenden Variablen für das i-te gewöhnliche Fahrzeug.
- $y_i^{(l)}$ ist das FNP für das i-te gewöhnliche Fahrzeug.
- $\mathbf{x}_j^{(u)}$ ist der Inputvektor bestehend aus den fünf beschriebenen clusterbildenden Variablen für das j-te Premiumfahrzeug.
- $y_j^{(u)}$ ist das vorhergesagte FNP für das j-te Premiumfahrzeug.

Der Funktionsraum d wird mit Hilfe eines quadratischen Kernels $K(\mathbf{x}_i, \mathbf{x}_i') = (\langle \mathbf{x}_i, \mathbf{x}_i' + d \rangle)^2$ erweitert, damit mehr Flexibilität durch nicht lineare Hyperebenen zwischen den Klassen geschaffen werden kann (Hastie, Tibshirani & Friedman, 2009; James et al., 2013). Um die Präzision der Vorhersagen von $y_i^{(l)} = f(\mathbf{x}_i^{(l)})$ zu messen und eine Überanpassung³ zu vermeiden, wird ein fünffaches Kreuzvalidierungsverfahren integriert. Dabei handelt es sich um eine statistische Methode, welche den Datensatz zufällig in fünf gleiche Teildatensätze aufteilt. Die Funktion $f(\cdot)$ wird dann auf der Basis eines Teildatensatzes trainiert und die restlichen Teildatensätze werden zur Evaluierung verwendet. Um die berechneten Hyperebenen und deren Ergebnisse zu validieren, wurde eine Konfusionsmatrix mit den finalen Parametereinstellungen gebildet. Die Konfusionsmatrix beinhaltet den Vergleich zwischen geschätzten und tatsächlichen Labels. Die Diagonale beschreibt richtig klassifizierte FNPs, während alle Werte abseits der Diagonale falsche Klassifizierungen darstellen.

³Überanpassung im Bereich des maschinellen Lernens bedeutet, dass das Modell sehr gute Vorhersagen für den trainierten Datensatz trifft, allerdings bei neuen Daten schlechte Ergebnisse erzielt. Deswegen wird der eigentliche Datensatz in Trainings-, Validierungs- und Testdaten aufgeteilt. Mit Hilfe der Trainings- und Validierungsdaten wird das Modell $f(\cdot)$ trainiert und angepasst. Der Testdatensatz prüft dann die „Generalisierung“ des Modells, da der Testdatensatz nicht in die Bestimmung von $f(\cdot)$ involviert war (Géron, 2019).

Fahrzeugnutzungsprofile		1	2	3	4	5	6	7	8	Tatsächliche Klassen
Stillstandfahrzeug	1	94 %			4 %	2 %	1 %			
Moderat genutztes Fahrzeug	2		97 %	1 %		1 %	2 %		<1 %	
Täglich genutztes Fahrzeug	3		< 1 %	96 %		< 1 %	1 %	1 %	1 %	
Werktagsfahrzeug	4	4 %	< 1 %		95 %				1 %	
Wochenendfahrzeug	5	2 %	1 %	< 1 %		94 %	2 %			
Langstreckenfahrzeug	6	< 1 %	2 %	< 1 %		1 %	93 %			
Kurzstreckenfahrzeug	7			2 %				99 %	2 %	
Allrounder	8		< 1 %	1 %	< 1 %	1 %	< 1 %		96 %	
Vorhergesagte Klassen										
Positiv vorhergesagter Wert		94 %	97 %	96 %	95 %	94 %	93 %	99 %	96 %	
Falscherkennungsrate		6 %	3 %	4 %	5 %	6 %	7 %	1 %	4 %	

Abbildung 3.2: Konfusionsmatrix

In Abbildung 3.2 werden die richtig und falsch klassifizierte Werte über alle acht Klassen aufgezeigt. Die niedrigste Präzision weist Klasse 6 (Langstreckenfahrzeug) mit 93 % auf. Diese Klasse hat gleichzeitig auch die geringste Größe. Die größte Präzision ist bei der Zuteilung von Kurzstreckenfahrzeugen mit 99 % vorhanden. Unter der Annahme, dass die Klassen der gewöhnlichen Fahrzeuge $\{(\mathbf{x}_i^{(l)}, y_i^{(l)})\}_{i=1}^{3340}$ unabhängig und identisch verteilt sind, kann mit Hilfe des trainierten Modells jedes Premiumfahrzeug eindeutig einem FNP zugewiesen und ein gelabelter Sensordatensatz $P^{(u)} = \{(\mathbf{x}_j^{(u)}, y_j^{(u)})\}_{j=1}^{74568}$ generiert werden. Unter der Annahme des One-vs-one-Ansatzes wurden $\binom{8}{2} = \frac{8!}{2!(8-2)!} = 28$ SVMs gebildet. Für die Entwicklung des Modells wurde die Software MATLAB verwendet.

3.1.5 Ergebnisse

Basierend auf dem Klassifikationsalgorithmus (quadratische SVM) konnten Premiumfahrzeuge in eines der acht FNPs eingeteilt werden. Um die generelle Vergleichbarkeit der Datenquellen (Umfrage- und Sensordaten) zu garantieren, werden die fünf clusterbildenden Variablen im Durchschnitt miteinander verglichen. Als Nächstes wird die kumulative Verteilung der zurückgelegten Distanzen über einen Beobachtungsraum von einem Monat dargestellt. Danach werden die Charakteristika der einzelnen FNPs von Premiumfahrzeugen beschrieben und die relativen Größen mit der Zuordnung der FNPs gewöhnlicher Fahrzeuge in Deutschland und

Kalifornien betrachtet. Als Letztes werden die zusätzlichen Informationen zur Sitzplatzbelegung für Premiumfahrzeuge genutzt, um die durchschnittliche Anzahl von Passagieren für die jeweiligen FNP's zu bestimmen.

Tabelle 3.3 beschreibt den relativen Durchschnitt der fünf clusterbildenden Variablen. Um vergleichbare Ergebnisse zu generieren, wurde die absolute Anzahl an Tagen für Tage0, Tage10 und Tage100 durch den jeweiligen Beobachtungszeitraum geteilt. Die relative Anzahl von Tagen, an denen das Fahrzeug nicht bewegt wird (Tage0), ist für gewöhnliche Fahrzeuge größer als für Premiumfahrzeuge. Das ist nachvollziehbar, da Sensordaten hauptsächlich junge Premiumfahrzeuge und Umfragedaten sowohl junge als auch alte gewöhnliche Fahrzeuge beinhalten. Die relative Anzahl von Tagen, an denen eine Tagesfahrleistung größer 100 km (Tage100) zurückgelegt wird, ist für Premiumfahrzeuge größer als für gewöhnliche Fahrzeuge. Auf der einen Seite kann dies durch den Premiumfaktor erklärt werden, auf der anderen Seite werden möglicherweise gewöhnliche Fahrzeuge nicht zwingend für den Fernverkehr eingesetzt (z. B. Nutzung als Zweitfahrzeug, kleine oder alte Fahrzeuge). Darüber hinaus handelt es sich bei den gewöhnlichen Fahrzeugen in Deutschland um einen modellierten Wert (CUMILE-GER), welcher womöglich den tatsächlichen Fernverkehr unterschätzt. Dieser Wert wird zufällig über die Erhebung INVERMO der modellierten Nutzung zugespielt. Diese Annahme wird durch den ähnlichen Wert zwischen Premium- und gewöhnlichen Fahrzeugen in Kalifornien, welcher nicht auf einer Modellierung basiert, bestätigt. Dasselbe Bild zeigt sich bei der Betrachtung der zurückgelegten Distanzen (VKT) (Premiumfahrzeuge in Kalifornien: 1523 km; gewöhnliche Fahrzeuge in Kalifornien: 1572 km). Das Verhältnis der Fahrzeugnutzung an Werktagen und Wochenenden ist für Premiumfahrzeuge und gewöhnliche Fahrzeuge ähnlich.

Tabelle 3.3: Relativer Durchschnitt der fünf clusterbildenden Variablen

Clusterbildende Variable	Premiumfahrzeuge		Gewöhnliche Fahrzeuge	
	Deutschland	Kalifornien	Deutschland	Kalifornien
Tage0 (relativ)	26 %	20 %	39 %	34 %
Tage10 (relativ)	7 %	7 %	17 %	11 %
Tage100 (relativ)	17 %	14 %	7 %	13 %
AntWerktag	77 %	77 %	79 %	75 %
VKT	1804 km	1523 km	1002 km	1.572 km

Abbildung 3.3 zeigt die Verteilung der zurückgelegten Distanzen von Premium- und gewöhnlichen Fahrzeugen über einen Beobachtungszeitraum von einem Monat. Allgemein haben die beiden Kurven einen ähnlichen Verlauf.

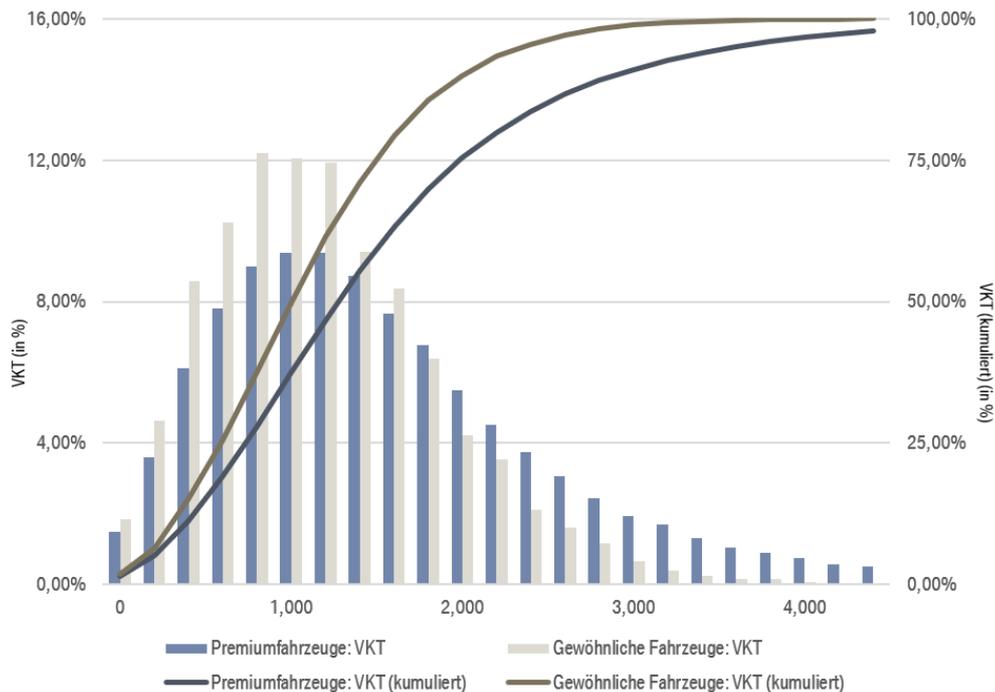


Abbildung 3.3: VKT zwischen Premium- und gewöhnlichen Fahrzeugen (monatlich)

Unterschiede ergeben sich im Wesentlichen, da Premiumfahrzeuge öfter größere Distanzen zurücklegen. Die kumulierte Verteilungsfunktion gewöhnlicher Fahrzeuge konvergiert wesentlich schneller gegen 100 % als die der Premiumfahrzeuge. Die Unterschiede könnten wiederum auf den Premiumfaktor und die potenzielle Unterschätzung des Fernverkehrs in Deutschland (CUMILE-GER) zurückzuführen sein.

In Abbildung 3.4 werden die Charakteristika für Premiumfahrzeuge anhand des Mittelwertes (MW) und der dazugehörigen Standardabweichung (StA) dargestellt.

- Das „Stillstandfahrzeug“ wird an nahezu der Hälfte der beobachteten Tage nicht bewegt. Wenn es aber genutzt wird, findet dies primär an Werktagen und im Fernverkehr statt (durchschnittliche Distanz pro Fahrt: 36 km; Monatsfahrleistung: 1485 km).
- Das „moderat genutzte Fahrzeug“ hat eine unterdurchschnittliche Nutzung. Dieses FNP weist sonst keine auffälligen Merkmale auf.
- Das „täglich genutzte Fahrzeug“ wird fast täglich genutzt. Dieses Cluster hat den niedrigsten Durchschnitt an Tagen, an denen das Fahrzeug nicht bewegt wurde.
- Das „Werktags-“ bzw. „Wochenendfahrzeug“ wird primär an Werktagen bzw. am Wochenende bewegt.
- Das „Langstrecken-“ bzw. „Kurzstreckenfahrzeug“ definiert sich über die zurückgelegten Tagesdistanzen. Langstreckenfahrzeuge legen an 13 von 30 beobachteten Tagen Tagesdistanzen von mehr als 100 km zurück. Zudem haben sie die zweithöchste durchschnittliche Fahrdistanz (35 km) und Anzahl von Fahrten im Beobachtungszeitraum (98 Fahrten).

Tabelle 3.4: Premiumfahrzeuge: Eigenschaften der verschiedenen Fahrzeugnutzungsprofile

Fahrzeugnutzungsprofil	Tage0		Tage10		Tage100	
	MW	(StA)	MW	(StA)	MW	(StA)
Stillstandfahrzeug	14	(5)	1	(1)	5	(4)
Moderat genutztes Fahrzeug	4	(3)	1	(1)	3	(2)
Täglich genutztes Fahrzeug	2	(2)	3	(1)	4	(3)
Werktagsfahrzeug	10	(4)	2	(2)	4	(3)
Wochenendfahrzeug	14	(4)	2	(1)	3	(3)
Langstreckenfahrzeug	3	(3)	1	(1)	13	(4)
Kurzstreckenfahrzeug	6	(4)	9	(2)	2	(2)
Allrounder	8	(3)	4	(1)	4	(4)
Durchschnitt	7	(5)	2	(2)	5	(5)

Fahrzeugnutzungsprofil	AntWerktag		VKT	
	MW	(StA)	MW	(StA)
Stillstandfahrzeug	88 %	(8 %)	1485	(1034)
Moderat genutztes Fahrzeug	77 %	(5 %)	1599	(631)
Täglich genutztes Fahrzeug	72 %	(3 %)	1716	(795)
Werktagfahrzeug	91 %	(4 %)	1451	(758)
Wochenendfahrzeug	68 %	(12 %)	1070	(744)
Langstreckenfahrzeug	75 %	(6 %)	3291	(1179)
Kurzstreckenfahrzeug	77 %	(7 %)	942	(662)
Allrounder	78%	(6 %)	1386	(949)
Durchschnitt	77 %	(9 %)	1695	(1088)

Allgemein gilt, dass einzelne FNP's sich inhaltlich überschneiden. Beispielsweise kann ein Wochenendfahrzeug gleichzeitig auch die Kriterien eines Langstreckenfahrzeugs erfüllen. Durch die Betrachtung der zusätzlichen Variable (VKT) können die FNP's noch präziser voneinander getrennt werden. Dadurch sind z. B. Wochenendfahrzeuge (durchschnittliche Distanz: 29 km) anhand der geringeren Distanzen besser von Langstreckenfahrzeugen (durchschnittliche Distanz: 35 km) abgrenzbar.

Der Vergleich (Tabelle 3.5) der relativen Größen der FNP's von Premiumfahrzeugen (PF) mit denen gewöhnlicher Fahrzeuge (GF) in Deutschland und Kalifornien zeigt einige Unterschiede auf. Die gewöhnlichen Langstreckenfahrzeuge (DE: 3 %; KA: 14 %) sind viel stärker in Kalifornien vertreten als in Deutschland. Dort ist der Schienenverkehr sehr gut ausgebaut, speziell im Fernverkehr zwischen größeren Städten (z. B. München und Stuttgart). Aufgrund dessen ist

es Individuen möglich im Fernverkehr ihr altes und nicht so komfortables Fahrzeug durch den Deutsche-Bahn (DB)-Fernverkehr zu ersetzen. Darüber hinaus könnte der Unterschied auch durch den höheren Anteil an geschäftlich genutzter Premiumfahrzeuge begründet werden. Die bessere Qualität des ÖPNV erklärt möglicherweise auch den größeren Anteil an täglich genutzter Fahrzeuge (DE: 12 %; KA: 18 %) in Kalifornien. Beispielsweise zeigen Magdolen, von Behren, Chlond, Hunecke und Vortisch (2019) auf, dass speziell in urbanen Gebieten Individuen in Deutschland auch auf andere Verkehrsmittel zurückgreifen können, um den täglichen Mobilitätsbedarf zu decken. Der Anteil von Werktagsfahrzeugen ist in Deutschland (14 %) größer als in Kalifornien (8 %). Eine Interpretation des Unterschiedes ist vielleicht der bereits angesprochene hohe Anteil an geschäftlich genutzter Fahrzeuge in Deutschland.

Generell werden Premiumfahrzeuge unabhängig von ihrem räumlichen Kontext ähnlich genutzt. Beispielsweise haben Kurzstreckenfahrzeuge (DE: 5 %; KA: 4 %), Langstreckenfahrzeuge (DE: 14 %; KA: 14 %) und täglich genutzte Fahrzeuge (DE: 8 %; KA: 8 %) ähnliche Clustergrößen für neue Premiumfahrzeuge in Deutschland und Kalifornien. Im Lebenszyklus eines Fahrzeugs durchläuft ein Fahrzeug verschiedene FNPs, besonders in Deutschland. Da es sich bei Premiumfahrzeugen hauptsächlich um junge Fahrzeuge handelt, werden diese auch relativ ähnlich genutzt. Der Zusammenhang zwischen dem Fahrzeugalter und der Fahrzeugnutzung ist ein weiteres Argument für die zusätzliche Einbindung der Variable „zurückgelegte Distanzen (VKT)“. Das ermöglicht eine klarere Trennung der Premiumfahrzeuge und deren Nutzung.

Tabelle 3.5: Vergleich von Premiumfahrzeugen und gewöhnliche Fahrzeugen

Fahrzeugnutzungsprofil	Deutschland (DE)		
	PF (Häufigkeit)	PF (Anteil)	GF (Anteil)
Stillstandfahrzeug	5187	11 %	16 %
Moderat genutztes Fahrzeug	12 971	28 %	17 %
Täglich genutztes Fahrzeug	3865	8 %	12 %
Werktagsfahrzeug	2182	5 %	14 %
Wochenendfahrzeug	6315	14 %	15 %
Langstreckenfahrzeug	6354	14 %	3 %
Kurzstreckenfahrzeug	2057	5 %	9 %
Allrounder	6659	15 %	14 %
Σ	45 590	100 %	100 %

Kalifornien (KA)

Fahrzeugnutzungsprofil	PF (Häufigkeit)	PF (Anteil)	GF (Anteil)
Stillstandfahrzeug	2842	10 %	13 %
Moderat genutztes Fahrzeug	10 142	35 %	22 %
Täglich genutztes Fahrzeug	2227	8 %	18 %
Werktagsfahrzeug	953	3 %	8 %
Wochenendfahrzeug	3737	13 %	16 %
Langstreckenfahrzeug	4010	14 %	14 %
Kurzstreckenfahrzeug	1224	4 %	3 %
Allrounder	3843	13 %	6 %
Σ	28 978	100 %	100 %

Bei Betrachtung der Verteilungen für Fahrdistanzen (Abbildung 3.4) und zurückgelegte Fahrten (Abbildung 3.5) werden die Unterschiede in der Fahrzeugnutzung zwischen Kalifornien (KA) und Deutschland (DE) sichtbar. Abbildung 3.4 zeigt eine logarithmische Normalverteilung der durchschnittlich zurückgelegten Distanzen pro Fahrt von Premiumfahrzeugen. Die durchschnittliche gefahrene Distanz in Deutschland und Kalifornien beträgt 25,7 km. Werden Premiumfahrzeuge räumlich getrennt betrachtet, zeigen sich Unterschiede. Während in Deutschland die durchschnittliche Fahrdistanz 30 km beträgt, liegt diese in Kalifornien bei nur 19 km. Dafür legen Premiumfahrzeuge in Kalifornien größere Tagesdistanzen zurück.

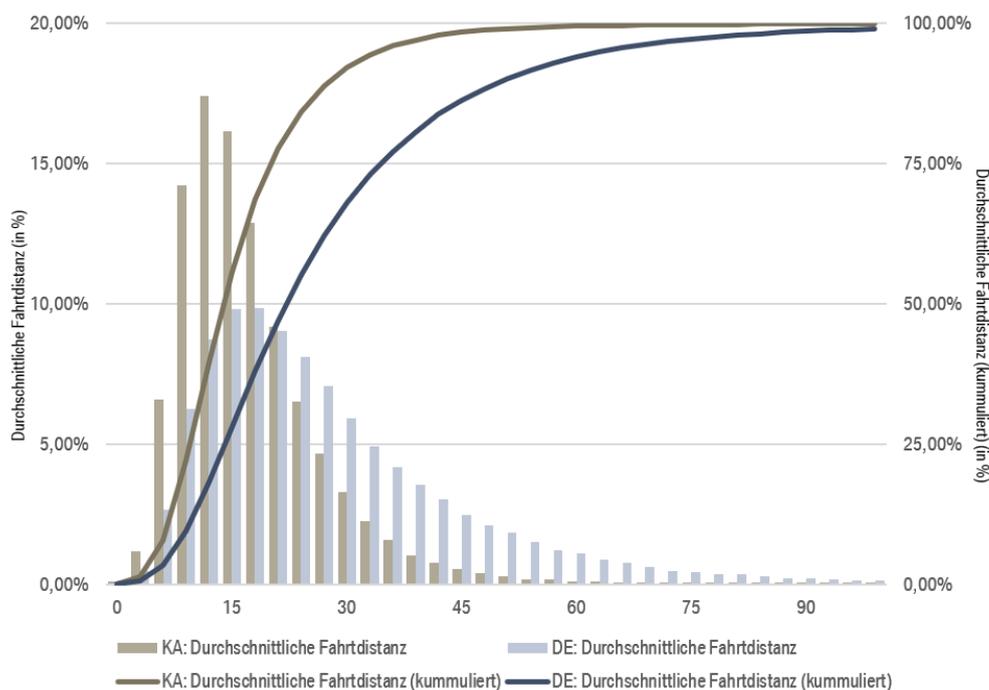


Abbildung 3.4: Premiumfahrzeuge: Verteilung für zurückgelegte Fahrdistanzen

Dieser Befund wird durch die Verteilung der zurückgelegten Fahrten in Abbildung 3.5 ergänzt. In Deutschland werden durchschnittlich 67,7 Fahrten in einem Beobachtungszeitraum von einem Monat zurückgelegt. In Kalifornien sind es hingegen 93,9 Fahrten. Dieser Unterschied könnte durch die räumlichen Gegebenheiten und die Stellung des Fahrzeugs in den jeweiligen Ländern erklärt werden. In Deutschland ist es nicht notwendig, alle Fahrtzwecke (z. B. Arbeit, Freizeit, Einkaufen) mit dem Fahrzeug zu erledigen, da der Zugang zum ÖPNV, vor allem in urbanen Räumen, besser ausgebaut ist als in Kalifornien. Das spiegelt sich in der geringen Anzahl zurückgelegter Fahrten und den längeren Fahrdistanzen wider. Darüber hinaus könnten die Kalifornier eine stärkere Orientierung zum Auto, besonders zu Premiumfahrzeugen, als die Deutschen aufweisen. Basierend auf soziodemografischen Informationen aus den USA nutzt Hildebrand (2003) ein hierarchisches Clustering zur Bestimmung der optimalen Anzahl von Clustern. Außerdem verwendet er den k-Means-Algorithmus, um Individuen in Cluster einzuteilen (die gleiche Methode verwendet auch Eisenmann und Buehler (2018) für die Identifikation der dargestellten FNP's). Bei allen Cluster wird die Mehrheit der Fahrten mit dem Auto zurückgelegt. Auf der anderen Seite zeigt die Clusteranalyse von Haustein und Nielsen (2016) europäische Mobilitätstypen wie die „ÖPNV-Nutzer“ und die „grünen Fahrradfahrer“, welche keine Orientierung zum Pkw aufweisen. Das bestätigt auch die Arbeit von von Behren et al. (2018), welche die Mobilität in urbanen Räumen in San Francisco, Berlin und Schanghai untersuchen und eine größere Orientierung zum Auto in San Francisco im Vergleich zu Berlin feststellen.

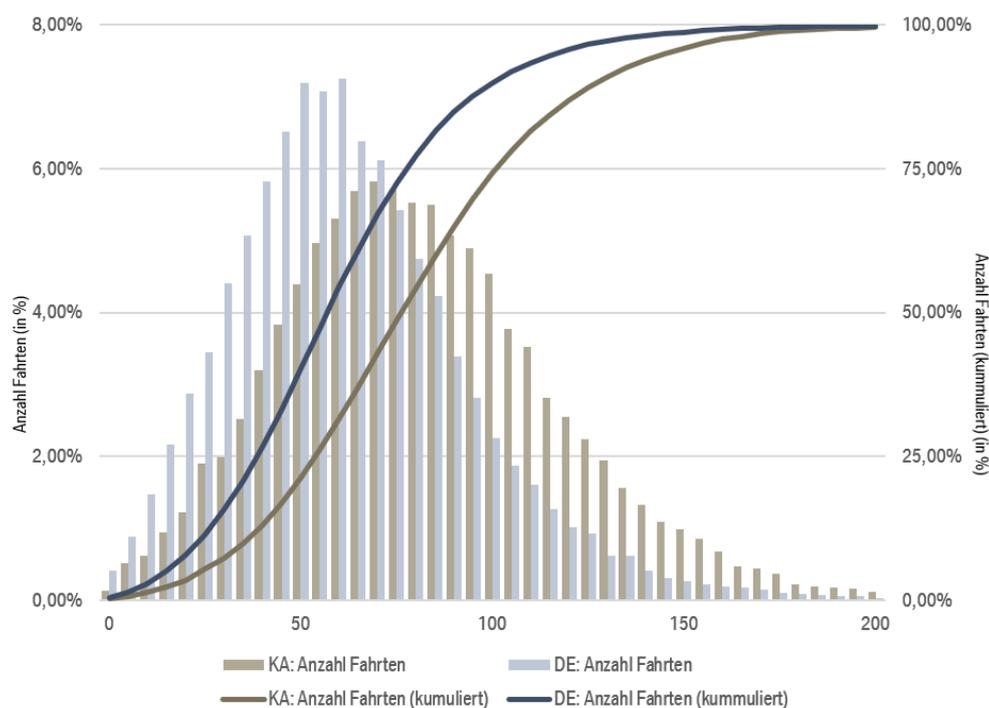


Abbildung 3.5: Premiumfahrzeuge: Verteilung der zurückgelegten Fahrten

Der Anteil der Stillstandfahrzeuge (DE: 11 %; KA: 10 %) ist für Premiumfahrzeuge geringer als für gewöhnliche Fahrzeuge (DE: 16 %; KA: 13 %). Trotzdem ist er relativ hoch vor dem Hintergrund, dass es sich hierbei um neue Premiumfahrzeuge handelt. Eine weitere Erklärung dafür könnte sein, dass neue Premiumfahrzeuge oftmals in Haushalten mit mehreren Fahrzeugen genutzt werden. Das würde die Pkw-Mobilitätsbedürfnisse eines Haushalts auf die verfügbaren Fahrzeuge verteilen. In Konsequenz ist die Nutzung der beobachteten Premiumfahrzeuge geringer. Moderat genutzte Premiumfahrzeuge (DE: 28 %; KA: 35 %) verzeichnen einen größeren Anteil in Kalifornien. Eine Erklärung dafür wäre der bereits beschriebene ÖPNV-Effekt. Kalifornier nutzen ihre Fahrzeuge in einem erweiterten Radius für verschiedene Zwecke (z. B. Arbeit, Einkaufen). Des Weiteren könnten gewöhnliche Fahrzeuge nur für bestimmte Fahrtzwecke eingesetzt werden, wenn sie beispielsweise als Zweitfahrzeug fungieren. Das würde auch den relativ geringen Anteil an moderat genutzten gewöhnlichen Fahrzeugen (DE: 17 %; KA: 22 %) erklären. Diese Annahme ist konsistent mit der Clustergröße von Werktagspremiumfahrzeugen (DE: 5 %; KA: 3 %). Der Anteil des Clusters ist sehr gering, weil Premiumfahrzeuge für verschiedene Fahrtzwecke und nicht speziell für Fahrten an Werktagen, wie gewöhnliche Fahrzeuge (DE: 14 %; KA: 13 %), eingesetzt werden. Der Anteil von Wochenendfahrzeugen (DE: 14 %, KA: 13 %) für Premiumfahrzeuge ist ähnlich zu gewöhnlichen Fahrzeugen. Langstreckenfahrzeuge und Allrounder (DE: 15 %; KA: 13 %) weisen premiumfahrzeugtypische Merkmale auf. Aufgrund des besseren Gesamtpakets (durch Komfort, Sicherheit, Technik, Material) bei Premiumfahrzeugen werden diese überdurchschnittlich für längere Distanzen genutzt. Das spiegelt sich auch im Vergleich der Anteile von Langstreckenfahrzeuge wider.

Die Einordnung der Premiumfahrzeuge in FNPs basierend auf gewöhnlichen Fahrzeugen bildet eine solide Grundlage, um Berechnungen mit weiteren Informationen aus Sensordaten zu machen. Im Folgenden werden Informationen zur Sitzplatzbelegung verwendet, um Unterschiede in der Belegungsrate entlang der einzelnen FNPs zu untersuchen. Die Anzahl der Passagiere definiert sich hier über die Anzahl der Personen, welche während der Fahrt den Sicherheitsgurt geschlossen haben. Sobald der Sicherheitsgurt während der Fahrt geschlossen wird, entspricht dies der Belegung eines Sitzplatzes. Schnallt sich ein Passagier während der Fahrt nicht an, spiegelt sich dies in der Sitzplatzbelegung nicht wider. Der Besetzungsgrad hat vier Ausprägungen: „1-0“: Fahrersitz ist belegt, „1-X“: Fahrersitz und mindestens ein Rücksitz sind belegt, „2-0“: Fahrer- und Beifahrersitz sind belegt, „2-X“: Fahrer- und Beifahrersitz sowie mindestens ein Rücksitz sind belegt. Abbildung 3.6 beschreibt die durchschnittliche Anzahl von Passagieren über alle beobachteten Premiumfahrzeuge und deren zugehöriges FNP. Mit Hilfe der Sensordaten werden Informationen zur Sitzplatzbelegung von jedem Fahrzeug für jede Fahrt aufgezeichnet. Die Fahrzeuge werden entsprechend ihrem FNP gruppiert und die durchschnittliche Anzahl von Passagieren für jedes FNP berechnet. 71 %

der Fahrten von Werktagspremiumfahrzeugen werden allein gemacht. Möglicherweise erfüllen die meisten Fahrten den Zweck des Pendelns (z. B. zur Arbeit). Diese Pendelbewegungen finden womöglich hauptsächlich allein statt. In Kalifornien finden nur 68,2% der Fahrten von Werktags Fahrzeugen allein statt. Die niedrigere Rate könnte durch nachfrageorientierte Maßnahmen (z. B. HOV-Lanes) erklärt werden. Durch HOV-Lanes wird speziell die Bildung von Fahrgemeinschaften durch einen gesonderten Fahrstreifen gefördert. Des Weiteren zeigen Werktagsfahrzeuge den geringsten Anteil an Fahrten mit Fahrer und Beifahrer (18,6%). Das Gegenteil zeigt sich bei Wochenendfahrzeugen. Diese haben den größten Anteil an Fahrten mit Fahrer und Beifahrer (29,6%) und den geringsten Anteil an Fahrten nur mit Fahrer (57,6%). Freizeitfahrten, welche vermehrt am Wochenende stattfinden, finden eher zu zweit als allein statt. Täglich genutzte Fahrzeuge haben den höchsten Anteil an Fahrten mit Belegung der Rückbank (15% = 5,1% + 9,9%). Hier könnte das Fahrzeug vor allem den Zweck des Holens und Bringens von Personen erfüllen. Die Zusammensetzung der Sitzplatzbelegung bei Langstrecken- und Kurzstreckenfahrzeugen unterscheidet sich nur minimal. Das könnte implizieren, dass die Anzahl der Passagiere stärker vom Zweck der Fahrt als von der zurückgelegten Distanz abhängt. Darüber hinaus bestätigen die zusätzlichen Informationen über die Sitzplatzbelegung für Premiumfahrzeuge die Intuition der FNPs. Des Weiteren untermauern die Ergebnisse, dass die Zuteilung der Premiumfahrzeuge in die FNPs robust ist.

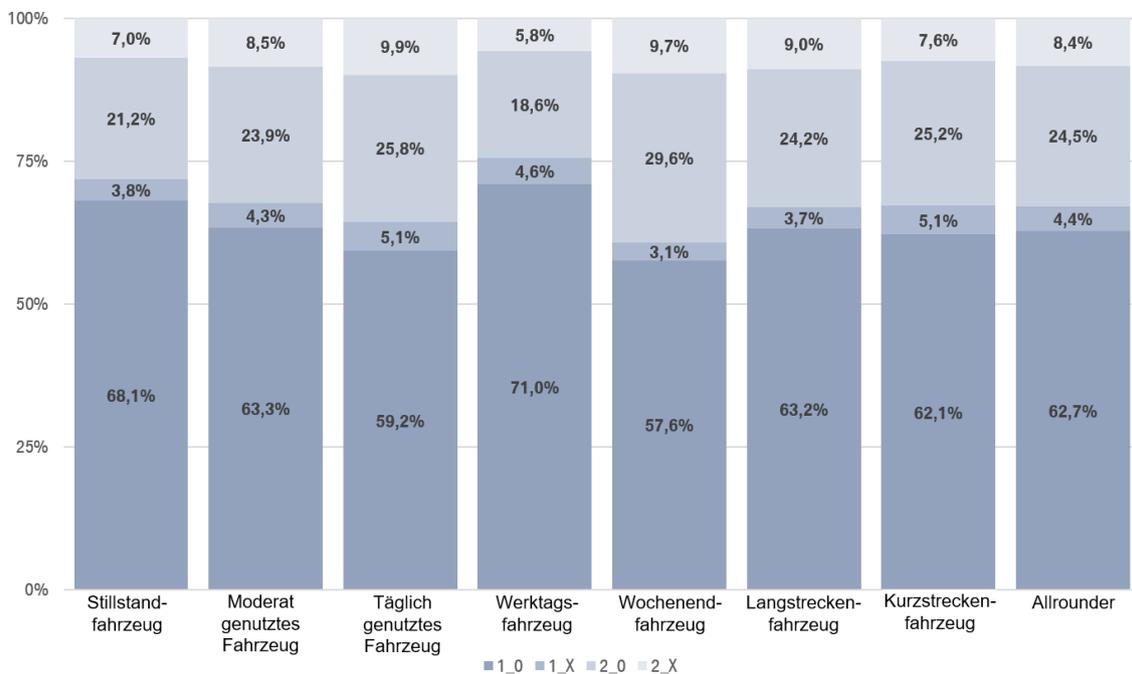


Abbildung 3.6: Premiumfahrzeuge: Anzahl Passagiere nach Fahrzeugnutzungsprofil

Beobachtung der Fahrzeugnutzung über ein Jahr

Mit dem Ziel, die zeitlichen und räumlichen Unterschiede in der Fahrzeugnutzung besser verstehen zu können, wird im Folgenden die Fahrzeugnutzung über ein Jahr (2018) für Frankreich, Italien, Spanien, Großbritannien sowie Österreich betrachtet. Die Bestimmung von FNPs stellt immer eine Momentaufnahme dar, welche von der Stichprobe und dem Beobachtungszeitraum abhängt. Durch die Längsschnittanalyse kann zum einen die saisonale Instabilität und zum anderen die Besonderheiten in der Fahrzeugnutzung über verschiedene Länder aufgezeigt werden. Die verwendeten Sensordaten sind von der Struktur her mit der Beschreibung aus Abschnitt 3.1.3 identisch. Insgesamt wurden 56 824 Fahrzeuge mit 43 279 403 Fahrten in den genannten Ländern beobachtet. Für die Aufnahme der Fahrzeuge in den Datensatz gelten dieselben Kriterien wie in Abschnitt 3.1.3 definiert. Für die Zuteilung der Premiumfahrzeuge zu FNPs wurde dasselbe Modell wie in Abschnitt 3.1.4 beschrieben verwendet. Aus diesem Grund wird auf die nähere Beschreibung der Daten und des Modells verzichtet.

Um Unterschiede über die vier Jahreszeiten zu identifizieren, wird die Fahrzeugnutzung der Premiumfahrzeuge für alle betrachteten Länder in verschiedene Zeiträume aufgeteilt: Frühjahr (F) (01.03.–31.05.), Sommer (S) (01.06.–31.08.), Herbst (H) (01.09.–30.11.) und Winter (W) (01.12.–28.02.). Durch die Erweiterung des Beobachtungszeitraums können nun die Veränderungen über die vier Jahreszeiten beschrieben werden. Tabelle 3.6 stellt die relativen Größen der in die FNPs eingeteilten Premiumfahrzeuge für die verschiedenen zeitlichen Dimensionen dar. Die Anteile der Stillstandfahrzeuge (F: 7,9 %; S: 7,8 %; H: 7,1 %; W: 4,9 %) nehmen über den Beobachtungszeitraum ab. Das ist durchaus nachvollziehbar. Zu Jahreszeiten mit wärmeren Temperaturen können andere Verkehrsmittel (z. B. Fahrrad) für bestimmte Wegezwecke verwendet werden, während Individuen bei kälteren Temperaturen lieber auf den Pkw zurückgreifen. Dies bestätigt sich auch bei den Anteilen der täglich genutzten Fahrzeuge (F: 10,7 %; S: 12,2 %; H: 10,6 %; W: 12,5 %) und Werktagsfahrzeuge (F: 2,7 %; S: 2,5 %; H: 2,8 %; W: 3,1 %). Der Anteil bei Wochenendfahrzeugen (F: 19,1 %; S: 19,7 %; H: 19,6 %; W: 15,1 %) ist im Winter am geringsten. Wochenendfahrzeuge könnten, wie bereits erwähnt, stärker für Freizeitaktivitäten (z. B. Ausflug in die Berge oder zum See) eingesetzt werden. Das Spektrum von Freizeitaktivitäten ist im Winter geringer als zu anderen Jahreszeiten, was sich am geringen Anteil von Wochenendfahrzeugen im Winter widerspiegelt. Der Anteil von Langstreckenfahrzeugen (F: 15,4 %; S: 15,3 %; H: 15,7 %; W: 15,6 %) und Kurzstreckenfahrzeugen (F: 4,9 %; S: 4,8 %; H: 4,9 %; W: 4,9 %) bleibt über die vier Jahreszeiten nahezu konstant. Möglicherweise werden Langstrecken- und Kurzstreckenfahrzeuge intensiver für Pflichtfahrten eingesetzt, wofür, unabhängig von der Jahreszeit, auf den Pkw zurückgegriffen werden muss.

Tabelle 3.6: Premiumfahrzeuge: Fahrzeugnutzungsprofile über die vier Jahreszeiten

Fahrzeugnutzungsprofil	Frühjahr	Sommer	Herbst	Winter
Stillstandfahrzeug	7,9 %	7,8 %	7,1 %	4,9 %
Moderat genutztes Fahrzeug	26,6 %	26,2 %	28,2 %	32,7 %
Täglich genutztes Fahrzeug	10,7 %	11,2 %	10,6 %	12,5 %
Werktagsfahrzeug	2,7 %	2,5 %	2,8 %	3,1 %
Wochenendfahrzeug	19,1 %	19,7 %	19,6 %	15,1 %
Langstreckenfahrzeug	15,4 %	15,3 %	15,7 %	15,6 %
Kurzstreckenfahrzeug	4,9 %	4,8 %	4,9 %	4,9 %
Allrounder	12,7 %	12,5 %	11,1 %	11,2 %
Σ	56 824	56 824	56 824	56 824

Tabelle 3.7 stellt die relativen Größen der FNP's für verschiedene Länder (Frankreich (FR), Italien (IT), Spanien (ES), Großbritannien (UK), Österreich (AT)) dar. Eine zusätzliche Information der Sensordaten ist das jeweilige Orderland des Fahrzeugs. Das Orderland entspricht dem Standort des Händlers. Unter der Annahme, dass sich der Händler im selben Land wie der Kunde befindet, kann das entsprechende Land bestimmt werden.

Bei der länderspezifischen Betrachtung der FNP's sind Unterschiede erkennbar. Der Anteil der Stillstandfahrzeuge ist in Großbritannien (7 %) am niedrigsten. Dies könnte vor allem daran liegen, dass Großbritannien unter den beobachteten Ländern die geringste Pkw-Dichte besitzt. Somit gibt es im Vergleich weniger Haushalte mit mehreren Autos, was dazu führt, dass der Anteil der Stillstandfahrzeuge dementsprechend klein ist. Das könnte auch die Erklärung dafür sein, dass der Anteil der täglich genutzten Fahrzeuge in Großbritannien am höchsten ist. Im Vergleich zu den anderen Ländern ist der Anteil von Wochenendfahrzeugen in Frankreich mit Abstand am höchsten. Langstreckenfahrzeuge haben in Italien die größte Clustergröße mit 26 %. Premiumfahrzeuge werden in Italien intensiv im Fernverkehr eingesetzt, wodurch sich der Premiumfaktor stark über längere Tagesdistanzen definieren könnte.

Tabelle 3.7: Premiumfahrzeuge: Fahrzeugnutzungsprofile über verschiedene Länder

Fahrzeugnutzungsprofil	FR	IT	ES	UK	AT
Stillstandfahrzeug	8 %	6 %	7 %	5 %	7 %
Moderat genutztes Fahrzeug	25 %	25 %	28 %	29 %	29 %
Täglich genutztes Fahrzeug	8 %	12 %	11 %	12 %	10 %
Werktagsfahrzeug	1 %	1 %	2 %	1 %	2 %
Wochenendfahrzeug	28 %	20 %	21 %	18 %	17 %
Langstreckenfahrzeug	15 %	26 %	15 %	15 %	20 %
Kurzstreckenfahrzeug	4 %	3 %	5 %	6 %	3 %
Allrounder	10 %	7 %	11 %	14 %	13 %
Σ	9643	5 870	3399	34 782	3130

3.1.6 Diskussion

Um die Fahrzeugnutzung von Premium- und gewöhnlichen Fahrzeugen zu vergleichen, wurde eine Technik des maschinellen Lernens, quadratische SVM, verwendet. Für den Vergleich wurden gewöhnliche Fahrzeuge aus Umfragedaten und Premiumfahrzeuge aus Sensordaten in existierende FNP eingeteilt. Auf dieser Basis war es möglich, die relativen Größen der FNPs miteinander zu vergleichen und premiumfahrzeugspezifische Nutzungscharakteristika, wie eine stärkere Nutzung im Fernverkehr, zu identifizieren. Über Informationen zur Sitzplatzbelegung bei Premiumfahrzeugen konnten die Abhängigkeiten zwischen der Anzahl der Passagiere und den FNPs beschrieben werden. Durch die Erweiterung des Beobachtungszeitraums wurden saisonale und länderspezifische Unterschiede untersucht.

Diese Erkenntnisse sind hilfreich, um Anforderungen für alternative Antriebstechnologien und Mobilitätsdienstleistungen im Premiumsektor abzuleiten:

- Premiumfahrzeuge werden intensiver (geringer Anteil von Stillstandfahrzeugen) und für längere Distanzen (großer Anteil von Langstreckenfahrzeugen) genutzt. Deshalb könnte sich der Premiumfaktor bei einem batterieelektrischen Fahrzeug (BEV) stärker über die Reichweite definieren. Beispielsweise hat Tesla mit dem Model S und Model X mit einer Reichweite von ca. 450 km im Winter und ca. 350 km im Sommer bereits BEVs am Markt (Stand 2019), welche überdurchschnittlich im Fernverkehr eingesetzt werden (Figenbaum & Nordbakke, 2019; Nicholas, Tal & Turrentine, 2017). Allerdings wird die Anzahl der BEVs am Markt mit entsprechender Reichweite immer größer (z. B. Hyundai Kona Electric (2018): Reichweite 449 km⁴; VW ID.3 (2020): Reichweite 550 km; Polestar 2 (2020): Reichweite 500 km), wodurch sich der Premiumfaktor für BEVs zukünftig verändert wird.

⁴Die Elektr Reichweite wurde unter WLTP (Worldwide harmonized Light vehicles Test Procedure) berechnet.

- Anbieter von Premium-Carsharing sollten Fahrzeuge mit einem erweiterten Radius anbieten, um die größeren Tagesdistanzen bei Premiumfahrzeugen zu berücksichtigen.
- Durch die höhere Zahlungsbereitschaft könnte auch der Wille zu einer breiten Nutzung, unabhängig von spezifischen Fahrtzwecken, vorhanden sein. Möglicherweise ist der Premiumfaktor die universelle Einsetzbarkeit eines Fahrzeugs für verschiedenste Zwecke (Fernverkehr, Stauraum, Komfort). Diese Annahme ist auch konsistent mit bestehender Literatur. Beispielsweise haben von Behren et al. (2020) gezeigt, dass die Nutzung von Premiumfahrzeugen in einem Haushalt intensiver ist als die von gewöhnlichen Fahrzeugen.
- Die Analyse der Sitzplatzbelegung hat ergeben, dass die Anzahl der Passagiere stärker vom Fahrtzweck als von der zurückgelegten Distanz bestimmt wird. Im Kontext der Klimadiskussionen könnten Carsharinganbieter zusammen mit der Regierung Anreize schaffen, um die Belegungsrate in einem Auto zu erhöhen. Beispielsweise könnte der Preis eines Mietwagens von der Anzahl der Passagiere abhängen. Ein weiterer Anreiz wäre die Erweiterung von HOV-Lanes, deren positiver Effekt bereits in verschiedenen Studien belegt wurde (Delhomme & Gheorghiu, 2016; Guensler et al., 2019).

Die dargestellten Ergebnisse und ihre Ableitungen stehen im Kontrast zu verschiedenen Limitationen, welche im Kontext der Analyse von FNPs auf der Basis von Sensordaten berücksichtigt werden müssen:

- Aus Gründen des Datenschutzes können bestimmte Informationen nur bei einer expliziten Zustimmung des Fahrzeugnutzers abgerufen werden. Beispielsweise ist es nicht möglich, zwischen privat und geschäftlich genutzten Pkws zu unterscheiden. Darüber hinaus wird durch einen Algorithmus die Fahrgestellnummer anonymisiert, was dazu führt, dass keine soziodemografischen Informationen verfügbar sind.
- GPS-Informationen können nur mit ausdrücklicher Zustimmung des Kunden analysiert werden. Die hier verwendeten Sensordaten erlauben deshalb nur eine Identifikation des Orderlandes über den Händler. Mit einem detaillierteren Blick auf den Aktionsraum eines Fahrzeugs (z. B. Unterscheidung von ländlichen und urbanen Gebieten) könnten die Unterschiede in der Fahrzeugnutzung zwischen den Ländern noch besser herausgearbeitet werden.
- Aufgrund des Datenschutzes sind auch keine Informationen hinsichtlich Halterwechseln und der Anzahl der Fahrzeugnutzer verfügbar. Oftmals werden Neuwagen als Firmenfahrzeug genutzt und danach im Privatautomobilmarkt als Jahreswagen angeboten. Das führt zu einem Wechsel der Besitzverhältnisse und damit auch zu einer Veränderung der Fahrzeugnutzung. Mit Hilfe dieser Informationen könnte die Nutzung in Abhängigkeit von den Lebensphasen eines Fahrzeugs untersucht werden.

- Die Aufzeichnung von Sensordaten zur Fahrzeugnutzung kann über verschiedene Übertragungskanäle zu unterschiedlichen Zwecken erfolgen. In den meisten Fällen ist der Übertragungskanal an bestimmte Ausstattungsmerkmale gebunden. Bei den verwendeten Sensordaten muss das Fahrzeug mit einem hochwertigen Navigationssystem ausgestattet sein. Sensordaten von Fahrzeugen ohne diese Ausstattung sind nicht beobachtbar. Als Folge könnte die beobachtete Fahrzeugnutzung verzerrt sein, wenn Fahrzeuge mit einer besseren Ausstattung auch generell intensiver genutzt werden.

Aus den Erkenntnissen und den Limitationen der Sensordaten ergeben sich zudem weitere Ansätze für zukünftige Forschungsvorhaben:

- Eine hohe Sitzplatzbelegungsrate kann das Verkehrsaufkommen und damit auch die Stauzeit reduzieren. Eine Analyse der Effekte von HOV-Lanes auf die Sitzplatzbelegungsrate würde neue Einblicke generieren, inwieweit Individuen bereit sind, ihr Mobilitätsverhalten zu verändern, um dadurch Fahrtzeit einzusparen.
- Sensordaten enthalten im Kontrast zu Umfragedaten keine zusätzlichen Fahrtinformationen (z. B. Fahrtzweck). Diese Informationen sind aber relevant, um die tatsächlichen Gründe für den Einsatz eines Fahrzeugs zu kennen. Zukünftige Forschungsarbeiten könnten sich damit beschäftigen, Sensordaten um zusätzliche Fahrtinformationen zu ergänzen. Beispielsweise wäre es möglich, über Nutzungsmuster Fahrtzwecke (z. B. Fahrt nach Hause, Fahrt zur Arbeit) zu approximieren.

Die in diesem Unterkapitel verwendete Methode zur Zuteilung der Premiumfahrzeuge in die FNP von Eisenmann und Buehler (2018) bietet Raum für weiterführende Diskussionen:

- Jedes Fahrzeug kommt im Laufe seiner Nutzung in jedes FNP. Langstreckenfahrzeuge werden auch für kurze Strecken genutzt und andersherum. Darüber hinaus ist die Definition der Profile nicht immer eindeutig. Beispielsweise könnte ein Langstreckenfahrzeug auch primär am Wochenende genutzt werden. Aus diesem Grund ist eine eindeutige Zuteilung der Fahrzeuge für bestimmte FNPs oftmals nicht sinnvoll. Deshalb werden in Unterkapitel 3.2 BEVs aus den gleichen Sensordaten probabilistisch zugeordnet. Zum einen findet damit eine Weiterentwicklung der angewandten Methode statt und zum anderen kann inhaltlich die Marktdurchdringung von BEVs näher untersucht werden.
- Die FNPs von Eisenmann und Buehler (2018) basieren auf der Fahrzeugnutzung von gewöhnlichen Fahrzeugen. Durch die unüberwachte, nicht regelbasierte Identifikation der FNPs hängt die Interpretation der Profile stark von der Verteilung der Datenpunkte ab. Werden also andere Daten zur Beschreibung der Fahrzeugnutzung verwendet, ergeben sich daraus auch ganz andere FNPs. Möglicherweise spielen für die Beschreibung der

Nutzung von Premiumfahrzeugen auch ganz andere Variablen eine wichtige Rolle. Aus diesem Grund werden in Kapitel 4 neue Variablen auf der Basis der drei Dimensionen Zeit, Distanz und Fahrtzweck gebildet, um neue FNPs zu entwickeln.

Alles in allem konnte die eingangs genannte Fragestellung, inwiefern sich die Fahrzeugnutzung von Premiumfahrzeugen von gewöhnlichen Fahrzeugen unterscheidet, mit Hilfe der FNPs von Eisenmann und Buehler (2018) und des daraus resultierenden Vergleichs der relativen Größen beantwortet werden. Dadurch wurde eine stärkere Relevanz im Fernverkehr für Premiumfahrzeuge identifiziert. Darüber hinaus werden Premiumfahrzeuge nicht nur für spezifische Fahrtzwecke, sondern sehr generell eingesetzt. Dies wird auch durch die ähnlichen Clustergrößen in Deutschland und Kalifornien bestätigt. Des Weiteren wurden Informationen zur Sitzplatzbelegung genutzt, um Unterschiede entlang der FNPs zu beobachten. Hierbei stellte sich heraus, dass die Anzahl der Passagiere stärker vom Zweck als von der Distanz abhängt.

Dieses Unterkapitel konnte ein besseres Verständnis der Marktbedürfnisse im Premiumsegment für neue Mobilitätsdienstleistungen (z. B. Carsharing) und alternative Antriebstechnologien (z. B. Elektromobilität) generieren.

3.2 Elektrofaktor

Dieses Unterkapitel basiert großteils auf der Veröffentlichung „Electric factor – A Comparison of Car Usage Profiles of Electric and Conventional Vehicles by a Probabilistic Approach“ (Niklas, von Behren, Chlond & Vortisch, 2020).

3.2.1 Motivation

Das steigende Bewusstsein für umweltverträgliche Mobilität auf der Nachfrageseite und striktere CO₂-Regulierungen auf der Angebotsseite machen BEVs höchst relevant für die Reduzierung von CO₂-Emissionen im Transportsektor. Dies führt zu Veränderungen in der Zusammensetzung von Antriebstechnologien im Automobilmarkt. Beispielsweise hat sich in Norwegen der Anteil zugelassener BEVs am Gesamtmarkt in den letzten Jahren vervielfacht (Abbildung 3.7). In Norwegen liegt der Marktanteil bei über 45 % (BMW, 2020a).

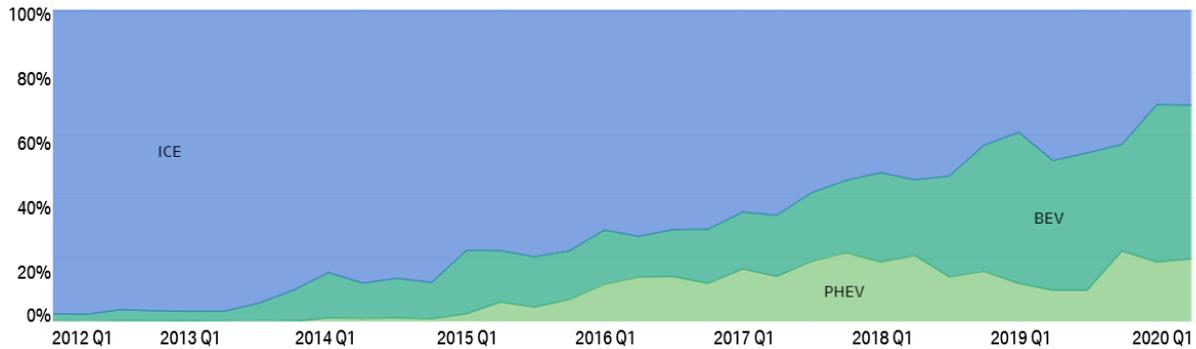


Abbildung 3.7: Anteil der Zulassungen nach Antrieben in Norwegen

Grund hierfür sind die technologischen Entwicklungen, welche zu einer Steigerung der verfügbaren Elektreichweiten geführt haben. Gegenwärtige (2020) BEVs am Markt erreichen bereits die erforderlichen Reichweiten (z.B. Polestar 2 mit einer Reichweite von bis zu 500 km, Mercedes EQC mit einer Reichweite von 412 km oder Audi e-tron mit einer Reichweite von 411 km⁵). Zum anderen führen Verbesserungen hinsichtlich des Ladeverhaltens (Ladeinfrastruktur und Ladegeschwindigkeit) zu einer sich angleichenden Nutzbarkeit von BEVs im Vergleich zu gewöhnlichen Fahrzeugen mit Verbrennungsmotoren.

Wie bereits in Abschnitt 3.1.2 erwähnt, ist es möglich, die Fahrzeugnutzung über traditionelle HTS- sowie Sensordaten zu untersuchen. Die Berücksichtigung der BEV-Nutzung ist bei HTS noch nicht so stark verbreitet, weshalb diese bei traditionellen Erhebungen gegenwärtig nur geringfügig abgebildet sind. Des Weiteren ist ein Vergleich der Fahrzeugnutzung zwischen verschiedenen Ländern nicht direkt abbildbar, da sich die Erhebungsmethoden von Land zu Land unterscheiden. Zudem müssten bei einer Eignungsprüfung von BEVs über HTS alle Nutzer eines Fahrzeugs über einen längeren Zeitraum beobachtet werden, was zusätzliche Annahmen erfordern würde.

Auch hier liefern die CUMILE-Daten einen zentralen Vorteil, da sie die Nutzung des Fahrzeugs und nicht das Verhalten von Individuen abbilden.

Um die Eignung von BEVs zu prüfen, werden die bereits in Abschnitt 3.1.3 beschriebenen CUMILE-Daten sowie Sensordaten, welche die spezifische Nutzung von BEVs betrachten, für Deutschland und Kalifornien verwendet.

Mit Hilfe der zugrundeliegenden Daten sollen folgende Forschungsfragen beantwortet werden:

1. In welchem Ausmaß werden relevante Nutzungsmuster von gewöhnlichen Fahrzeugen bereits von BEV-Nutzern als aufstrebende Antriebstechnologie praktiziert?
2. Unterscheidet sich die Nutzung von BEVs in Deutschland und Kalifornien?

⁵Die elektrische Reichweite wurde unter WLTP berechnet.

Ziel des Unterkapitels

In der Vergangenheit wurde in der Literatur viel über mögliche elektrische Reichweiten zur Befriedigung der alltäglichen Pkw-Mobilität diskutiert. Kriterien zur Bestimmung geeigneter elektrischer Reichweiten basieren primär auf objektiven Reichweiten, welche aus der aktuellen Nutzung von Fahrzeugen mit Verbrennungsmotor (Abschnitt 3.2.2 – Literaturübersicht) abgeleitet sind. Diese Studien können allerdings nicht beantworten, inwiefern BEVs am Markt bereits wie gewöhnliche Fahrzeuge genutzt werden. Nur wenn Individuen BEVs wie gewöhnliche Fahrzeuge nutzen, werden diese auch in der langen Frist gewöhnliche Fahrzeuge ersetzen können. Im Umkehrschluss können BEVs sich nicht flächendeckend am Markt durchsetzen, da die Nachfrage lediglich aufgrund von monetären Anreizen (z. B. geringe Parkgebühren für BEVs) oder Regulierungen (z. B. Fahrverbote für Fahrzeuge mit Verbrennungsmotor (z. B. Paris und London) oder Lizenz-Obergrenzen (z. B. in Peking)) existiert.

Deutschland und Kalifornien sind gut entwickelte Wirtschaftsräume mit hoher Lebensqualität und vergleichbaren geografischen Größen (Eisenmann, 2019; Eisenmann & Buehler, 2018; Kühne, Mitra & Saphores, 2018). Beide geografischen Gebiete haben eine hohe Motorisierungsrate und ein gut ausgebautes Autobahnnetz (Buehler, 2011). Deutschland und Kalifornien weisen allerdings auch Unterschiede, wie beispielsweise die unterschiedliche Qualität des ÖPNVs und die unterschiedlichen Kosten für die Nutzung verschiedener Verkehrsmittel (z. B. Benzinpreis) (IEA, 2019) auf. Beide Länder sind gut zum Vergleich geeignet, da sie sowohl ähnlich entwickelt sind (z. B. Lebensstandard und Motorisierungsrate), aber trotzdem spezielle Merkmale mit sich bringen (z. B. in Bezug auf Verkehrspolitik und räumliche Entwicklung) (Eisenmann & Buehler, 2018; Haustein & Nielsen, 2016; Hildebrand, 2003).

Im nächsten Schritt wird eine Literaturübersicht zur Eignungsprüfung von BEVs gewährt. Da die verwendeten CUMILE-Daten sowie die Erfassung und Speicherung der Sensordaten bereits in Abschnitt 3.1.3 dargestellt wurden, wird auf die Datenbeschreibung nicht eingegangen. Insgesamt wurden für diesen Ansatz 1073 BEVs mit 66 672 Fahrten im Mai 2018 für Deutschland und Kalifornien über Sensordaten erfasst. Im darauffolgenden Abschnitt wird die Funktionsweise der angewandten Methode (Softmax-Regression (SR)) näher beschrieben. Danach werden die berechneten Ergebnisse diskutiert. Eine Zusammenfassung der Ergebnisse, ein Ausblick sowie Möglichkeiten für weitere Forschungsarbeiten werden im abschließenden Abschnitt präsentiert.

3.2.2 Literaturübersicht

Es gibt eine Vielzahl von Studien, welche sich damit beschäftigen, inwieweit sich die Nutzung von Fahrzeugen mit bestimmten Elektreichweiten deckt. Manche umfragebasierte Studien messen die Eignung von BEVs für vordefinierte Eigenschaften (Reichweite, Ladezeit) auf

der Basis von Schwellenwerten in Bezug auf Fahrt- oder Tagesdistanzen. Andere Studien zielen darauf ab zu untersuchen, an wie vielen Tagen bestimmte elektrische Reichweiten nicht die Anforderungen der zurückgelegten Distanzen erfüllen. Weiss, Chlond, Heilig und Vortisch (2014) definieren, basierend auf dem MOP, drei Stadien: (1) Eignung von BEVs ohne Anpassungen, (2) Eignung von BEVs mit leichten Anpassungen, (3) keine Eignung von BEVs. Ein Fahrzeug ist nicht durch ein BEV ersetzbar, sobald das Fahrzeug eine Tagesfahrleistung von mehr als 100 km an mindestens fünf Tagen im Jahr überschreitet. Im gleichen Jahr wurde eine internationale Vergleichsstudie, basierend auf webbasierten Wegetagebüchern in Deutschland, Frankreich, Großbritannien, Italien, Polen und Spanien, durchgeführt. Eine beobachtete Tagesfahrleistung von 40 km in Großbritannien und 80 km in Polen ergibt, dass die meisten Fahrzeuge innerhalb gegebener Elektreichweiten liegen. Des Weiteren sind mehr als die Hälfte aller Wegekettens kürzer als 50 km. Aussagen darüber, ob das Fahrzeug im Haushalt als Erst- oder Zweitwagen genutzt wird, wurden nicht gemacht (Pasaoglu et al., 2014). Jakobsson et al. (2016) unterscheiden zwischen Erst- und Zweitwagen und untersuchen die technisch-ökologische Nachhaltigkeit von BEVs in Schweden (400 Fahrzeuge) und Deutschland (6300 Fahrzeuge), basierend auf dem Swedish car movement data (SCMD) und dem MOP. In dieser Studie wurden nur Fahrzeuge aus SCMD mit mindestens 30 Tagen glaubwürdigen GPS-Messungen berücksichtigt. Mit einer angenommenen elektrischen Reichweite von 230 km konnte die Hälfte der Fahrzeuge ihre Tagesdistanzen auch mit einem BEV zurücklegen. Bei Jakobsson et al. (2016) ist die Annahme, dass Individuen keinerlei Anpassungen ihres Nutzungsverhaltens machen und BEVs über Nacht geladen werden können. Basierend auf dem U.S. NHTS aus dem Jahr 2009 zeigt van Haaren (2011), dass 99 % der Fahrt- bzw. Tagesdistanzen unter 70 Meilen bzw. 250 Meilen liegen. Die Daten des U.S. NHTS wurden über die Erhebungsmethode CATI an einem Tag erhoben. Khan und Kockelman (2012) nehmen in ihrer Arbeit eine elektrische Reichweite von 120 Meilen an. Es zeigt sich, dass 95 % der Fahrzeuge aus einem Haushalt mit nur einem Fahrzeug an 95 % der beobachteten Tage ihre zurückgelegten Strecken auch mit einem BEV hätten bewältigen können. In dieser Studie wurde ein Datensatz verwendet, welcher Nutzungsdaten von 445 Fahrzeugen (aus 264 Haushalten) an durchschnittlich 341 Tagen in Seattle (Puget Sound Regional Council) beinhaltet. Im selben Jahr setzen auch Tamor und Milačić (2015) diesen Datensatz für ihre Untersuchungen ein. Der Fokus liegt hier auf der Eignung von BEVs in Haushalten mit mehreren Fahrzeugen. Im Ergebnis benötigen 90 % der Haushalte mit mehreren Fahrzeugen lediglich 60 Meilen elektrische Reichweite, um ihre Pkw-Mobilitätsbedürfnisse zu befriedigen. Eine weitere Studie zeigt auf, dass 75 % der beobachteten Fahrzeuge durch BEVs mit einer Reichweite von 200 Meilen ersetzt werden können, vorausgesetzt, es werden Anpassungen an sechs Tagen im Jahr gemacht (Pearre et al., 2011). Die Daten wurden durch die Installation

von sogenannten Datenloggern im Fahrzeug generiert. Der Datensatz aus dem Jahr 2000 beinhaltet 363 Fahrzeuge aus Atlanta mit einem Beobachtungszeitraum von einem bis drei Jahren (Wolf, Guensler, Washington & Frank, 2000).

Viele Studien, welche sich mit der Eignung von BEVs beschäftigen, sind bereits nach einigen Jahren durch den technologischen Fortschritt in der Batterieentwicklung überholt. Beispielsweise sind gegenwärtig elektrische Reichweiten weit über 200 km gängig, wodurch die Ergebnisse vieler Studien nicht mehr verwertbar sind. Zusammenfassend lässt sich sagen, dass es kein gemeinsames Verständnis gibt, ab welchen Kriterien BEVs die Anforderungen der Fahrzeugnutzer erfüllen und gewöhnliche Fahrzeuge mit Verbrennungsmotoren durch BEVs ersetzt werden können.

3.2.3 Methodik

In Abschnitt 3.1.4 wurde mit Hilfe des gelabelten Datensatzes für gewöhnliche Fahrzeuge $C^{(l)} = \{(\mathbf{x}_i^{(l)}, y_i^{(l)})\}_{i=1}^{3340}$ eine Funktion $f(\cdot)$ erlernt, welche die gewöhnlichen Fahrzeuge mit den Eigenschaften $\mathbf{x}_i^{(l)}$ in das entsprechende FNP $y_i^{(l)}$ einteilen kann. Mit Hilfe der Funktion $f(\cdot)$ konnte jedes Premiumfahrzeug j mit den Eigenschaften $\mathbf{x}_j^{(u)}$ einem FNP $y_j^{(u)} = f(\mathbf{x}_j^{(u)})$ eindeutig zugeordnet werden. Im Ergebnis wurde dadurch ein gelabelter Datensatz für Premiumfahrzeuge $P^{(u)} = \{(\mathbf{x}_j^{(u)}, y_j^{(u)})\}_{j=1}^{74568}$ generiert. In einem Fahrzeugleben gerät aber jedes Fahrzeug einmal in eines der acht FNPs. Beispielsweise kann ein Fahrzeug, welches dem Typus Langstreckenfahrzeug angehört, auch mal eine Kurzstrecke zurücklegen. Darüber hinaus ist die Zuteilung der FNPs nicht eindeutig, da die Nutzung eines Fahrzeugs die Kriterien für mehrere FNPs (z. B. Langstrecken- und Wochenendfahrzeug) erfüllen kann. Deshalb werden die vorliegenden BEVs nicht mehr eindeutig, sondern probabilistisch zugeteilt. Jedes BEV j^6 erhält nun auf der Basis der fünf Eigenschaften \mathbf{x}_j eine Wahrscheinlichkeit der Zugehörigkeit $\sum_{k=1}^8 P(\mathbf{x}_j | y_{j;k})$ für jedes FNP k , sodass im Ergebnis ein gelabelter Datensatz für Elektrofahrzeuge $BEV^{(u)} = \{(\mathbf{x}_j, \sum_{k=1}^8 P(\mathbf{x}_j | y_{j;k}))\}_{j=1}^{1073}$ generiert wird.

Aufgrund der unterschiedlichen Beobachtungszeiträume (gewöhnliche Fahrzeuge: 1 Jahr; BEVs: 1 Monat) werden die Daten, wie in Abschnitt 3.1.4, über eine z -Transformation standardisiert ($z_i^{(GF)} = \frac{x_i^{(GF)} - \mu}{\sqrt{\sigma^2}}$). Um die BEVs in die acht FNPs einzuteilen, wird eine Softmax-Regression (SR) verwendet. Die SR ist in der Lage, multiple Klassifikationsprobleme zu lösen, und stellt eine Generalisierung der bekannten logistischen Regression, welche binäre Klassifikationsprobleme löst, dar (Géron, 2019; Hastie et al., 2009; VanderPlas, 2016).

$$s_k(\mathbf{z}) = \mathbf{z}^\top \theta^{(k)} \quad (3.1)$$

⁶Zur Vereinfachung werden die BEVs, genauso wie Premiumfahrzeuge, mit j deklariert.

Für jedes Fahrzeug i und für jedes FNP k berechnet die SR einen Parameter $s_k(\mathbf{z})$. Dieser Parameter wird in einer $n \times k$ -Matrix \mathbf{s} ($\mathbf{s} = \sum_{k=1}^K s_k(\mathbf{z})$) hinterlegt. \mathbf{z} ist die standardisierte $n \times m$ Matrix über alle Fahrzeuge n und die fünf Inputvariablen m . $\theta^{(k)}$ ist der Parametervektor des k ten FNP. K entspricht den acht FNPs. Alle Vektoren finden sich in der $k \times m$ -Matrix Θ wieder.

$$\Theta = \begin{array}{c} \left(\begin{array}{cccccc} \theta_{11} & \theta_{12} & \cdots & \cdots & \theta_{1m} \\ \theta_{21} & \ddots & \vdots & \cdots & \theta_{2m} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{k1} & \theta_{k2} & \cdots & \cdots & \theta_{km} \end{array} \right) \end{array} \begin{array}{l} \text{Stillstandfahrzeug} \\ \text{Moderat genutztes Fahrzeug} \\ \text{Täglich genutztes Fahrzeug} \\ \text{Werktagsfahrzeug} \\ \text{Wochenendfahrzeug} \\ \text{Langstreckenfahrzeug} \\ \text{Kurzstreckenfahrzeug} \\ \text{Allrounder} \end{array} \quad (3.2)$$

Nachdem alle Parameter für jedes FNP berechnet worden sind, kann die Wahrscheinlichkeit \hat{p}_k der Zugehörigkeit zu einem FNP k geschätzt werden. Dies erfolgt durch Integration der Parameter $s_k(\mathbf{z})$ in die Softmax-Funktion:

$$\hat{p}_k = h_{\theta}(s_k(\mathbf{z})) = \frac{\exp(s_k(\mathbf{z}))}{\sum_{k=1}^K \exp(s_k(\mathbf{z}))} \quad (3.3)$$

Die Softmax-Funktion normalisiert jeden Parameter zu einem Wert zwischen 0 und 1. Große Parameter ($+\infty$) konvergieren gegen 1 und niedrige Parameter ($-\infty$) konvergieren gegen 0. Anders ausgedrückt: Die Werte von $s_k(\mathbf{z})$ werden zu Wahrscheinlichkeiten transformiert. Jedes Fahrzeug i erhält somit eine Wahrscheinlichkeit $\sum_{k=1}^8 P(\mathbf{x}_i | y_{i,k})$ der Zugehörigkeit für jedes FNP k . Die Summe aller Wahrscheinlichkeiten der acht FNPs eines Fahrzeugs ergibt 1 ($\sum_{k=1}^8 \hat{p}_k = 1$). $h_{\theta}(s_k(\mathbf{z}))$ ist die geschätzte Wahrscheinlichkeit, dass ein Fahrzeug zu einem FNP, unter der Annahme der Parameter $s_k(\mathbf{z})$, gehört. Dieser Zusammenhang ist ein Vorteil der SR gegenüber anderen überwachten Lernverfahren (z. B. SVM), welche nur eindeutige Zuteilungen bieten. Die SR liefert eine Wahrscheinlichkeit der Zuteilung für jedes FNP. Speziell bei der Verwendung verschiedener Datenquelle ist eine Wahrscheinlichkeit hilfreich, um die Qualität der Zuteilung für jedes BEV zu bestimmen. Zusätzlich ist es möglich, Fahrzeuge zu identifizieren, welche nicht in eines der acht FNPs passen. Um eine eindeutige Zuteilung zu generieren, wird das FNP mit der höchsten Zuteilungswahrscheinlichkeit ausgewählt:

$$\hat{y} = \underset{k}{\operatorname{argmax}} \hat{p}_k = \underset{k}{\operatorname{argmax}} h_{\theta}(s_k(\mathbf{z})), \quad (3.4)$$

wobei \hat{y} den Vektor aus den jeweiligen FNPs mit den höchsten Wahrscheinlichkeiten darstellt. Zur Optimierung des Modells werden die Parameter der Matrix Θ so angepasst, dass die Kreuzentropiekostenfunktion $J(\Theta)$ minimal ist:

$$J(\Theta) = -\frac{1}{m} \sum_{i=1}^m \sum_{k=1}^K y_k^{(i)} \log(\hat{p}_k^{(i)}) \quad (3.5)$$

Da $J(\Theta)$ keine bekannte geschlossene Lösung aufweist, aber eine konvexe Funktion⁷ ist, kann das globale Optimum über den Optimierungsalgorithmus stochastisches durchschnittliches Gradientenabstiegsverfahren (SAG) iterativ bestimmt werden. Das SAG ist eine leichte Abwandlung des stochastischen Gradientenabstiegsverfahrens (SGD) und berechnet den Durchschnitt der Parametervektoren über die Zeit. Um diesen zu berechnen, wird die 1. Ableitung der Kreuzentropiekostenfunktion $J(\Theta)$ nach θ_k benötigt:

$$\nabla_{\theta_k} J(\Theta) = \frac{1}{m} \sum_{i=1}^m (\hat{y}_k^{(GF)(i)} - y_k^{(GF)(i)}) \mathbf{z}^{(i)} \quad (3.6)$$

Um das Modell zu validieren und die Generalisierung zu gewährleisten, wurde der gelabelte Datensatz in einen Trainings- (70 %) und Testdatensatz (30 %) aufgeteilt. Insgesamt stimmen 91,21 % der vorhergesagten FNPs \hat{y}^{GF} mit der tatsächlichen Zuteilung y^{GF} überein. Die geschätzten Parameter für die Zuteilung sind in der $k \times m$ -Matrix Θ in Tabelle 3.8 aufgeführt:

Tabelle 3.8: Entscheidungsfunktionen der Fahrzeugnutzungsprofile

Fahrzeugnutzungs- profil	Θ				
	Tage0	Tage10	Tage100	AntWerktag	VKT
Stillstandfahrzeug	6,461	-4,771	1,320	5,123	-0,045
Moderat genutztes Fahrzeug	-1,245	-4,077	-1,999	-1,560	0,412
Täglich genutztes Fahrzeug	-6,437	2,170	-1,093	-4,237	0,643
Werktagsfahrzeug	1,192	-0,338	0,860	6,777	-0,579
Wochenendfahrzeug	4,728	-2,340	-0,905	-4,175	0,143
Langstreckenfahrzeug	-1,639	-4,185	3,075	-1,307	-0,044
Kurzstreckenfahrzeug	-3,752	10,078	-1,049	-1,241	-0,573
Allrounder	0,692	3,463	-0,210	0,620	0,044

Jede Zeile beschreibt die Entscheidungsfunktion für das FNP k und jede Spalte zeigt die Parameter für die fünf Inputvariablen auf. Die Parameter Θ beinhalten die gebündelte Schätzung für beide Länder, um die gleichen Kriterien für die Zuteilung der BEVs zu den FNPs, unabhängig vom Land, zu garantieren. Anhand der Matrix Θ wird sichtbar, wie sich die jeweiligen FNPs in den Inputvariablen widerspiegeln. Tage, an denen das Fahrzeug

⁷Eine Funktion ist streng konvex, wenn gilt: $f(tx + (1-t)y) < tf(x) + (1-t)f(y)$.

nicht bewegt wurde (Tage0), hat den größten positiven Einfluss (6,461) auf die Zuteilung zu einem Stillstandfahrzeug und den größten negativen Einfluss ($-6,437$) auf die Zuteilung zu einem täglich genutzten Fahrzeug. Je weniger ein Fahrzeug bewegt wird, desto geringer ist die Wahrscheinlichkeit für die Zuteilung zu einem täglich genutzten Fahrzeug. Eine Tagesfahrleistung unter 10 Meilen (Tage10) hat einen sehr starken Einfluss (10,078) auf die Zuteilung zu einem Kurzstreckenfahrzeug. Auch die zusätzliche Zuteilungsvariable VKT hat einen Einfluss auf die Zuteilung. Auch wenn dieser Einfluss nicht sehr hoch ist, werden die VKT vor allem bei der Zuteilung von BEVs zu FNPs relevant. Besonders die zurückgelegten Distanzen eines BEV sind im Kontext der Diskussionen zu elektrischen Reichweiten wichtig. Um die Wahrscheinlichkeiten $\hat{p}_k^{(BEV)}$ der FNPs für jedes BEV zu berechnen, werden die gleichen fünf Inputvariablen $x^{(BEV)}$ verwendet und mit Hilfe der z -Transformation \mathbf{z}^{BEV} standardisiert. Danach werden die Parameter $s_k(\mathbf{z})^{(BEV)}$ auf der Basis der geschätzten Matrix Θ bestimmt und in die Softmax-Funktion transferiert, sodass die Zuteilungswahrscheinlichkeit $\hat{p}_k^{(BEV)}$ für jedes FNP k und jedes BEV j bestimmt werden kann. Basierend auf den Wahrscheinlichkeiten kann dann die Zuteilung $\hat{y}^{(BEV)}$ der BEVs zu einem FNP erfolgen. Das Verfahren wurde unter Verwendung von Python und der zugehörigen Paketen (pandas, numpy, scikit-learn, matplotlib) realisiert.

3.2.4 Ergebnisse

Um die Zuverlässigkeit der Einordnung zu untersuchen, werden im ersten Schritt die Zuteilungswahrscheinlichkeiten der BEVs $\hat{p}_k^{(BEV)}$ und deren Varianzen untersucht. Basierend auf der Zuverlässigkeit werden nur noch BEVs mit einer Zuteilungswahrscheinlichkeit größer 75 % (BEV75) betrachtet. Danach findet ein Vergleich der prognostizierten Zuteilungen $\hat{y}^{(BEV75)}$ der BEVs und der gewöhnlichen Fahrzeuge $y^{(GF)}$ zu einem bestimmten FNP für Deutschland und Kalifornien statt. Durch die Softmax-Funktion ist es möglich die durchschnittliche Wahrscheinlichkeit (med) und deren Varianz (var) für die Zuteilung zu den acht FNPs (Abbildung 3.8) zu betrachten. Dies bietet die Möglichkeit zu überprüfen, wie gut die FNPs mit der Nutzung der BEVs übereinstimmen. Je höher die Wahrscheinlichkeit, desto besser passt die Nutzung der BEVs zu den jeweiligen FNPs. Des Weiteren kann über die Varianz die Unterscheidbarkeit zwischen den BEVs und den FNPs untersucht werden. Je geringer die Varianz der FNPs, desto eindeutiger sind die FNPs.

Stillstandfahrzeuge (med: 97 %; var: 2,1 %), Kurzstreckenfahrzeuge (med: 97 %; var: 2,0 %) sowie Wochenendfahrzeuge (med: 95 %; var: 2,4 %) haben eine hohe durchschnittliche Wahrscheinlichkeit und eine niedrige Varianz. Das bedeutet, dass die Zuteilungen der sensorbasierten BEVs zu den umfragebasierten Nutzungsprofilen gut passen. Moderat genutzte Fahrzeuge (med: 89 %; var: 2,9 %) und Langstreckenfahrzeuge (med: 93 %; var: 4,2 %) haben zwar einen hohen Median, aber die Varianz, speziell für Langstreckenfahrzeuge, ist doppelt so groß wie

die für Stillstandfahrzeuge. Anders ausgedrückt: Es gibt zwar BEVs, welche klar dem Langstreckenfahrzeug zugeordnet werden können (hohe Zuteilungswahrscheinlichkeit), aber es gibt auch BEVs, deren Qualität der Zuordnung nicht so hoch ist (niedrige Zuteilungswahrscheinlichkeit). Täglich genutzte Fahrzeuge haben die geringste durchschnittliche Zuteilungswahrscheinlichkeit (86 %) und die zweithöchste Varianz (2,8 %). Jedoch entsprechen lediglich 6 % der 1073 BEVs den Kriterien eines täglich genutzten Fahrzeugs. Diese geringe Anzahl könnte wiederum den Median und die Varianz beeinflussen. Des Weiteren sind täglich genutzte Fahrzeuge primär durch die Inputvariablen Tage0 (−6,43686) und AntWerktag (−4,23679) determiniert. Diese sind auch für andere FNP (z. B. Kurzstreckenfahrzeug: −3,75175 (Tage0), Wochenendfahrzeug: −4,17534 (AntWerktag)) relevant, was wiederum zu Überlappungen der FNPs führt. Werktagsfahrzeuge (med: 87 %; var: 2,0 %) und Allrounder (med: 87 %; var: 2.8 %) haben eine relativ niedrige Zuteilungswahrscheinlichkeit, aber eine niedrige Varianz. Das könnte bedeuten, dass Werktags-BEVs nicht perfekt mit dem Nutzungsprofil übereinstimmen, aber im Vergleich zu den anderen Profilen immer noch am besten zu den Werktagsfahrzeugen passen.

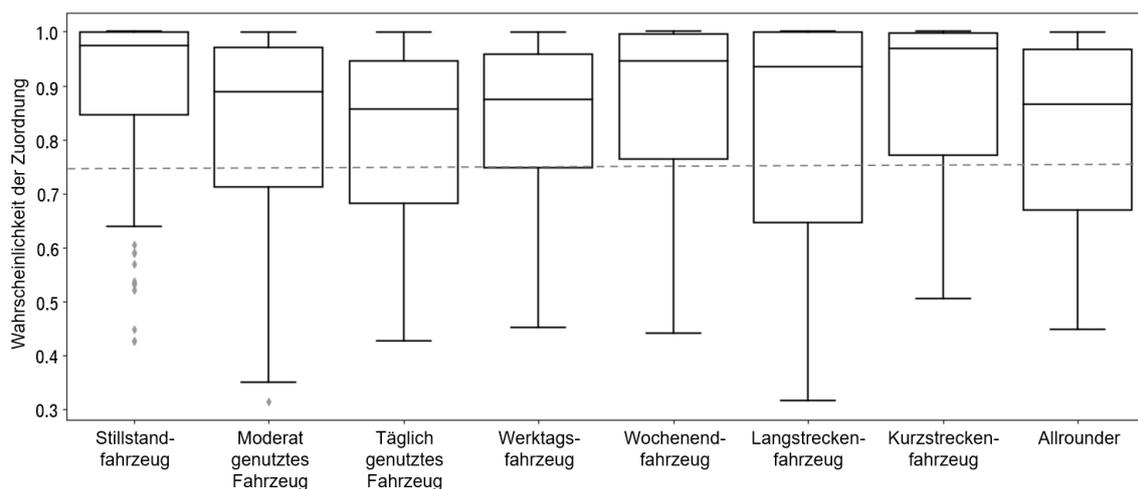


Abbildung 3.8: Wahrscheinlichkeit der Zuteilung zu den Fahrzeugnutzungsprofile

Da lediglich Fahrzeuge betrachtet werden, deren Nutzungen gut mit den einzelnen FNPs übereinstimmen, werden beim Vergleich nur BEVs mit einer Zuteilungswahrscheinlichkeit größer oder gleich 75 % berücksichtigt. Dieses Kriterium reduziert die Stichprobe von 1073 (BEV100) auf 787 (BEV75) BEVs. Die Verteilung entlang der Nutzungsprofile zwischen BEV100 und BEV75 verändert sich dadurch nur geringfügig. Die Anteile der täglich genutzten Fahrzeuge, Wochenendfahrzeuge und Langstreckenfahrzeuge haben die gleichen relativen Größen. Die Anteile der Stillstandfahrzeuge, moderat genutzten Fahrzeuge sowie Kurzstreckenfahrzeuge haben sich leicht erhöht, während die Anteile der Werktagsfahrzeuge und Allrounder leicht gesunken sind.

Vergleich von gewöhnlichen Fahrzeuge und BEVs innerhalb von Deutschland und Kalifornien

Tabelle 3.9 zeigt die Verteilung der umfragebasierten gewöhnlichen Fahrzeuge (GFs) und der sensorbasierten BEVs (BEV75) entlang der FNPs.

Tabelle 3.9: Vergleich von Elektrofahrzeugen und gewöhnliche Fahrzeugen

Fahrzeugnutzungsprofil	Deutschland		Kalifornien	
	GF	BEV75	GF	BEV75
Stillstandfahrzeug	16 % ↗	25 %	13 % ↘	11 %
Moderat genutztes Fahrzeug	17 % ↗	22 %	22 % ↗	47 %
Täglich genutztes Fahrzeug	12 % ↘	5 %	18 % ↘	7 %
Werktagsfahrzeug	14 % ↘	13 %	8 % →	8 %
Wochenendfahrzeug	15 % ↘	14 %	16 % ↘	6 %
Langstreckenfahrzeug	3 % ↗	5 %	14 % ↘	10 %
Kurzstreckenfahrzeug	9 % ↘	2 %	3 % ↗	4 %
Allrounder	14 % →	14 %	6 % ↗	7 %
Σ	1659	516	1681	271

In Deutschland ist der bemerkenswerteste Faktor, dass die relativen Größen von Kurzstreckenfahrzeugen für BEVs (2 %) geringer sind als für GF (9 %). BEVs werden also nicht intensiver für Tagesfahrleistungen unter 10 km als GFs genutzt. Auf der anderen Seite ist der Anteil der Langstreckenfahrzeuge für BEVs (5 %) größer im Vergleich zu GFs (3 %). Das ist möglicherweise ein Hinweis für die Nutzung von BEVs im Fernverkehr. Der Fernverkehr entspricht hier einer zurückgelegten Tagesdistanz von mehr als 100 km. Die beobachteten BEVs können mit einer elektrischen Reichweite von ca. 200 km durchaus die Kriterien des Fernverkehrs erfüllen. Ein weiterer Grund könnte sein, dass es sich bei der Abbildung des Fernverkehrs unter CUMILE um einen modellierten Wert handelt, wodurch dieser möglicherweise unterschätzt wird. In Kalifornien zeigt sich ein anderes Bild. Langstreckenfahrzeuge von BEVs (10 %) haben einen geringeren Anteil im Vergleich zu GFs (14 %). Generell müssen in Kalifornien größere Distanzen aufgrund der Raumstruktur abgedeckt werden. Deshalb werden Fahrzeuge insgesamt stärker im Fernverkehr eingesetzt. In Konsequenz müssen Fahrzeuge eher Distanzen über die Elektroreichweiten (200 km) hinaus zurücklegen, was wiederum den geringeren Anteil von BEVs im Vergleich zu GFs in Kalifornien erklären könnte. Der Anteil der Werktagsfahrzeuge ist nahezu gleich in Deutschland (GF: 14 %; BEV75: 13 %) und Kalifornien (GF: 8 %; BEV75: 8 %). Somit werden BEVs genauso wie GFs für Pendel- und Dienstzwecke eingesetzt. Der generell niedrigere Anteil bei täglich genutzten Fahrzeugen (Deutschland: GF: 12 %; BEV75: 5 %, Kalifornien: GF: 18 %, BEV75: 7 %) und der gleichzeitig hohe Anteil von Wochenendfahrzeugen in Deutschland (Deutschland: GF: 15 %; BEV75: 14 %) bestätigen

die Annahme, dass BEVs eher für spezifische Zwecke eingesetzt werden. Täglich genutzte Fahrzeuge müssen universal einsetzbar für verschiedenste Zwecke (z. B. Freizeitfahrt mit mehreren Passagieren, Stadtverkehr oder Kofferraumvolumen für Großeinkäufe) sein. Da im Datensatz kleine Fahrzeuge die Nutzung von BEVs darstellen, ist beispielsweise der Transport großer Gegenstände schwierig. Insgesamt ist der Anteil von täglich genutzten Fahrzeuge in beiden Ländern sehr gering im Vergleich zu GFs. Alternativ werden BEVs eher als Zweitwagen in einem Haushalt oder als Werbefahrzeug verwendet und finden sich deshalb weniger in der alltäglichen Fahrzeugnutzung wieder. Das könnte wiederum auch eine Erklärung für den hohen Anteil von Stillstandfahrzeugen in Deutschland (GF: 16 %; BEV75: 22 %) sein. In Kalifornien ist der Anteil von Stillstandfahrzeugen für BEVs (11 %) geringer im Vergleich zu GFs (13 %). Hier liegt die Vermutung nahe, dass BEVs in Kalifornien eine stärkere Marktdurchdringung haben als in Deutschland.

Vergleich zwischen den BEVs in Deutschland und Kalifornien

Der wesentlichste Faktor im Vergleich zwischen den BEVs in Deutschland und Kalifornien ist, dass der Anteil von Stillstandfahrzeugen in Deutschland (25 %) mehr als doppelt so groß ist wie in Kalifornien (11 %). Möglicherweise findet ein Einsatz von BEVs in Deutschland primär für spezifische Fahrtzwecke (z. B. zur Arbeit) statt. Dies wird auch bei Betrachtung der zweithöchsten Zuteilungswahrscheinlichkeit bestätigt. Entspricht die höchste Zuteilungswahrscheinlichkeit dem Stillstandfahrzeug, dann ist zu 83 % die zweithöchste Zuteilungswahrscheinlichkeit das Werktagsfahrzeug. Eine Erklärung hierfür wäre, dass BEVs in Deutschland oftmals als Zweitfahrzeug eingesetzt werden. Eine weitere Interpretation für den geringen Anteil von Stillstandfahrzeugen in Kalifornien könnte die Wirksamkeit von BEV-unterstützenden Maßnahmen (z. B. Nutzung von HOV-Lanes für BEVs, Clean Vehicle Rebate Program (CVRP)) sein. In Konsequenz sind BEVs mehr etabliert und werden deshalb auch stärker genutzt. Die Anteile von Langstreckenfahrzeugen (Deutschland: 5 %; Kalifornien: 10 %) und moderat genutzten Fahrzeugen (Deutschland: 22 %; Kalifornien: 47 %) sind in Kalifornien doppelt so groß wie in Deutschland. Dies kann, wie bereits erwähnt, durch länderspezifische Eigenschaften (z. B. Topografie) und die Rolle andere Verkehrsmittel zur Bewältigung der Mobilität erklärt werden. In Kalifornien ist das Auto das mit Abstand am meisten genutzte Verkehrsmittel, weshalb viele Wege auch mit dem Fahrzeug zurückgelegt werden. Es ist offensichtlich, dass die beobachteten BEVs in Kalifornien ihre Rolle als moderat genutztes Fahrzeug gefunden haben. Knapp die Hälfte aller BEVs landet in diesem Nutzungsprofil. Hingegen ist der Anteil von Wochenendfahrzeugen in Deutschland (14 %) mehr als doppelt so groß wie in Kalifornien (6 %). Das könnte bedeuten, dass in Kalifornien eher Fahrzeuge mit Verbrennungsmotor für Freizeitaktivitäten, welche primär am Wochenende stattfinden, stark bevorzugt werden.

3.2.5 Diskussion

In diesem Unterkapitel wurde ein probabilistischer Ansatz gewählt, um die Nutzung von gewöhnlichen Fahrzeugen (von Umfragedaten) mit der von BEVs (von Sensordaten) über FNPs zu vergleichen. Die Sensordaten der BEVs wurden in die gleiche Datenstruktur transformiert, damit sie den Kriterien der FNPs entsprechen. Durch eine SR erfolgte die Berechnung der Wahrscheinlichkeiten der Zuteilung für jedes BEV für jedes FNP. Basierend auf diesen Wahrscheinlichkeiten war eine Messung der Qualität der Zuteilung möglich. Anschließend wurden nur noch BEVs mit einer hohen Zuteilungswahrscheinlichkeit berücksichtigt und mit den gewöhnlichen Fahrzeugen in Deutschland und Kalifornien verglichen.

Die Gegenüberstellung von BEVs und gewöhnlichen Fahrzeugen machte Unterschiede in der Nutzung sichtbar. Die Indikationen werden im Folgenden nochmals stichpunktartig zusammengefasst:

BEVs ...

- ... in Deutschland haben relativ zu gewöhnlichen Fahrzeugen einen vergleichbaren Anteil von Langstreckenfahrzeugen (BEV75: 5 %; GF: 3 %).
- ... in Deutschland haben einen großen Anteil an Stillstandfahrzeugen.
- ... werden in Kalifornien nur wenig am Wochenende genutzt.
- ... haben sowohl in Deutschland als auch in Kalifornien im Vergleich zu gewöhnlichen Fahrzeugen weniger Relevanz in der täglichen Nutzung.
- ... haben sowohl in Deutschland als auch in Kalifornien ähnliche Relevanz in Bezug auf Werktagsnutzung wie gewöhnliche Fahrzeuge.
- ... haben ihre Rolle als moderat genutztes Fahrzeug in Kalifornien gefunden.
- ... werden sowohl in Deutschland als auch in Kalifornien nicht hauptsächlich für kurze Distanzen (Tagesfahrleistung unter 10 km) eingesetzt.

Diese Studie liefert mögliche Hinweise darauf, dass die Marktdurchdringung von BEVs im Fernverkehr gegeben ist. Diese Hypothese gilt es in zukünftigen Forschungsarbeiten weiter zu prüfen. Jedoch ist die Nutzung eines BEV für jegliche Fahrtzwecke, besonders in Deutschland, nicht gegeben. Möglicherweise sind deshalb BEVs für Carsharingkonzepte gut geeignet, da dort spezifische Nutzungszwecke bedient werden sollen. Es zeigt sich somit eine Tendenz, dass BEVs die notwendigen Reichweiten erfüllen. Dies spiegelt sich auch durch die Nutzung von BEVs im Fernverkehr wider. Möglicherweise wurde die in der Literatur viel diskutierte Reichweitenangst (Bonges III & Lusk, 2016; King, Griggs, Wirth & Shorten, 2013) über die Zeit und mit dem Vertrauen abgeschwächt. Warum BEVs immer noch nicht in der breiten Masse und in der alltäglichen Nutzung angekommen sind, kann verschiedene Gründe haben. Auf der einen Seite dauert das Laden von BEVs immer noch länger als das Tanken bei Fahrzeugen mit Verbrennungsmotor. Auch wenn sich die Ladezeit über die Zeit erheblich verkürzt hat,

könnte das immer noch die universelle Nutzung von BEVs beschränken. Auf der anderen Seite ist das Vertrauen in BEVs nicht so stark. Variationen in der elektrischen Reichweite bei extremen Temperaturen und die umweltschädliche Produktion und das Recycling der Batterien unterstützen nicht die Kaufentscheidung für ein BEV. Aufgrund dieser Unsicherheiten präferieren Individuen immer noch zu Fahrzeugen mit Verbrennermotoren. Die Ladezeit wird sich womöglich niemals mit der Tankzeit angleichen. Fahrzeugnutzer müssen in diesem Kontext ihr gewohntes Verhalten (z. B. durch das Laden über Nacht) verändern, wenn sie einen BEV in Erwägung ziehen. Die anderen genannten Punkte könnten jedoch durch eine transparente Politik gelöst werden. Eine konkrete Zusage (z. B. durch weitere Subventionen oder eine bessere Ladeinfrastruktur) der Regierung zur Elektromobilität würde zu mehr Stabilität hinsichtlich der Dauerhaftigkeit von BEVs führen. In Kalifornien zeigen die staatlichen Maßnahmen (z. B. HOV-Lanes) bereits einen positiven Effekt (Delhomme & Gheorghiu, 2016; Guensler et al., 2019). Darüber hinaus sind der Elektromotor (und die dazugehörige Ladetechnologie) und die Elektrifizierung größerer Fahrzeuge (z. B. mit größerem Stauraum) im Vergleich zu traditionellen Antriebssystemen (Benzin, Diesel) noch in der Entwicklung. Diese Verbesserungen könnten zukünftig die Marktdurchdringung von BEVs in beiden Ländern verbessern.

Der Fokus in dieser Studie liegt auf dem Vergleich von BEVs und gewöhnlichen Fahrzeugen über FNPs. BEVs mit einer niedrigen Zuteilungswahrscheinlichkeit wurden bei den Analysen nicht berücksichtigt. Zukünftige Arbeiten könnten genau diese BEVs näher untersuchen und neue elektrospezifische FNPs identifizieren. Um die Vergleichbarkeit zu gewährleisten, fand eine Zuordnung von BEVs zu bestehende FNPs statt. In weiterführenden Untersuchungen sollten alle Fahrzeuge (BEVs und gewöhnliche Fahrzeuge) gemeinsam geclustert werden, um zu überprüfen, wie viele der Fahrzeuge im selben Cluster landen. Ein weiterer Ansatz, um länderspezifische Unterschiede und Gemeinsamkeiten in der Zuteilung zu FNPs zu identifizieren, wäre, zwei Modelle für Deutschland und Kalifornien (Θ^{GER} , Θ^{CAL}) zu schätzen und die Gewichte miteinander zu vergleichen. In diesem Unterkapitel wurde die Nutzung im Kontext von Carsharingkonzepten nicht betrachtet. Weitere Untersuchungen könnten den gleichen Ansatz wählen, um FNPs zu identifizieren und daraus auf die Eignung von BEVs im Kontext von Carsharing zu schließen. Darüber hinaus ist die Stichprobengröße der BEVs besonders in Kalifornien gering. Zukünftige Studien sollten die Stichprobengröße und den Beobachtungszeitraum der BEVs erhöhen, um damit die Robustheit der Ergebnisse zu überprüfen. Die daraus entstandenen Resultate könnten wiederum als Input für bestehende Arbeiten (Seign & Bogenberger, 2013) eingesetzt werden. Insgesamt ist es gelungen, Unterschiede und Gemeinsamkeiten in der Nutzung von BEVs und gewöhnlichen Fahrzeugen

zu identifizieren und neue Erkenntnisse (z. B. Akzeptanz von BEVs im Fernverkehr und in der täglichen Nutzung) zu generieren. Darauf aufbauend konnten neue Implikationen entwickelt werden, um die Akzeptanz für den breiten Einsatz von BEVs zu erhöhen.

3.3 Zwischenfazit

Für die Entwicklung eines Ansatzes zur nutzungsorientierten Fahrzeugsegmentierung wurden in Unterkapitel 3.1 Premiumfahrzeuge auf der Basis bestehender FNPs mit den Dimensionen Zeit und Distanz eindeutig zugeteilt. Dadurch konnte die Notwendigkeit der Nutzung von internen Sensordaten zur Abbildung der Fahrzeugnutzung überprüft und bestätigt werden.

Da sich FNPs inhaltlich überschneiden (z.B. Langstreckenfahrzeuge können auch Wochenendfahrzeuge sein) und jeder Pkw im Laufe seines Fahrzeuglebens in jedes FNP kommt, wurden BEVs in Unterkapitel 3.2 probabilistisch den bestehenden FNPs zugeordnet. Dieser Ansatz liefert eine methodische Weiterentwicklung von einer eindeutigen zu einer probabilistischen Zuteilung.

Die Betrachtung der Fahrzeugnutzung über die Dimensionen Zeit und Distanz zeigt Erkenntnisse hinsichtlich der Unterschiede in der Nutzung zwischen Premiumfahrzeugen/BEVs und gewöhnlichen Fahrzeugen auf. Gründe für diese Abweichungen sind oftmals mit Hypothesen formuliert worden. Beispielsweise gibt es keine Evidenz, ob ein Werktagspremiumfahrzeug oder ein Werktags-BEV auch tatsächlich für Pendelwege eingesetzt wird. Das Gleiche gilt für ein Wochenendfahrzeug. Es ist unklar, für welche Aktivitäten ein Fahrzeug genutzt wird und wie lange diese andauern (z. B. mehrtägige Urlaubsreise). Des Weiteren war es nicht möglich zu untersuchen, für welche Zwecke Stillstandfahrzeuge speziell eingesetzt werden. All diese Punkte sind nicht über die FNPs von Eisenmann und Buehler (2018) erklärbar. Zwar bildet der CUMILE-Datensatz auch Fahrtzwecke ab, um diese Zusammenhänge näher zu untersuchen; da sich aber die Nutzung von Premiumfahrzeugen/BEVs und gewöhnlichen Fahrzeugen unterscheidet, können auch völlig andere Charakteristika und damit auch andere FNPs existieren. Aus diesem Grund werden in Kapitel 4 neue Variablen gebildet und unabhängig von den FNPs von Eisenmann und Buehler (2018) geclustert.

4 Zeit-Distanz-Fahrtzweck- Fahrzeugnutzungsprofile

Im folgenden Kapitel werden Fahrzeuge auf der Basis der drei Dimensionen Zeit, Distanz und Fahrtzweck betrachtet. In Unterkapitel 4.1 erfolgt die Approximation der Fahrtzwecke anhand von Regelmäßigkeiten in der Fahrzeugnutzung und der Verknüpfung von Point-of-Interest (POI)-Informationen. Zur Validierung der Fahrtzwecke werden diese anhand des bereits verwendeten CUMILE-Datensatzes geschätzt und mit den approximierten Fahrtzwecken verglichen. In Unterkapitel 4.2 finden die Einteilung der Fahrzeuge über die drei Dimensionen in FNPs sowie deren Interpretation statt.

4.1 Identifikation des Fahrtzwecks

4.1.1 Motivation

Bereits die Analysen von Zeit-Distanz-FNPs lassen vermuten, dass der Fahrtzweck ein wichtiger Differenzierungsfaktor für die Fahrzeugnutzung sein könnte. Beispielsweise stellt die Abbildung 3.6 dar, dass die Anzahl der Passagiere stärker durch den Fahrtzweck als durch die Distanz bestimmt werden könnte. Auch die Tabelle 3.6 zeigt auf, dass die relativen Anteile über die vier Jahreszeiten für distanzorientierte FNPs (Kurzstrecken- und Langstreckenfahrzeug) eher konstant bleiben, während zeitorientierte FNPs (Stillstand- und Wochenendfahrzeug) Veränderungen aufweisen. Unter der Annahme, dass Wochenendfahrzeuge primär für Freizeitaktivitäten eingesetzt werden, bestätigt sich, dass der Fahrtzweck ein wichtiges Kriterium für die Fahrzeugnutzung ist. Die bisher betrachteten FNPs berücksichtigen aber nicht die Fahrtzwecke als clusterbeschreibende Dimension.

Um die Relevanz des Fahrtzwecks zu verdeutlichen, werden in Abbildung 4.1⁸ ein Fahrzeug und dessen Fahrzeugnutzung schematisch an zwei Tagen in den drei Dimensionen dargestellt. Die x- und y-Achsen skizzieren den Aktionsraum eines Fahrzeugs, während die z-Achse die Zeit charakterisiert. An beiden Tagen wurde eine Tagesdistanz von 50 km jeweils an einem Werktag zurückgelegt. Nach den Eigenschaften der FNPs von Eisenmann und Buehler (2018) wäre die Nutzung des Fahrzeugs identisch. Wird die Fahrzeugnutzung in drei Dimensionen betrachtet, sind durchaus Unterschiede zwischen den beiden Tagen ersichtlich. An Tag 1

⁸Grafik in Anlehnung an Chen, Ma, Susilo, Liu und Wang (2016)

wurde das Fahrzeug um 07:00 Uhr für die Fahrt zur Arbeit verwendet, um 12:00 Uhr für die Fahrt zum Mittagessen, um 13:00 Uhr für die Fahrt zurück in die Arbeit und gegen 18:30 Uhr für die Fahrt wieder nach Hause. An Tag 2 wurde das Fahrzeug um 09:00 Uhr für die Fahrt zum Einkaufen eingesetzt. Nachdem es dort knapp 4 Stunden gestanden hatte, wurde gegen 16:00 Uhr die Heimreise angetreten. Die Anforderungen der Individuen an das Fahrzeug sind für beide Tage völlig unterschiedlich. Das Ziel einer Fahrt und der daraus abgeleitete Fahrtzweck bieten somit einen detaillierteren Blick auf die Fahrzeugnutzung.

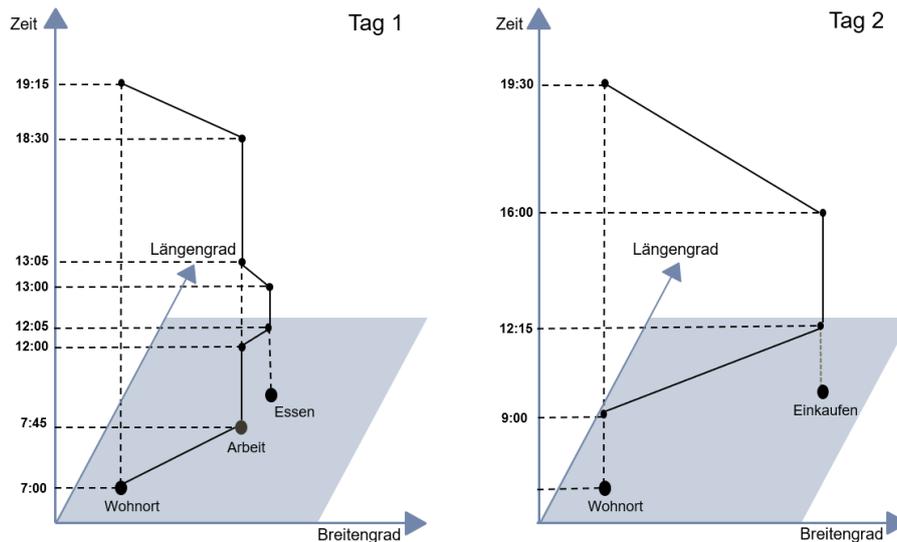


Abbildung 4.1: Schematische Darstellung der Fahrzeugnutzung in drei Dimensionen

Wie bereits diskutiert, ist der Fahrtzweck aus Sensordaten, im Gegensatz zu herkömmlichen Umfragedaten, nicht direkt ablesbar. Abbildung 4.1 zeigt aber schematisch auf, dass mit Hilfe von geografischen Informationen die Betrachtung der Fahrzeugnutzung durch Fahrtzwecke differenzierter ist. Aus diesem Grund werden im Folgenden Sensordaten mit GPS-Informationen (Längen- und Breitengrad) verwendet. Die Sensordaten stimmen in der Datenerfassung und -speicherung mit den bisherigen Sensordaten nicht überein, weswegen sie später im Detail erklärt werden.

4.1.2 Literaturübersicht

In der Literatur existieren einige Studien, welche sich mit der Identifikation von Wege- bzw. Fahrtzwecken beschäftigen. Wie schon in Abbildung 4.1 dargestellt, sind Kenntnisse über den Standort zwingend erforderlich, um Wege- bzw. Fahrtzwecke zu bestimmen. Primär stammen die angewandten Methoden aus dem Feld der Mobilfunkdatenanalyse. Hierbei wird der Standort von Individuen entweder direkt über einen Logger (z. B. über eine App) oder über den nächstgelegenen Funkmast ermittelt. Je nach Zweck kann sich die Herangehensweise zur Identifikation unterscheiden. Ein Wohnort oder ein fester Arbeitsort ist für gewöhnlich

ein regelmäßig angesteuertes Ziel, welches über die Routine identifiziert werden kann. Andere Fahrtzwecke, wie Einkaufen, lassen sich über die Verknüpfung mit Geoinformationssystem (GIS)-Daten ermitteln. Befindet sich am Zielort ein Einkaufszentrum, könnte darüber der Fahrtzweck „Einkaufen“ approximiert werden.

Der **Wohn- und Arbeitsort** wird in vielen Studien regelbasiert ermittelt. Alexander, Jiang, Murga und González (2015) berechnen auf der Basis von acht Milliarden anonymisierten Mobilfunkaufzeichnungen von ungefähr zwei Millionen Nutzern mit einem regelbasierten Ansatz den Wohn- und Arbeitsort der Menschen. Der Wohnort definiert sich als der Ort mit den meisten Aufenthalten zwischen 19:00 Uhr und 8:00 Uhr. Der Arbeitsort wird, ausgehend vom Wohnort, durch die größte absolute Distanz zum Wohnort beschrieben. Dabei gilt $\max(d \cdot n)$, wobei n die Anzahl der Aufenthalte an Werktagen zwischen 08:00 Uhr und 19:00 Uhr und d die Distanz zwischen dem Wohn- und dem Standort darstellt. Die verwendeten Kriterien werden auch durch die Arbeiten von Levinson und Kumar (1994) und Schafer (2000) gestützt. Isaacman et al. (2011) nutzen Mobilfunkdaten, um wichtige Standorte, wie den Wohn- und den Arbeitsort, zu bestimmen. Die Daten umfassen 400 Millionen Datenpunkte von 170 000 eindeutigen Mobilfunkgeräten. Zusätzlich machten 37 freiwillige Personen Angaben zu ihren wichtigen Standorten, welche als Grundwahrheit für das Modell dienen. Damit wird eine logistische Regression geschätzt und auf die verbliebenen Daten angewandt. Die Schätzung zeigt, dass der Wohnort durch Aufenthalte zwischen 19:00 Uhr und 07:00 Uhr und der Arbeitsort durch Aufenthalte zwischen 13:00 Uhr und 17:00 Uhr determiniert ist. Somit decken sich diese Annahmen mit den Kriterien von Alexander et al. (2015). Auch Calabrese, Diao, Di Lorenzo, Ferreira Jr. und Ratti (2013) versuchen, mittels Mobilfunkdaten Informationen zur individuellen Mobilität zu extrahieren, und nutzen einen regelbasierten Ansatz zur Bestimmung der Wohnorte. Hierbei werden Verbindungen gezählt, an denen sich der Nutzer am gleichen Mast in einem Zeitintervall von 18:00 Uhr bis 08:00 Uhr eingewählt hat.

Zur Bestimmung **anderer Zwecke** verknüpfen viele Studien GIS-Daten mit GPS-Lokationen. GIS-Daten enthalten Informationen zur Landnutzung, welche durch POIs gekennzeichnet sind (Gong, Liu, Wu & Liu, 2016). Landnutzung definiert sich über die Nutzung von Böden und Landflächen durch Menschen, wie die Unterteilung der Fläche in Industrie- und Wohnfläche. POIs sind GPS-Punkte, welche für Individuen von Relevanz sind und z.B. über Open-Street-Map (OSM) ausgelesen werden können. Ein POI kann die GPS-Koordinaten eines Restaurants, eines Einkaufszentrums, eines Kinos oder einer anderem Ort sein. POIs werden in der Praxis hauptsächlich in Navigationssystemen und Routenplanern eingesetzt. Phithakkitnukoon, Horanont, Di Lorenzo, Shibasaki und Ratti (2010) nutzen POIs, um verschiedene Freizeitaktivitäten (Essen, Einkaufen, Unterhaltung und Erholung) zu identifizieren. Der Beobachtungsraum wird in $500\text{ m} \times 500\text{ m}$ große Zellen aufgeteilt und jede Zelle wird mit POI-Informationen aus den vier genannten Kategorien befüllt. Über den

Satz von Bayes wird die wahrscheinlichste Aktivität durch die Anzahl der POIs und deren Zugehörigkeit zu einer der vier Freizeitaktivitäten innerhalb der Zelle berechnet. Auch Gong, Liu et al. (2016) nutzen POIs und Taxidaten aus Shanghai zur Erkennung von Wege- bzw. Fahrtzwecken. Ähnlich wie Phithakkitnukoon et al. (2010) schätzen Gong, Liu et al. (2016) für $1\text{ km} \times 1\text{ km}$ große Zellen mit POI-Informationen eine Wahrscheinlichkeit des Zwecks bei gegebenem Ausladepunkt. Hierbei wird zwischen acht Aktivitätskategorien (zu Hause, Arbeit, ÖPNV, Essen, Einkaufen, Erholung, Schule, Laden und Medizin) unterschieden. Je weiter der POI vom Ausladepunkt entfernt ist, desto unwahrscheinlicher ist das Aufsuchen des POI. Bohte und Maat (2009) berechnen die Wegezwecke von Individuen, indem sie einen 50-m-Radius um den Zielort ziehen. Befindet sich ein POI innerhalb des Radius, wird unterstellt, dass dieser POI besucht wurde. Bei mehreren POIs innerhalb des Radius wird der nächstgelegene POI zur Bestimmung des Zwecks verwendet. Auch Wolf, Guensler und Bachman (2001) bestimmen die Fahrtzwecke von Fahrzeugen aus einer GPS-gestützten Umfrage in Atlanta mittels Informationen zur Landnutzung. Hierbei werden Flächen speziellen Kategorien (z. B. Restaurant) zugeordnet. Endet eine Fahrt in dieser Fläche, wird der Fahrt der entsprechende Fahrtzweck (z. B. Essen) zugewiesen.

Eine Vielzahl von Studien (Deng & Ji, 2010; Gong, Kanamori & Yamamoto, 2018; Gong, Yamamoto & Morikawa, 2016; Xiao, Juan & Zhang, 2016) nutzt überwachte Verfahren, um einen Zusammenhang zwischen Inputfaktoren und den Fahrtzwecken herzustellen. Allerdings ist der Fahrtzweck im Datensatz bereits vorhanden, weshalb dieser Ansatz für die zugrundeliegenden Daten nicht geeignet ist. Zudem ist die Stichprobe bei den genannten Studien sehr gering, da die Datensammlung (passives Speichern von GPS-Informationen und die aktive Angabe des Wegezwecks durch die Probanden) sehr aufwendig und teuer ist.

Gong, Morikawa, Yamamoto und Sato (2014) liefern einen umfassenden Überblick über bestehende Methoden zur Anreicherung von GPS-Daten mit weiteren Informationen, wie beispielsweise dem Wege- bzw. Fahrtzweck. Gong et al. (2014) zeigen auf, dass der regelbasierte Ansatz sowie die Verknüpfung von GIS-Informationen zur Identifikation von Wege- bzw. Fahrtzwecken die beliebtesten Methoden sind. Dabei sind über GPS-Daten Informationen zur Standzeit, dem Zeitpunkt am Zielort und der Regelmäßigkeit extrahierbar. GIS-Daten enthalten Informationen zur Landnutzung und zu POIs und können mit GPS-Mobilitätsdaten verknüpft werden. Auch Shen und Stopher (2014) bieten einen Überblick über existierende GPS-Umfragen und deren Vorgehen zur Identifikation von Wege- bzw. Fahrtzwecken. Gleichmaßen sind hier regelbasierte Ansätze sowie die Verknüpfung mit GIS-Daten die dominierenden Methoden.

Zusammenfassend kann gesagt werden, dass der Wohn- und Arbeitsort oft über das routinierte Verhalten und die zeitliche Abfolge bestimmt werden. Andere, weniger routinierte Zwecke (Einkaufen, Restaurantbesuche etc.) werden mit der Anbindung von POI-Informationen

ermittelt. Die aufgelisteten Studien beschäftigen sich entweder mit der Identifikation des Wohn- und Arbeitsortes oder mit der Identifikation von anderen Zwecken. Im Folgenden werden die vorgestellten Sensordaten verwendet, um sowohl routinierte als auch nicht routinierte Fahrtzwecke zu approximieren.

4.1.3 Daten

Die verwendete Datenquelle ermöglicht es, aus einer Vielzahl von Signalen flexibel zu wählen und diese im Rahmen eines Skriptes abzurufen. Das verwendete Skript zur Analyse der Fahrzeugnutzung auf der Basis von drei Dimensionen berücksichtigt auch räumliche Aspekte, weshalb es sich von der Datenstruktur und den berücksichtigten Variablen mit den in Kapitel 3 verwendeten Daten teilweise unterscheidet (Tabelle 4.1).

Die Variablen „anonymisierte Fahrgestellnummer“ (*van*), „Fahrt-ID“ (*driveID*) und „Kilometerstand“ (*mileage*) sind mit den verwendeten Variablen aus Kapitel 3 identisch. Der „Zeitstempel“ (*timestamp*) berücksichtigt das Datum und die Uhrzeit und wird in Unixzeit bereitgestellt. Die Unixzeit zählt die vergangenen Sekunden seit dem 01.01.1970 00:00:00 Uhr (koordinierte Weltzeit) und kann mit gängigen Softwarepaketen aus Python in ein lesbares Zeitstempelformat (yyyy-MM-dd HH:mm:ss) umgewandelt werden. Mit Hilfe der „GPS-Position“ (*lat, lng*) lässt sich die lokale Ortszeit (Greenwich Mean Time (GMT)) bestimmen. Die GPS-Position entspricht dem Längen- und Breitengrad und wird aus den Daten des Navigationssystems extrahiert. Während einer Fahrt werden mehrere Datenpakete versendet, wodurch eine Fahrt *i* auch mehrere Datenpunkte besitzt. Ein Datenpaket wird verschickt, sobald zehn Minuten seit dem Versenden des letzten Paketes vergangen sind oder eine Datengröße von 120 Megabyte überschritten ist.

Tabelle 4.1: Datenstruktur: GPS-basierte Sensordaten

Variable	Name	Datenformat
<i>van</i>	Anonymisierte Fahrgestellnummer	String
<i>driveID</i>	Fahrt-ID	String
<i>mileage</i>	Kilometerstand	Integer
<i>timestamp</i>	Zeitstempel	yyyy-MM-dd HH:mm:ss
<i>lat, lng</i>	GPS-Position	Float

Wird nun nach dem frühesten $start_i$ - und dem spätesten end_i -Zeitstempel einer Fahrt *i* gefiltert und die Events in eine Zeile transformiert, so entsteht daraus ein Datensatz auf Fahrtebene mit Informationen zu Fahrtbeginn und -ende. Die Differenz des Kilometerstandes

und die des Zeitstempels entspricht der Fahrdistanz $distance_i$ bzw. der Fahrdauer $duration_i$. Die Abweichung des Zeitstempels am Ende der Fahrt i vom Zeitstempel zu Beginn der darauffolgenden Fahrt $i + 1$ definiert die Standzeit $standingtime_i$ nach der Fahrt i .

Um die Daten zu bereinigen und unrealistische Fahrten aus dem Datensatz zu entfernen, wurden diese auf einer Karte visualisiert und verschiedene Metriken (durchschnittliche Fahrdistanz, Fahrtzeit, Korrelation Fahrdistanz/Fahrtzeit) betrachtet. Fahrten mit folgenden Kriterien wurden aus dem Datensatz entfernt: (1) Fahrten mit einer Fahrtzeit von mehr als drei Stunden sowie einer Distanz kleiner 10km; (2) Fahrten mit einer Fahrtzeit von mehr als acht Stunden.

Von den 90 165 verfügbaren Fahrten erfüllen 517 (0,57 %) die Ausschlusskriterien. Insgesamt sind 625 Fahrzeuge mit 89 648 validen Fahrten in Deutschland im Datensatz enthalten. Die mittlere Beobachtungsdauer pro Fahrzeug beträgt 36 Tage, 3 Stunden, 43 Minuten und 19 Sekunden. Die mittlere Anzahl der Fahrten pro Fahrzeug beläuft sich auf 134 Fahrten.

In Abbildung 4.2 sind die Fahrtverflechtungen der zugrundeliegenden Daten dargestellt. Jeder Punkt auf der Karte repräsentiert einen GPS-Punkt am Anfang oder Ende einer Fahrt. Die Linien sind die Verbindungen zwischen Fahrtenanfang und -ende. Hierbei wird auch der Fernverkehr zwischen größeren Städten (z. B. Berlin und Hamburg) gut sichtbar. Die Darstellung wurde mit kepler.gl, einem Tool zur interaktiven Visualisierung von geografischen Informationen von Uber, erstellt. Kepler.gl kann als Paket in die Python-Umgebung integriert und verwendet werden.

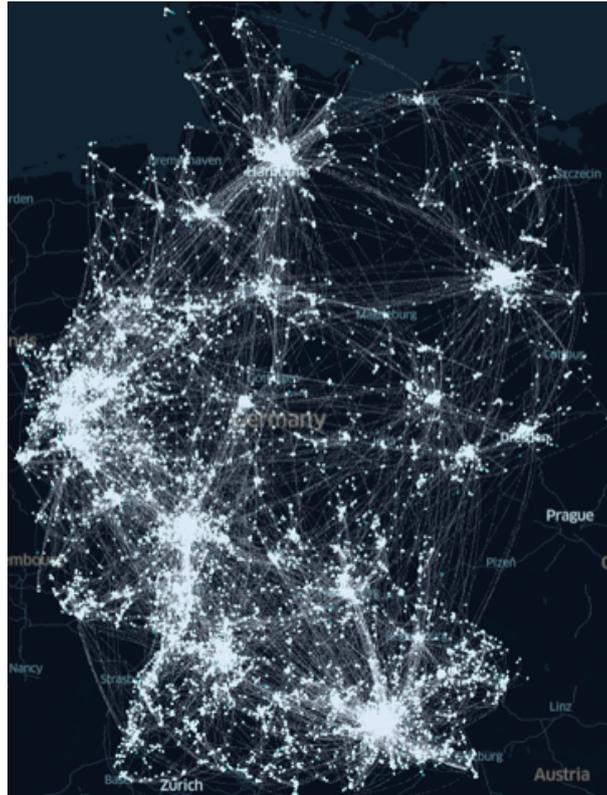


Abbildung 4.2: Darstellung der Fahrtverflechtungen

4.1.4 Methodik

Approximation des Fahrtzwecks

Gong et al. (2014) sowie Shen und Stopher (2014) haben gezeigt, dass der regelbasierte Ansatz und die Verknüpfung von POI-Informationen mit GPS-Koordinaten einer Fahrt die gängigsten Methoden zur Approximation der Wege- bzw. Fahrtzwecke sind.

Ziel dieses Abschnitts ist es, eine Fahrttabelle mit zusätzlichen Informationen zum Fahrtzweck zu erstellen. Diese Informationen werden wiederum auf eine Fahrzeugtabelle aggregiert. Auf der Basis der Fahrzeugtabelle werden im darauffolgenden Unterkapitel 4.2 dreidimensionale FNP's bestimmt.

Die Approximation des Fahrtzwecks besteht aus insgesamt neun Schritten, wobei das Vorgehen im Umgang mit **routinierten** und **nicht routinierten Fahrten** getrennt werden kann:

Routinierte Fahrten:

1. Zusammenführung von Fahrten
2. Wohnort
3. Fahrtzweck „nach Hause“
4. Aktionsradius eines Fahrzeugs
5. Touren

6. Wichtige Standorte eines Fahrzeugs
7. Arbeitsort
8. Fahrtzweck „zur Arbeit“

Nicht routinierte Fahrten:

9. Weitere Fahrtzwecke durch Verknüpfung mit POI-Informationen

Im **1. Schritt** findet eine Zusammenführung der Fahrten mit geringen Standzeiten statt. Die Standzeit st_t zum Zeitpunkt t wird ermittelt, indem die Startzeit der zukünftigen Fahrt $time_{t+1}^{start}$ von der Zielzeit der gegenwärtigen Fahrt $time_t^{end}$ subtrahiert wird. Ist die Standzeit geringer als ein bestimmter Schwellenwert δ_{time} , werden die Fahrten zu den Zeitpunkten t und $t + 1$ zusammengeführt. Durch die Fusionierung von Fahrten ist speziell der Fernverkehr besser beobachtbar, da bei längeren Fahrten in der Regel Pausen eingelegt werden. Ohne die Berücksichtigung von kurzen Standzeiten würden zukünftige Analysen den Fernverkehr unterschätzen. Darüber hinaus sollten die zugrundeliegenden Daten möglichst vergleichbar mit traditionellen HTS sein. Stopher et al. (2007) zeigen auf, dass 7,4% der Fahrten in HTS nicht berichtet werden, da diese von den Teilnehmern, wahrscheinlich aufgrund von geringen Standzeiten, kombiniert werden. Durch das Zusammenbringen von Fahrten verringert sich die Anzahl der Fahrten von 89 648 auf 78 995 (Δ 11,9%). Die relativ hohe Anzahl fusionierter Fahrten ist für Sensordaten durchaus nachvollziehbar, da jeder Fahrtstopp als neue Fahrt gemessen wird. Wird eine Fahrt kombiniert, stammen Fahrtinformationen (Kilometerstand, Zeitstempel, GPS-Koordinaten) zu Fahrtbeginn von der Fahrt tr_t , während die Fahrtinformationen zu Fahrtende von der Fahrt tr_{t+1} entnommen werden. Dieser Prozess wird so lange wiederholt, bis keine Fahrt mehr eine Standzeit von weniger als δ_{time} besitzt. Die Variable $paused_{count}$ gibt die Anzahl der verknüpften Fahrten an, $paused_{time}$ beinhaltet die Summe der Standzeiten von integrierten Fahrten.

Im **2. Schritt** wird der Wohnort ($home_{lat}$, $home_{lng}$) über die Anzahl der Fahrzeugübernachtungen je Fahrzeug bestimmt. Die GPS-Koordinaten am Ende aller Fahrten, welche zum Zeitpunkt t enden und die darauffolgende Fahrt zum Zeitpunkt $t + 1$ an einem anderen Tag beginnt, werden markiert. Steht ein Fahrzeug mehrere Tage am gleichen Standort, wird diese Lokation um die entsprechende Anzahl an Tagen gewichtet. Aus den markierten GPS-Koordinaten wird für jedes Fahrzeug die durchschnittlichen Längen- und Breitengrade ($home_{lat}$, $home_{lng}$) berechnet.

Im **3. Schritt** wird der Fahrtzweck „nach Hause“ auf der Basis des approximierten Wohnortes ($home_{lat}$, $home_{lng}$) bestimmt. Um den berechneten Wohnort wird ein Radius von δ_{radius}^h gezogen. Alle Fahrten, welche in diesem Radius enden, werden mit dem Fahrtzweck „nach Hause“ deklariert.

Im **4. Schritt** wird der Aktionsradius (1–10 km, 11–30 km, 31–100 km, > 100 km) eines Fahrzeugs auf der Basis des berechneten Wohnortes ($home_{lat}$, $home_{lng}$) bestimmt. Analog zur Approximation des Fahrtzwecks „nach Hause“, werden Radien in der Größe von δ_{radius}^{fa} um den Wohnort gezogen. Alle Fahrten, welche innerhalb dieser Radien stattfinden, werden dementsprechend vermerkt.

Im **5. Schritt** werden mit Hilfe des berechneten Wohnortes ($home_{lat}$, $home_{lng}$) Touren bestimmt. Unter der Annahme, dass ein Fahrzeug vermehrt zu Hause geladen wird, ist die Betrachtung von Touren speziell im Kontext der Elektromobilität von Relevanz. Für die Definition von Touren werden zwei unterschiedliche Ansätze verwendet. Variante 1: Eine Tour beginnt und endet zu Hause. Variante 2: Eine Tour beginnt zu Hause und endet, sobald das Fahrzeug übernachtet. Durch die Berücksichtigung beider Varianten ist es möglich zu untersuchen, wie oft ein Fahrzeug Touren nicht zu Hause beendet hat. Darüber hinaus liefert die durchschnittliche Anzahl der Fahrten bei einer Tour Rückschlüsse auf die Komplexität der Fahrzeugnutzung.

Im **6. Schritt** werden die wichtigen Standorte eines Fahrzeugs mit dem sogenannten Density-based Spatial Clustering of Applications with Noise (DBSCAN) bestimmt (Ester, Kriegel, Sander, Xu et al., 1996; Géron, 2019). Hierbei werden die Breiten- und Längengrade als Eingangsvariablen verwendet. Genauso wie beim k-Means-Algorithmus handelt es sich beim DBSCAN um ein unüberwachtes Verfahren. Die zentrale Idee des DBSCAN ist, hochverdichtete Regionen, welche durch niedrigverdichtete Regionen getrennt sind, im Datenraum zu identifizieren. Formal ausgedrückt kann ein Punkt p drei Zustände annehmen: (1) Kernobjekt, (2) Randobjekt oder (3) Ausreißer. Ausgehend von einem Punkt p wird in einem ersten Schritt der Abstand $dist(p, q)$ (z. B. über die euklidische Distanz) zu einem beliebigen Punkt q berechnet, sodass gilt:

$$N_{\delta_{dbscan}}(p) = \{q \in dist(p, q) \leq \delta_{dbscan}\}, \quad (4.1)$$

wobei $N_{\delta_{dbscan}}(p)$ der Anzahl von Punkten, deren Distanz $dist(p, q)$ einen bestimmten Schwellenwert δ_{dbscan} nicht überschreitet, entspricht.

Auf dieser Basis wird ein Punkt p als „Kernobjekt“ bezeichnet, wenn gilt:

$$|N_{\delta_{dbscan}}(p)| \geq MinPts, \quad (4.2)$$

wobei $MinPts$ die Mindestanzahl von Punkten innerhalb eines Radius δ_{dbscan} von Punkt p darstellt. Erfüllt ein Punkt diese Bedingung, wird der Punkt mit all seinen Punkten innerhalb des Radius δ_{dbscan} als ein Cluster deklariert.

Der Punkt p wird als „Randobjekt“ bezeichnet, wenn gilt:

$$\begin{aligned} |N_{\delta_{dbscan}}(p)| &< MinPts \\ p &\in |N_{\delta_{dbscan}}(q)| \end{aligned} \quad (4.3)$$

$|N_{\delta_{dbscan}}(p)|$ ist kleiner als eine gegebene Mindestanzahl $MinPts$ von Punkten innerhalb eines Radius δ_{dbscan} und gleichzeitig befindet sich der Punkt p innerhalb des Radius eines anderen Kernobjektes q . Ist ein Randobjekt p Element von $|N_{\delta_{dbscan}}(q)|$ sowie $|N_{\delta_{dbscan}}(d)|$ der Kernobjekte q und d , werden die Cluster den Kernobjekten q und d zusammengefasst.

Der Punkt p wird als „Ausreißer“ bezeichnet, wenn gilt:

$$\begin{aligned} |N_{\delta_{dbscan}}(p)| &< MinPts \\ p &\notin |N_{\delta_{dbscan}}(q)| \end{aligned} \quad (4.4)$$

$|N_{\delta_{dbscan}}(q)|$ ist kleiner als eine gegebene Mindestanzahl $MinPts$ von Punkten innerhalb eines Radius δ_{dbscan} und gleichzeitig befindet sich der Punkt p außerhalb des Radius eines anderen Kernobjektes q . Dieser Prozess wird mit allen Punkten durchgeführt. Im Vergleich zum distanzbasierten k-Means-Algorithmus handelt es sich beim DBSCAN um ein dichtebasiertes Verfahren, wodurch auch nicht lineare Zusammenhänge entdeckt werden können. Ein weiterer Vorteil gegenüber dem k-Means-Algorithmus ist, dass das DBSCAN selbstständig die Anzahl der Cluster bestimmt. Dies ist speziell bei der Bestimmung wichtiger Standorte eines Fahrzeugs von Relevanz, da sich die Anzahl der wichtigen Standorte von Fahrzeug zu Fahrzeug unterscheiden kann. Ein weiterer Vorteil gegenüber anderen unüberwachten Verfahren ist, dass das DBSCAN nicht alle Datenpunkte zwingend einem Cluster zuordnet. Die nicht zugeordneten Fahrten (Ausreißer) können dann wiederum als nicht routinierte Fahrten interpretiert werden. Beispielsweise nutzen Magdolen, Ecke, Hilgert, Chlond und Vortisch (2020) einen regelbasierten Ansatz, um nicht routinierte Touren zu identifizieren. Über den DBSCAN wäre es möglich, dies automatisiert zu ermitteln. Das DBSCAN kommt bereits in vielen Studien aus der Mobilitätsanalyse zum Einsatz, um die wichtigen Standorte eines Individuums zu lokalisieren (Gong, Sato, Yamamoto, Miwa & Morikawa, 2015; Hidayah & Wulandari, 2018; Panahandeh & Åkerblom, 2015). Abbildung 4.3 zeigt beispielhaft die Identifikation der wichtigen Standorte sowie den approximierten Wohnort eines Fahrzeugs. Aus datenschutzrechtlichen Gründen wurden aus der Darstellung die Breiten- und Längengrade entfernt. Die Fahrzeugübernachtungen sind in der Abbildung rot markiert. Aus diesen GPS-Koordinaten ergibt sich im Median der approximierte Wohnort (schwarz). Die wichtigen Standorte (Standort 1 (blau), Standort 2 (grün)) sowie unregelmäßige Fahrten (grau) können über das DBSCAN bestimmt werden. Dieses Fahrzeug hat einen wichtigen Standort (blau)

relativ nahe am approximierten Wohnort und einen wichtigen Standort (grün) etwas weiter entfernt. Die unregelmäßigen Fahrten verteilen sich großflächig und sind, im Vergleich zu den wichtigen Standorten, weiter vom approximierten Wohnort entfernt.

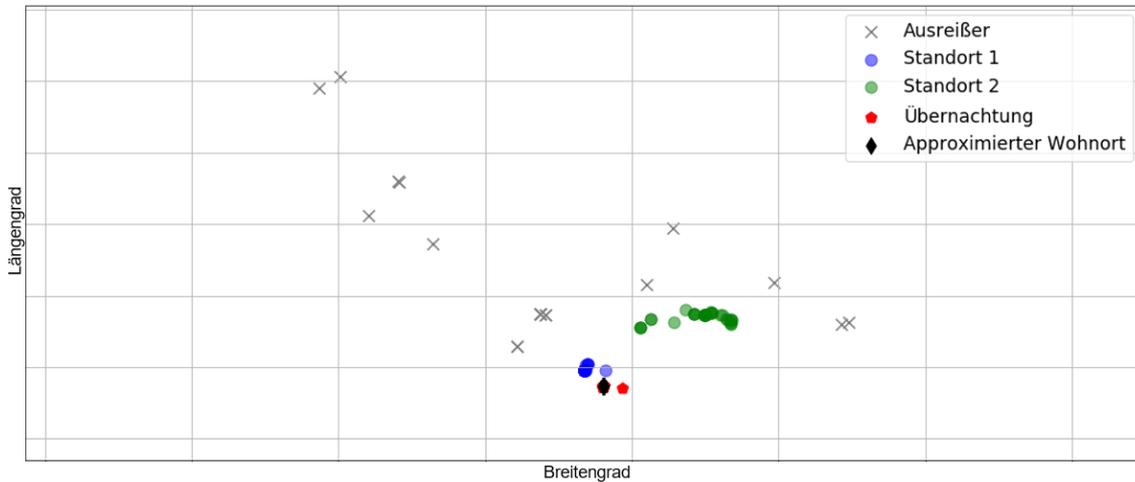


Abbildung 4.3: Wichtige Standorte und approximierter Wohnort eines Fahrzeugs

Im **7. Schritt** kann auf der Basis der wichtigen Standorte eines Fahrzeugs der Arbeitsort regelbasiert erfasst werden. Zunächst werden nur Fahrten betrachtet, die nicht mit dem Fahrtzweck „nach Hause“ deklariert sind und einen wichtigen Standort ansteuern. Die Approximation des Arbeitsortes erfolgt über die Standzeit (δ_{time}^w) und die Regelmäßigkeit ($MinPts^w$) an einem Standort. Steht ein Fahrzeug an mindestens zwei Werktagen in der Woche ($MinPts^w$) drei Stunden (δ_{time}^w) während des Tages an einem der wichtigen Standorte, wird dieser Standort als Arbeitsort ($work_{lat}, work_{lng}$) ausgewiesen. Erfüllen mehrere Standorte die genannten Kriterien, wird der Standort gewählt, welcher die Kriterien am häufigsten erfüllt. Im **8. Schritt** wird der Fahrtzweck „zur Arbeit“ bestimmt. Analog zur Approximation des Fahrtzwecks „nach Hause“ wird auch hier um die mittlere GPS-Koordinate des Arbeitsortes ein Radius von 1 km gezogen. Alle Fahrten, welche in diesem Radius enden, werden als Fahrtzweck „zur Arbeit“ deklariert.

Sind die routinierten Fahrten als Fahrten nach Hause und zur Arbeit bestimmt, kann für die restlichen Fahrten der Fahrtzweck ausgewiesen werden. Zur Identifikation weiterer Fahrtzwecke werden POI-Informationen an die GPS-Koordinaten am Ende nicht routinierten Fahrten angespielt.

Im **9. Schritt** werden „weitere Fahrtzwecke“, wie „Einkaufen“ oder „Freizeit“, welche nicht über die Routine oder die Standzeit bestimmbar sind, mit Hilfe von POI-Daten approximiert. Wie bereits erwähnt, wird jeder POI durch eine GPS-Koordinate mit entsprechender Referenz zu einem Layer oder zu einer Klasse repräsentiert. Die Benennung der Layer und Klassen ist durch die jeweilige Datenquelle (OSM) vorgegeben. Beispielsweise enthält der Layer „Catering“ die Klassen „Restaurant“, „Fast Food“, „Café“, „Pub“ und „Bar“. Für diese Arbeit wurden

insgesamt 13 relevante Layer mit 130 Klassen verwendet. Hierbei ist das Ziel, einen Fahrtende-GPS-Punkt einem Layer bzw. einer Klasse zuzuweisen. Wird z. B. einem Fahrtende-GPS-Punkt die Klasse „Supermarkt“ mit dem entsprechenden Layer „Shopping“ zugeteilt, kann mit dieser Kenntnis über das Ziel der Fahrt der Fahrtzweck „Einkaufen“ zugewiesen werden. Besteht $\sum_{i=1}^K q_i$ aus K unterschiedlichen Klassen, werden Wahrscheinlichkeiten, basierend auf der Entfernung zum Fahrtende-GPS-Punkt, berechnet. Durch Einsatz des k-Nearest Neighbors (kNN)-Algorithmus, können den Fahrtende-GPS-Punkten jeweils Layer und Klassen automatisiert zugewiesen werden. Der kNN-Algorithmus ist eines der simpelsten überwachten Klassifikationsverfahren (Fix, 1951). Über die nicht parametrische Methode werden als Erstes die k_{knn} nächsten POI-GPS-Punkte q_i zum Fahrtende-GPS-Punkt p identifiziert. Danach findet, analog zum DBSCAN, die Berechnung der euklidischen Distanzen $d(p, q_i)$ zwischen p und q_i statt. Darauf aufbauend werden die Inversen der euklidischen Distanzen $d(p, q_i)^{-1}$ bestimmt und aufsummiert ($\sum_{i=1}^k d(p, q_i)^{-1}$). Durch eine Normalisierung lassen sich die Gewichte auf 1 ($\sum_{i=1}^k \theta_i = 1$) summieren:

$$\theta_i = \frac{d(p, q_i)^{-1}}{\sum_{i=1}^k d(p, q_i)^{-1}} \quad (4.5)$$

Wie bei Gong, Liu et al. (2016) gilt auch hier: Je weiter der POI-Punkt im Vergleich zu den restlichen $k^{knn} - 1$ POI-Punkten vom Fahrtende-GPS-Punkt entfernt ist, desto geringer ist die Wahrscheinlichkeit. Der Layer bzw. die Klasse mit der höchsten Gewichtung wird daraufhin dem Fahrtende-GPS-Punkt \hat{y}_{Layer} bzw. \hat{y}_{Klasse} zugewiesen. Die Zuordnung erfolgt allerdings nur dann, wenn die höchste Gewichtung größer als ein bestimmter Schwellenwert δ_{knn} und die Distanz zwischen POI-Punkt und Fahrtende-GPS-Punkt geringer als ein bestimmter Schwellenwert d_{knn} ist. Bei Nichterfüllung der Kriterien entspricht die Fahrt dem Fahrtzweck „sonstiges“. Da nur Fahrtzwecke mit einer hohen Wahrscheinlichkeit und einer bestimmten Distanz zwischen POI-Punkt und Fahrtende-GPS-Punkt verwendet werden, wird die willkürliche Zuordnung von Fahrtzwecken zu Fahrten minimiert. Die zentrale Annahme bei diesem Ansatz ist, dass das Ziel der Fahrt sich auch am Fahrtende-GPS-Punkt befindet.

Im Folgenden werden die bestimmten Fahrtzwecke zusammengefasst und den aus HTS erhobenen Fahrtzwecken (Tabelle 2.1) gegenübergestellt. Die Fahrtzwecke „nach Hause“ und „zur Arbeit“ werden über die Routine bestimmt. Die Fahrtzwecke „Einkaufen“, „Freizeit“, „Essen“ sowie „Erledigungen“ können über die Verknüpfung mit POI-GPS-Punkten und die beschriebene kNN-Methode approximiert werden. Durch die Verknüpfung der POI-GPS-Punkte und der in ihnen enthaltenen Informationen zu Layern und Klassen ist es möglich, beliebige Fahrtzwecke (z. B. Besuch beim Arzt, Übernachtung, Tourismus etc.) und deren Abhängigkeiten zu den Dimensionen Zeit und Distanz zu untersuchen. Hier sind dem Detaillierungsgrad nahezu keine Grenzen gesetzt. Beispielsweise sind beim MOP und der MiD Restaurantbesuche in Freizeitaktivitäten enthalten. Dieser Aspekt ist mit der

vorgestellten Methodik differenziert beobachtbar. Zusätzlich ist es potenziell realisierbar, bei Freizeitaktivitäten zwischen eintägigen und mehrtägigen Reisen zu unterscheiden. Des Weiteren sind bei der MiD und U.S. dem NHTS im Fragebogen keine Informationen zu Tank- und Ladevorgängen enthalten. Das MOP fragt zwar über die Tankbucherhebung Tankvorgängen ab, allerdings nicht Ladevorgänge. Bei Sensordaten ist eine Beobachtung von öffentlichen Ladevorgängen theoretisch möglich. Die Kriterien für die Fahrtzwecke „zur Arbeit“ und „Ausbildung“ sind sich sehr ähnlich, wodurch eine Differenzierung nur eingeschränkt möglich ist. Aus diesem Grund werden diese Fahrtzwecke bei den Sensordaten im Fahrtzweck „zur Arbeit“ zusammengefasst. Die Fahrtzwecke „dienstlich“ und „Holen/Bringen“ folgen weder einer Routine, noch können sie über POI-GPS-Punkte ermittelt werden. Deshalb ist derzeit keine Approximation dieser Fahrtzwecke möglich.

Ein weiterer Vorteil bei der Approximation der Fahrtzwecke gegenüber den betrachteten Erhebungen ist, dass die Definition der Fahrtzwecke für alle Fahrzeuge identisch ist, da sie nicht von der Interpretation der Probanden abhängt. Beispielsweise könnte eine Person Einkaufen/Shopping als „Freizeit“ deklarieren, während eine andere Person diese Aktivität als „Einkaufen“ interpretiert.

In Tabelle 4.2 sind die verwendeten Hyperparameter aufgelistet. Insgesamt werden für die Approximation der Fahrtzwecke durch routinierte Fahrten acht und für die durch nicht routinierte Fahrten drei Hyperparameter benötigt. Zudem sind in Tabelle 4.2 die initialen Werte für die verwendeten Hyperparameter angegeben. Je nach Anwendungsfall können die Parameter beliebig angepasst werden.

Tabelle 4.2: Hyperparameter zur Approximation des Fahrtzwecks

Variablen	Beschreibung	Initialisierung
Routinierte Fahrten		
δ_{time}	Standzeit für das Zusammenführen von Fahrten	5 min
δ_{radius}^h	Radius für den Fahrtzweck „nach Hause“	1 km
δ_{radius}^a	Aktionsradien eines Fahrzeugs	1–10 km, 11–30 km, 31–100 km, > 100 km
δ_{dbscan}	Radius für die wichtigen Standorte	100 m
$MinPts$	Anzahl Fahrten für die wichtigen Standorte	0,1 · Anzahl Fahrten
δ_{time}^w	Standzeit am Arbeitsplatz	3 Stunden
$MinPts^w$	Anzahl Fahrten am Arbeitsplatz	2-mal in der Woche
δ_{radius}^w	Radius Fahrtzweck „Arbeit“	1 km
Nicht routinierte Fahrten		
k^{knn}	Anzahl der k nächsten Nachbarn	5
δ_{knn}	Schwellenwert für Zuweisung	0,45
d_{knn}	Distanz zu POI mit höchster Gewichtung	1 km

Zur Anpassung der Methodik und zur Festlegung der initialen Werte wurden zum einen Richtwerte aus der Literatur verwendet und zum anderen ein zweiwöchiges Pkw-Wegetagebuch von acht freiwilligen Personen geführt. Diese gaben jeweils am Anfang und Ende der Pkw-Fahrt Informationen zur GPS-Position, dem Kilometerstand und dem Zeitstempel plus den dazugehörigen Fahrtzweck an. Dabei wurde der Fahrtzweck über die beschriebene Methodik approximiert und mit dem angegebenen Fahrtzweck aus dem Wegetagebuch verglichen. Stimmten der approximierten und der angegebene Fahrtzweck nicht überein, wurden die initialen Parametereinstellungen dementsprechend angepasst, um die tatsächlichen Fahrtzwecke so gut wie möglich zu repräsentieren.

Schätzung des Fahrtzwecks

Die folgenden Ausführungen in diesem Abschnitt sowie die dargestellte Literaturübersicht zur Identifikation von Fahrtzwecken basiert größtenteils auf der Veröffentlichung „Identifying Trip Purposes on Trip Level for Vehicle Sensor Data“ (Niklas, Magdolen, von Behren & Vortisch, 2021).

Ein Ansatz zur Identifikation von Fahrtzwecken und zur zusätzlichen Validierung der vorgestellten Methodik ist das Schätzen der Fahrtzwecke auf der Basis von CUMILE. Der verwendete Datensatz aus Kapitel 4 liefert Informationen auf Fahrtebene mit dem jeweiligen Fahrtzweck. Im CUMILE-Datensatz sind insgesamt 4 559 288 Fahrten mit elf Fahrtzwecken (nach Hause

(40,05 %), zur Arbeit (12,76 %), dienstlich (4,09 %), Fahrt zum Kindergarten/Schule/Universität (0,57 %), Einkaufen (18,45 %), Freizeit (13,83 %), Holen/Bringen (8,56 %), andere (0,08 %), Fahrt zurück zum Hotel (1,50 %), Fahrt zum Zweitwohnsitz (0,08 %), Fahrten, dessen Start- und Zielort identitisch ist (0,04 %)) hinterlegt. Damit Fahrtzwecke valide geschätzt werden können, muss eine gewisse Anzahl von Fahrten mit dem jeweiligen Fahrtzweck vorhanden sein. Aus diesem Grund werden im Folgenden nur diejenigen Fahrtzwecke betrachtet, welche mehr als 1,5 % der Fahrten an den Gesamtfahrten darstellen. Für die Modellschätzung wurden alle Fahrten nach den Fahrtzwecken „nach Hause“, „zur Arbeit“, „dienstlich“, „Freizeit“, „Einkaufen“ und „Holen/Bringen“ gefiltert. Dies entspricht insgesamt 4 455 750 Fahrten. Da sich die Regelmäßigkeit bei der Approximation des Fahrtzwecks über das wiederholte Aufsuchen eines Standorts definiert und keine GPS-Informationen im CUMILE-Datensatz enthalten sind, müssen Fahrtzwecke bei diesem Ansatz auf Fahrtebene geschätzt werden. Der geschätzte wird dann mit dem approximierten Fahrtzweck abgeglichen. Stimmt die Vorhersage in beiden Ansätzen überein, spricht dies für die Robustheit der Approximation. Um das Modell auf die Sensordaten anwenden zu können, wird darauf geachtet, dass nur solche Inputvariablen eingesetzt werden, welche auch über die Sensordaten generierbar sind. Für die Bestimmung der Inputvariablen und die Schätzung von Fahrtzwecken auf Fahrtebene ist es in einem ersten Schritt sinnvoll, die Verteilungen der Fahrtzwecke mit verschiedenen kontinuierlichen Variablen (Fahrtdistanz, Startzeit und Standzeit) visuell zu betrachten. Die Darstellung der Fahrtdistanzen (Abbildung 4.4) bringt für die Schätzung des Fahrtzwecks auf Fahrtebene keinen Mehrwert, da sich diese für bestimmte Fahrtzwecke von Fahrzeug zu Fahrzeug unterscheiden kann und keiner übergreifenden Regelmäßigkeit folgt.

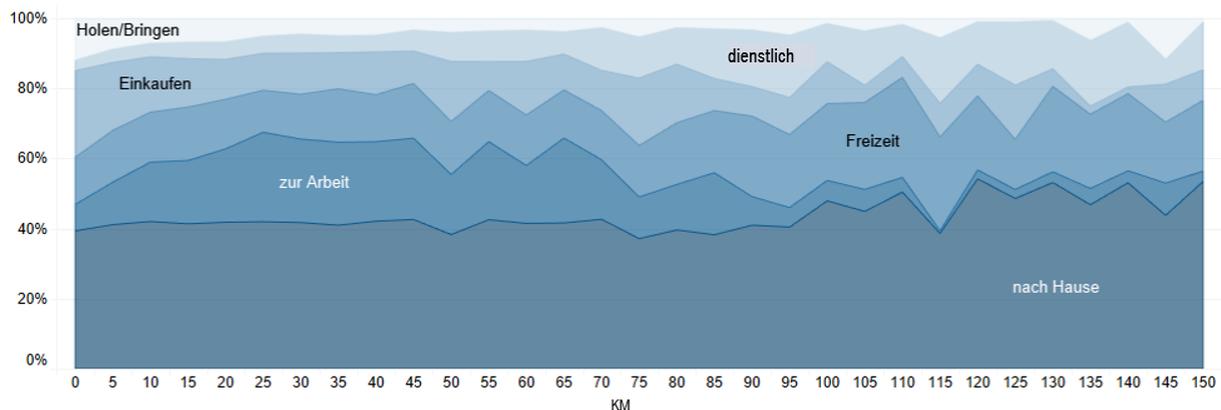


Abbildung 4.4: Fahrtzwecke nach Fahrtdistanz

Abbildung 4.5 zeigt die Verteilung der Fahrtzwecke nach Standzeit, nachdem eine Fahrt mit einem bestimmten Fahrtzweck beendet worden ist. Zur besseren Darstellung werden alle Fahrten mit einer Standzeit größer 1110 Minuten (18,5 Stunden) abgeschnitten. Hier ist Folgendes eindeutig zu erkennen: Je länger die Standzeit nach einer Fahrt andauert, desto

wahrscheinlicher wird der Fahrtzweck „nach Hause“ absolviert. Beträgt die Standzeit nach einer Fahrt zwischen vier und 10,5 Stunden ist „zur Arbeit“ der dominierende Fahrtzweck. Bei sehr kurzen Standzeiten bestimmen „Einkaufen“ und „Holen/Bringen“ das Geschehen. Interessant bei „Einkaufen“ ist, dass bei einer Standzeit von ca. 6,5 Stunden der Anteil noch einmal stark zunimmt. Das könnte theoretisch ein Shoppingtag in einem großen Einkaufszentrum oder einer größeren Stadt darstellen. Dieser Effekt wird nochmals verstärkt, wenn nur Fahrten am Wochenende betrachtet werden. „dienstlich“ findet erst ab einer Standzeit von ca. 30 Minuten statt.

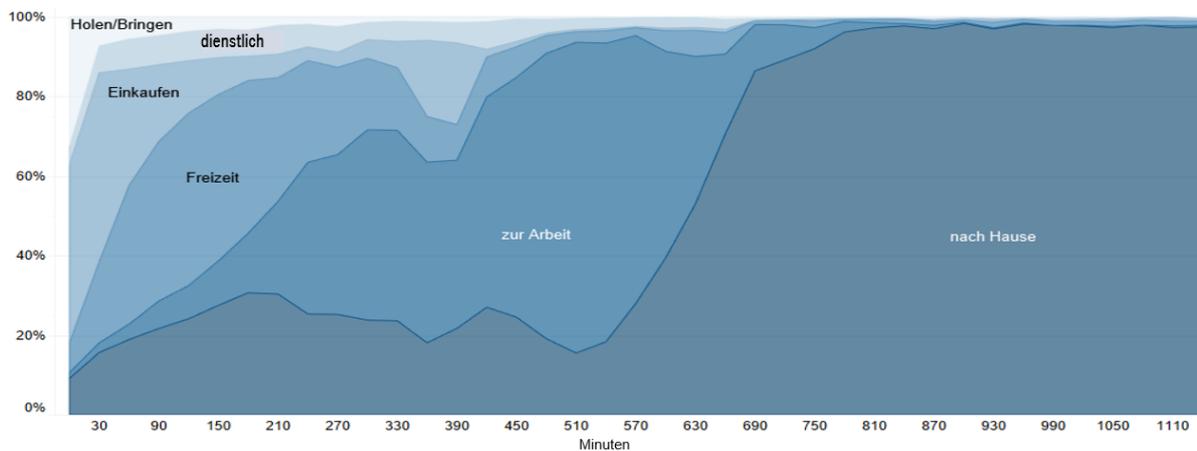


Abbildung 4.5: Fahrtzwecke nach Standzeit

Auch die Verteilung der Fahrtzwecke nach Startzeit (Abbildung 4.6) zeigt Unterschiede auf. Wird eine Fahrt ganz früh am Tag (00:00 Uhr–04:00 Uhr) begonnen, handelt es sich hauptsächlich um „nach Hause“. Je später die Startzeit einer Fahrt an einem Tag, desto eher entspricht es einer Fahrt „nach Hause“. Der Fahrtzweck „zur Arbeit“ dominiert stark in den Morgenstunden zwischen 04:00 Uhr–07:00 Uhr. „Einkaufen“ findet hauptsächlich vormittags (08:00 Uhr–11:00 Uhr) statt. „dienstlich“ hat interessanterweise einen Ausreißer um 03:00 Uhr. Allerdings ist dieser Anteil mit Vorsicht zu genießen, da allgemein sehr wenig Fahrten (1171 Fahrten) um 03:00 Uhr stattfinden. Der Anteil der Fahrten mit dem Fahrtzweck „Holen/Bringen“ ist über die Uhrzeit relativ konstant verteilt.

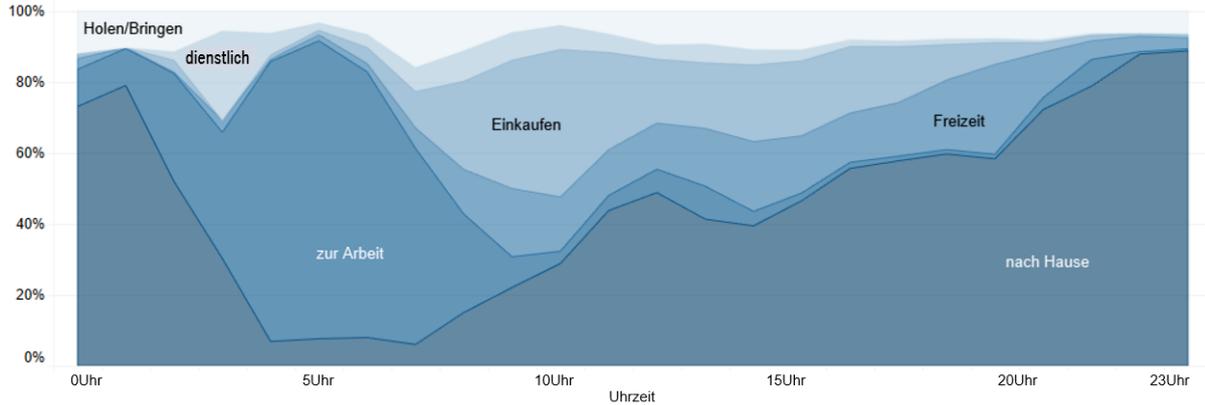


Abbildung 4.6: Fahrtzwecke nach Startzeit

Die vorangegangenen Auswertungen sind für eine erste Einschätzung des Einflusses der Stand- und Startzeit auf die Fahrtzwecke nützlich und werden bei der Auswahl der Eingangsvariablen berücksichtigt. Für die Schätzung wird die bereits angewandte Softmax-Regression (SR) verwendet, weswegen auf die formale Beschreibung des Verfahrens an dieser Stelle verzichtet wird (siehe Unterkapitel 3.2 sowie Niklas, von Behren, Chlond und Vortisch (2020)).

Die Eingangsvariablen beinhalten Informationen zur Standzeit sowie Startzeit. Damit das Verfahren bessere Entscheidungsfunktionen bilden kann, werden die kontinuierlichen Variablen in Intervalle eingeteilt, um binäre Variablen zu generieren. Zur Festlegung der Intervallgrenzen für die Standzeit (0–30 min, 30–60 min, 60–120 min, 120–240 min, 240–360 min, 360–480 min, mehr als 480 min) und die Startzeit (0:00–06:00 Uhr, 06:00–09:00 Uhr, 09:00–12:00 Uhr, 12:00–15:00 Uhr, 15:00–18:00 Uhr, 18:00–21:00 Uhr, 21:00–24:00 Uhr) dienen die Abbildungen 4.5 und 4.6 als Grundlage. Darüber hinaus werden die Informationen einer Fahrt nicht nur zum Zeitpunkt t verwendet, sondern auch zum Zeitpunkt $t - 1$ und $t + 1$. Diese Zusammenhänge haben durchaus eine Relevanz für die Klassifizierung, da im Prozess der Variablenauswahl mehrere Varianten getestet wurden. Somit wird die Startzeit einer Fahrt zu den Zeitpunkten t und $t - 1$, sowie die Standzeit nach einer Fahrt zu den Zeitpunkten t , $t - 1$ und $t + 1$ betrachtet. Darüber hinaus sind noch die Informationen, ob die Fahrt an einem Werktag stattgefunden und ob das Fahrzeug nach der Fahrt übernachtet hat, in das Modell eingeflossen. Nachdem Fahrten mit fehlenden Werten entfernt worden waren, konnten insgesamt 4 398 678 Fahrten mit den genannten Variablen generiert werden.

Der Modellinput entspricht dem gelabelten Datensatz von CUMILE mit Informationen auf Fahrtebene und dem jeweiligen modellierten Fahrtzweck $C^{(mp)} = \{(\mathbf{x}_i^{(mp)}, y_i^{(mp)})\}_{i=1}^{4\,398\,678}$ mit $\mathbf{x}_i^{(mp)} = (x_{1;i}^{(mp)}, \dots, x_{37;i}^{(mp)}) \in \mathbb{R}^{37}$ und $y_i^{(mp)} = (y_{1;i}^{(mp)}, \dots, y_{s;i}^{(mp)}) \in \{1, 2, \dots, 6\}$. Im Optimierungsschritt wird wieder die 6×37 -Matrix der Parameter Θ so angepasst, dass die Verlustfunktion $J(\Theta)$ minimal ist. Im Ergebnis entsteht dann eine Funktion $y_i^{(mp)} = f(\mathbf{x}_i^{(mp)})$, welche eine Fahrt i dem jeweiligen Fahrtzweck $y_i^{(mp)}$ zuordnen kann. Diese Funktion $f(\cdot)$ wird

auf die Sensordaten mit den gleichen Eingangsvariablen adaptiert und den Fahrten j wird jeweils ein Fahrtzweck y_j^p probabilistisch zugeordnet. Im Endergebnis entsteht ein gelabelter Sensordatensatz auf Fahrtebene mit den geschätzten Fahrtzwecken $S^{(p)} = \{(\mathbf{x}_j^{(p)}, y_j^{(p)})\}_{j=1}^m$. Die Validierung des Modells wird ähnlich wie in Abschnitt 3.1.4 über eine Konfusionsmatrix bestimmt (Tabelle 4.3). Durch die Konfusionsmatrix können die „Precision“ und der „Recall“ bestimmt werden. Die „Precision“ gibt an, wie viele der prognostizierten Fahrtzwecke mit dem tatsächlichen Fahrtzweck übereinstimmen. Beispielsweise sind 86 % des geschätzten Fahrtzwecks „nach Hause“ auch tatsächlich der Fahrtzweck. Der „Recall“ gibt an, wie viele der tatsächlichen Fahrtzwecke identisch zum geschätzten Fahrtzweck sind. Beispielsweise entsprechen 93 % der tatsächlichen Fahrtzwecke „nach Hause“ auch den prognostizierten Fahrtzwecken. Der „F1-Score“ ist eine Kombination aus beiden Metriken, welche über das harmonische Mittel $2 \cdot (Precision \cdot Recall) / (Precision + Recall)$ bestimmt wird. Das harmonische Mittel führt dazu, dass niedrige Werte stärker gewichtet werden. Dies macht sich vor allem bei der Identifikation des Fahrtzwecks „dienstlich“ bemerkbar. Während die „Precision“ 44 % entspricht, hat „Recall“ nur einen Wert von 8 %. In Konsequenz hat der F1-Score lediglich einen Wert von 13 %. Die Spalte „Anzahl“ gibt die Stichprobengröße für die jeweiligen Klassen an. Zur Validierung des Modells wurden die 4 398 678 Fahrten in einen Trainingsdatensatz (67 %) und einen Testdatensatz (33 %) aufgeteilt. Aus diesem Grund berücksichtigt die Tabelle 4.3 zur Validierung des Modells nur die Stichprobe des Testdatensatzes (1 451 564). Die „Precision“ bewertet das Modell in seiner Gesamtheit. 73 % aller prognostizierten Fahrtzwecke sind mit dem tatsächlichen Fahrtzweck gleich. Der „Marko-Durchschnitt“ repräsentiert den Durchschnitt der jeweiligen Metriken. Beim „gewichteten Durchschnitt“ werden die Metriken um die Anzahl der tatsächlichen Fahrtzwecke gewichtet (Géron, 2019).

Insgesamt lässt sich sagen, dass „nach Hause“ (0,90) und „zur Arbeit“ (0,80) gut identifizierbar sind. Auch bei „Einkaufen“ (0,61) und „Freizeit“ (0,59) kann immer noch über die Hälfte aller Fahrtzwecke richtig zugeordnet werden. Bei „dienstlich“ (0,13) und „Holen/Bringen“ (0,37) ist die Genauigkeit unterdurchschnittlich. Dies hat sich bereits bei der Verteilung des Fahrtzwecks nach Stand- und Startzeit (Abbildungen 4.5 und 4.6) angedeutet, da diese keine Besonderheiten aufweisen konnten. Auch rein logisch betrachtet sind „dienstlich“ und „Holen/Bringen“ sehr spezielle Fahrtzwecke, weshalb es schwierig ist über alle Fahrzeuge hinweg Gemeinsamkeiten auf Fahrtebene zu identifizieren. Vor dem Hintergrund, dass Regelmäßigkeiten in der Fahrzeugnutzung bewusst nicht berücksichtigt werden und das berechnete Modell nur auf Fahrtinformationen zu den Zeitpunkten t , $t - 1$ und $t + 1$ basiert, ist die Präzision des Gesamtmodells (73 %) durchaus akzeptabel.

Tabelle 4.3: Validierung des Modells zur Fahrtzweckschätzung

Fahrtzweck	Precision	Recall	F1-Score	Anzahl
Nach Hause	0,86	0,93	0,90	600 599
zur Arbeit	0,80	0,80	0,80	191 847
Dienstlich	0,44	0,08	0,13	60 551
Einkaufen	0,55	0,67	0,61	275 147
Freizeit	0,61	0,58	0,59	207 565
Holen/Bringen	0,47	0,30	0,37	115 855
Präzision			73 %	1 451 564
Makro-Durchschnitt	0,62	0,56	57 %	1 451 564
Gewichteter Durchschnitt	0,71	0,73	71 %	1 451 564

Das Modell mit den berechneten Parametern Θ kann nun abgespeichert und für die Approximation der Fahrtzwecke zur Validierung verwendet werden.

Validierung der Fahrtzweckapproximation

Im Folgenden können nun die approximierten mit den geschätzten Fahrtzwecken validiert werden. Wie bereits beschrieben, werden nicht routinierte Fahrten über POI-Informationen einem Fahrtzweck zugeordnet. Der Fahrtzweck ist somit vom jeweiligen POI-Layer bzw. POI-Klasse abhängig. Für die Validierung werden die Fahrtzwecke so zugeordnet, dass sie auch mit den Fahrtzwecken aus CUMILE vergleichbar sind. Alle Fahrten, welche den Layern „Einkaufen“, „Geld“, „Gesundheit“, „Tanken“ und „Laden“ zugerechnet werden, entsprechen dem Fahrtzweck „Einkaufen“. Alle Fahrten, welche in den Layer „Freizeit“, „Unterkunft“, „Restaurant“, „öffentlich“ sowie „Tourismus“ einsortiert werden, entsprechen dem Fahrtzweck „Freizeit“. Im nächsten Schritt kann das berechnete Modell mit den Parametern Θ verwendet werden, um die Fahrtzwecke der Sensordaten zu prognostizieren. Die geschätzten Fahrtzwecke werden im nächsten Schritt mit den approximierten Fahrtzwecken und der Wahrscheinlichkeit der Zuordnung verglichen. Die konkrete Hypothese ist nun folgende: Je größer die Wahrscheinlichkeit der Zuordnung ist, desto eher stimmen die approximierten mit den geschätzten Fahrtzwecken überein.

Abbildung 4.7 stellt die Präzision der Schätzung (approximierter Fahrtzweck stimmt mit dem geschätzten Fahrtzweck überein) entlang der Wahrscheinlichkeit der Zuordnung dar. Die Balken zeigen die durchschnittliche Präzision entlang der Wahrscheinlichkeit der Zuteilung. Beispielsweise stimmen in 81 % der Fälle die approximierten mit den geschätzten Fahrtzwecken überein, wenn die Wahrscheinlichkeit größer 99 % ist. Die orange Linie repräsentiert die kumulierte Summe der Fahrten entlang der Zuordnungswahrscheinlichkeit. Die Grafik legt dar,

dass der approximierten mit dem geschätzten Fahrtzweck umso eher übereinstimmt, je größer die Wahrscheinlichkeit der Zuordnung ist. Somit stellen die approximierten Fahrtzwecke durchaus robuste Ergebnisse dar.

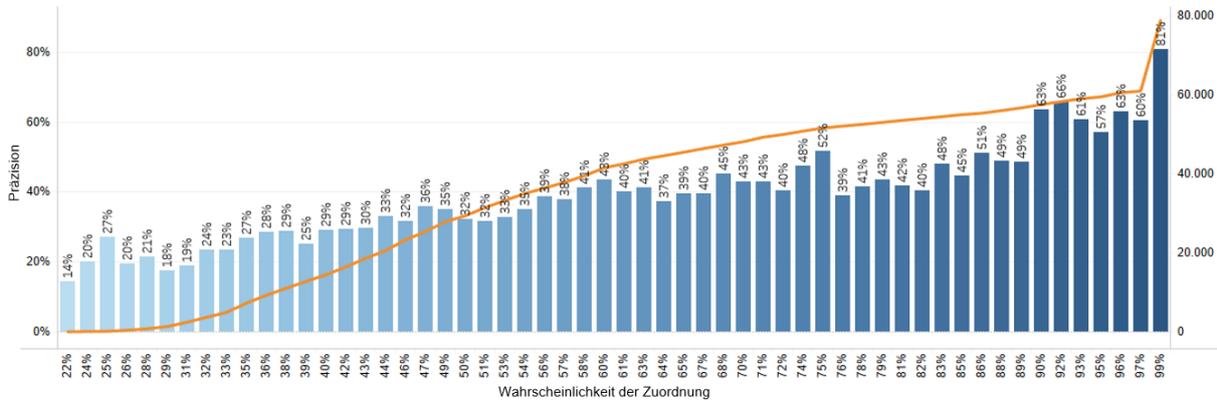


Abbildung 4.7: Präzision entlang der Wahrscheinlichkeit der Zuordnung

Insgesamt wurden die Fahrtzwecke von 78 995 Fahrten von 625 Fahrzeugen in Deutschland über einen Beobachtungszeitraum von ca. einem Monat approximiert. Dabei wurden 37 % dem Fahrtzweck „nach Hause“, 9 % „Arbeit“, 24 % „Einkaufen“ und 28 % dem Fahrtzweck „Freizeit“ zugeordnet. Lediglich 2 % der Fahrten konnten keinem Fahrtzweck zugeordnet werden. Bei einem Vergleich der Anteile mit den Ergebnissen aus verschiedenen HTS, ergibt sich ein stimmiges Bild. Bei der MiD entspricht die Aufteilung der Pkw-Verkehrsleistung auf Fahrtzwecke zu 24 % der Fahrten dem Fahrtzweck „Freizeit“ und 21 % (13 % + 8 %) dem Fahrtzweck „Einkauf“ und „Erledigung“. 28 % bzw. 21 % der Fahrten entsprechen dem Fahrtzweck „Arbeit“ bzw. „dienstlich“. Bei der MiD werden alle Rückwege und alle Wege nach Hause dem Zweck des zuvor genannten Weges zugeordnet (Nobis & Kuhnimhof, 2018). Bei der Approximation der Fahrtzwecke wird der Fahrtzweck „nach Hause“ aber separat ausgewiesen, da er gerade für das Laden zu Hause relevant ist. Um die Werte vergleichbar zu machen, müssen die MiD-Werte durch zwei dividiert werden. Das MOP betrachtet die Verteilung der Verkehrsleistung nach Wegezwecken. Dabei sind 45 % der berichteten Wege dem Zweck „nach Hause“, 13 % dem Zweck „Arbeit, dienstlich oder geschäftlich“, 26 % (18 % + 8 %) dem Zweck „Besorgung und Service“ oder „sonstige private Erledigungen“ sowie 13 % dem Fahrtzweck „Freizeit“ zuzuordnen (Ecke et al., 2019). Der Anteil bei „Freizeit“ ist bei der Approximation im Vergleich zum MOP um 15 % Prozentpunkte höher. Diese Differenz könnte sich durch die speziellen Eigenschaften der Sensordaten erklären. Bei Sensordaten wird, auch bei geringen Fahrdistanzen, der Fahrt ein Fahrtzweck zugewiesen, während solche kurzen Fahrten bei HTS möglicherweise gar nicht auftauchen. Bei der amerikanischen U.S. NHTS verteilt sich die jährliche Anzahl an individuellen Wegen mit dem Fahrzeug mit 20 % auf „von und zur Arbeit“, mit 4 % auf „dienstlich“, mit 21 % auf „Einkaufen“, mit 15 % auf „Schule/Kirche“, mit 35 %

auf „Erholung/Soziales“ sowie mit 6 % auf „andere“. Auch hier sind Rückwege von der Arbeit mit dem Arbeitszweck gleichzusetzen. Es sollte noch angemerkt werden, dass die Verteilungen nicht direkt vergleichbar sind, da HTS sie auf Individualebene ausweisen, während Sensordaten Fahrtzwecke aus der Fahrzeugperspektive betrachten. Insgesamt sind die Verteilungen der Fahrtzwecke aber ähnlich und keine auffälligen Ausreißer erkennbar. Darüber hinaus ist bereits im Vergleich zwischen den traditionellen HTS ersichtlich, dass die Verteilungen der Fahrtzwecke nicht exakt übereinstimmen.

4.1.5 Limitationen

Die Approximation von Fahrtzwecken ist mit gewissen Limitationen und Annahmen verbunden, die im Folgenden betrachtet werden. Aus diesem Grund wird auch bewusst von einer Approximation und nicht von einer Identifikation der Fahrtzwecke gesprochen.

- Wie in Tabelle 4.2 dargestellt, ist eine Festlegung verschiedener Kriterien zur Approximation der Fahrtzwecke erforderlich. Entspricht der Zusammenhang zwischen Fahrzeugnutzung und Fahrtzweck nicht den definierten Kriterien, kann es zu Fehlklassifikationen kommen. Um dem entgegenzuwirken, wurden, so weit wie möglich, die regelbasierten Annahmen anhand von Experteneinschätzungen, bestehender Literatur und durch die zweiwöchigen Wegetagebücher festgelegt. Darüber hinaus wurden mit Hilfe der Kriterien zur Stand- und Tageszeit lediglich der durchschnittliche Wohn- und Arbeitsort geografisch bestimmt. Durch das Ziehen von Radien können die definierten Stand- und Tageszeiten zur Approximation des Fahrtzwecks wieder vernachlässigt werden. Des Weiteren stellen die initialen Werte der Hyperparameter keine harten Kriterien dar. Die Herangehensweise zur Approximation der Fahrtzwecke wurde allgemein beschrieben, sodass die Hyperparameter je nach Anwendungsfall, subjektiver Einschätzung und allgemeinen Veränderungen der Fahrzeugnutzung über die Zeit angepasst werden können.
- Die Approximation der Fahrtzwecke anhand der Verknüpfung mit GPS-POI-Punkten ist mit Annahmen verbunden. Die stärkste Annahme ist, dass der GPS-Punkt am Ende einer Fahrt in unmittelbarer Nähe zum Ziel und damit zum Fahrtzweck liegen muss. Speziell in urbanen Räumen mit Parkplatzproblemen muss der GPS-Punkt am Ende einer Fahrt nicht direkt am Zielpunkt liegen, da die örtliche Flexibilität eines Parkplatzes oftmals nicht vorhanden ist. Des Weiteren können sich sehr viele POIs auf einer kleinen Fläche sammeln, was die eindeutige Zuordnung erschwert. Durch die Einteilung der POIs in Layer und Klassen ist es möglich, verschiedene POIs zusammenzufassen. Für die zugrundeliegende Fragestellung ist es weniger relevant, welche Art von Restaurant ein Individuum angesteuert hat. Trotzdem kann es dazu führen, dass die Approximation der Fahrtzwecke durch die Verknüpfung mit GPS-POI-Punkten von Fahrzeugen, welche

sich hauptsächlich in ländlichen Räumen bewegen, eher funktioniert als mit Fahrzeugen, welche sich hauptsächlich in urbanen Räumen bewegen. In Konsequenz können Fahrten in urbanen Räumen weniger einem Fahrtzweck zugeordnet werden. Dies würde wiederum das Clustering zur Identifikation von dreidimensionalen FNP's beeinflussen. Aus diesem Grund werden für die Bestimmung von dreidimensionalen FNP's Fahrtzwecke, welche über die Verknüpfung von GPS-POI-Punkten approximiert wurden, in „Freizeit“ und „Einkaufen“ zusammengefasst. Dadurch ist eine Validierung der Ergebnisse mit Fahrtzweckverteilungen aus HTS möglich.

- Bei der Methodik zur Approximation der Fahrtzwecke werden nur das Fahrzeug und dessen Nutzung betrachtet. Intermodale Mobilität kann gegenwärtig nicht beobachtet werden. Dieser Umstand bringt gewisse Einschränkungen mit sich. Beispielsweise könnte ein Individuum ein Fahrzeug regelmäßig für Fahrten zum Bahnhof einsetzen, um von dort aus mit der Bahn zum Arbeitsplatz zu gelangen. Da das Fahrzeug regelmäßig mit einer gewissen Standzeit am Bahnhof stünde, würde der Bahnhof die Kriterien des Arbeitsortes erfüllen. Natürlich könnte der bestehende Ansatz weiterentwickelt werden, indem überprüft wird, ob sich bei einem wichtigen Standort ein Bahnhof, eine Haltestelle oder ein Parkplatz (Park and Ride) in der Nähe befindet. Allerdings wäre auch dieses Prozedere mit starken Annahmen verbunden, zumal der Fahrzeugnutzer theoretisch auch im Bahnhof arbeiten bzw. der Arbeitsort sich in der Nähe des Bahnhofes befinden kann. Darüber hinaus ist es möglicherweise für die ganzheitliche Betrachtung der Fahrzeugnutzung und die Ableitung von strategischen Potenzialen weniger relevant, ob es sich bei dem wichtigen Standort um den tatsächlichen Arbeitsort handelt. Vielmehr steht die Identifikation von regelmäßigen Pendelstrecken eines Fahrzeugs im Zentrum der Untersuchung.
- Ein allgemeines Problem bei der Approximation des Fahrtzwecks ist, dass, wie auch in vielen anderen Studien, keine Grundwahrheit existiert (Shen & Stopher, 2014). Somit kann die Präzision der Zuordnung von Fahrtzwecken zu Fahrten nur schwer überprüft werden. Durch die zweiwöchigen Wegetagebücher wurden zwar die Logiken geprüft und geschärft, allerdings ist die Stichprobe zu klein und nicht repräsentativ für eine robuste Aussage zur Präzision. Um diesem Problem weiter entgegenzuwirken, wurde der CUMILE-Datensatz für die Validierung der approximierten Fahrtzwecke verwendet.

4.2 Bestimmung von Zeit-Distanz-Fahrtzweck-Fahrzeugnutzungsprofilen

Im folgenden Unterkapitel werden dreidimensionale FNP's auf der Basis von Zeit, Distanz und Fahrtzweck erstellt und interpretiert. Vor der tatsächlichen Erstellung der FNP's wird ein Überblick gewährt, mit welchen Verfahren existierende Literatur ein ähnliches Verhalten bzw. eine ähnliche Nutzung in Gruppen einteilt. Danach werden die für das Clustering relevanten Daten inhaltlich vorgestellt. Im nächsten Schritt wird das Vorgehen zur Bestimmung dreidimensionaler FNP's präsentiert. Abschließend finden die Beschreibung und Interpretation der berechneten Cluster statt.

4.2.1 Literaturübersicht

Es gibt eine Vielzahl von Arbeiten, welche sich damit beschäftigen, Individuen oder auch Fahrzeuge auf der Basis ihres Verhaltens oder ihrer Nutzung in Gruppen einzuteilen. Hintergrund ist die Annahme, dass ähnliche Gruppen auch ähnliche Anforderungen an ihre Mobilität oder ihr Fahrzeug haben. Ziel bei allen Studien dieser Art ist es, die Homogenität innerhalb der Gruppen und die Heterogenität zwischen den Gruppen zu maximieren. Für die Identifikation der Cluster muss zum einen die Anzahl der Gruppen, unter Berücksichtigung der vorhandenen Daten, bestimmt und zum anderen die Objekte (Individuen/Fahrzeuge) in entsprechende Gruppen eingeteilt werden. Einige Verfahren (z. B. DBSCAN) benötigen nicht die Angabe einer spezifischen Anzahl von Gruppen. Allerdings müssen dann andere Kriterien (z. B. Mindestanzahl von Objekten zur Bildung einer Gruppe) festgelegt werden. Zur Identifikation von ähnlichen Nutzergruppen gibt es eine Vielzahl unüberwachter Verfahren, die zum Einsatz kommen können. Sind die Nutzergruppen bestimmt, kann mit verschiedenen Metriken die Qualität der Einteilung gemessen werden.

Eine der verbreitetsten Verfahren zur Identifikation von Nutzergruppen in der Mobilitätsforschung ist der k-Means-Algorithmus. Um die Anzahl der Cluster zu bestimmen, nutzen Eisenmann und Buehler (2018) und Magdolen et al. (2019) eine Varianzanalyse (Ward-Methode und Cubic-Cluster-Kriterium). Die Anzahl der Cluster fließt wiederum bei beiden Arbeiten als Input in den k-Means-Algorithmus ein. Haustein und Nielsen (2016) nutzen zwar auch ein k-Means-Clustering, allerdings stellen sie bei der Bestimmung der Anzahl der Cluster die Interpretierbarkeit in den Vordergrund. Sie teilen die Objekte in Gruppen ein und variieren dabei die Anzahl der Cluster von vier bis neun. Die Lösung mit den plausibelsten Clustern wird dann verwendet. Die Qualität der Cluster wird anschließend über einen χ^2 -Test und eine Varianzanalyse geprüft. Morton, Anable und Nelson (2017) berücksichtigen zur Qualitätssicherung das Varianz-Ratio-Kriterium. Auch Sodenkamp et al. (2019) und Olafsson, Nielsen und Carstensen (2016) wenden das k-Means-Verfahren an. Während Sodenkamp et

al. (2019) keine Aussagen zur Bestimmung der Anzahl der Cluster und der Qualität der Einteilung machen, darf bei Olafsson et al. (2016) die Clustergröße nicht kleiner als 100 sein. Bevor Wittwer (2014) die Clusteranalyse über ein Dendrogramm und k-Means-Algorithmus durchführt, nutzt er eine Hauptkomponentenanalyse (PCA) um die Dimensionen zu reduzieren. Lanzendorf (2002) nutzt für die Bestimmung von Mobilitätsstilen den k-Means-Algorithmus und eine Faktorenanalyse, um die latenten Variablen der Individuen zu Einstellung, Werten und Orientierung zu identifizieren.

Im Vergleich zur bisher genannten Literatur verwenden Molin, Mokhtarian und Kroesen (2016) ein latentes Klassenmodell zur Clusteranalyse. Im Gegensatz zum k-Means-Algorithmus können hier, ähnlich wie bei der SR aus Unterkapitel 3.2, die Objekte probabilistisch zugeordnet werden. Das Modell wird für eine unterschiedliche Anzahl von Clustern mehrmals berechnet und mit Hilfe des bayesianischen Informationskriteriums (BIC) werden die Clusterlösungen verglichen. Auch Alonso-González, Hoogendoorn-Lanser, van Oort, Cats und Hoogendoorn (2020) nutzen ein latentes Klassenmodell, um die Einstellungen zu Mobilität als Dienstleistung zu untersuchen. Auch hier wird das Modell für verschiedene Clusterlösungen berechnet und mit Hilfe von Akaike-Informationskriteriums (AIC) und des BIC die beste Clusterlösung bestimmt. Um die Anzahl der Variablen zu reduzieren und latente Strukturen zu identifizieren, haben Alonso-González et al. (2020) vor dem Clustering eine explorative Faktorenanalyse durchgeführt.

4.2.2 Daten

Infolge der Approximation der Fahrtzwecke kann die Tabelle 4.1 um weitere Informationen ergänzt werden, welche in der Tabelle 4.4 aufgeführt sind. Die „Anzahl der zusammengeführten Fahrten“ zählt die Fahrten, zwischen denen eine Standzeit geringer als ein bestimmter Schwellenwert δ_{time} ist. Die „Standzeit zusammengeführter Fahrten“ summiert die Standzeiten zwischen den zusammengeführten Fahrten auf, damit keine Information verloren geht. Die „Fahrdistanz“ und „Fahrtzeit“ berechnen sich jeweils aus der Differenz des Kilometerstandes bzw. Zeitstempels am Ende und Anfang der Fahrt. Durch die Zusammenführung von Fahrten ist eine Unterscheidung zwischen „Fahrtzeit (brutto)“ und „Fahrtzeit (netto)“ möglich, welche sich um die Standzeit der zusammengeführten Fahrten unterscheiden. Der „Fahrtzweck“ bestimmt sich über die beschriebene Herangehensweise zur Approximation der Fahrtzwecke und hat fünf Ausprägungen (nach Hause, zur Arbeit, Einkaufen, Freizeit, Sonstiges). Die „Entfernung zum Wohnort“ entspricht der euklidischen Distanz zwischen Wohnort ($home_{lat}, home_{lng}$) und den GPS-Punkt am Ende einer Fahrt (lat_{end}, lng_{end}). Gerade die Entfernung zum Wohnort und damit die Bestimmung des Aktionsradius eines Fahrzeugs ist für viele Fragestellungen von Relevanz. Beispielsweise kann mit Hilfe dieser Information eine Aussage

getroffen werden, inwieweit es möglich ist, das Fahrzeug über Nacht zu Hause zu laden. Die Variable $tour_{id}$ ordnet Fahrten bestimmten Touren zu. Daraus kann wiederum die durchschnittliche Anzahl der Fahrten pro Tour berechnet werden.

Tabelle 4.4: *Zusätzlich generierte Fahrtinformationen*

Variable	Beschreibung
$paused^{count}$	Anzahl zusammengeführter Fahrten
$paused^{time}$	Standzeit zusammengeführter Fahrten
$time^{end} - time^{start}$	Fahrtzeit (brutto)
$time^{end} - time^{start} - time^{paused}$	Fahrtzeit (netto)
$purpose$	Fahrtzweck
$homedistance$	Entfernung zum Wohnort
$tour_{id}$	Tour-ID

Die Fahrttabelle wird dann zu einer Fahrzeugtabelle transformiert, indem die Fahrten für jedes Fahrzeug nach bestimmten Kriterien gruppiert werden. Die Fahrzeugtabelle (Tabelle 4.5) enthält generelle Informationen zum **Fahrzeug**, welche über die Approximation des Fahrtzwecks generiert wurden (Wohnort, Arbeitsort, Anzahl der wichtigen Standorte, durchschnittliche Pendeldistanz sowie durchschnittliche Fahrten pro Tour). Die Variablen in den Dimensionen Zeit, Fahrtzweck und Distanz stellen das Variablenset für die darauffolgende Clusteranalyse dar. Alle Variablen werden relativ dargestellt, was zum einen die Interpretierbarkeit der Variablen erleichtert und zum anderen eine Standardisierung der Daten redundant macht. Mit Hilfe der Information zur Entfernung zum Wohnort lassen sich fahrzeugspezifisch Radien um den Wohnort ziehen. Dadurch kann der Aktionsradius eines Fahrzeugs ermittelt werden. Die Intervalle der Aktionsradien decken die Kurzstrecken (1–10 km), Mittelstrecken (11–30 km, 31–100 km) sowie Langstrecken (> 100 km) ab. Die Wahl der Aktionsradien orientiert sich an der möglichen Wahl des Verkehrsmittels. Für die Kurzstrecken sind nicht motorisierte Verkehrsmittel (z. B. Fahrrad) eine gute Alternative zum Pkw. Bei den Mittelstrecken bieten sich, vor allem in urbanen Räumen, der ÖPNV und die Bahn an, während für die Langstrecken über 100 km auch der Fernverkehr auf der Schiene oder das Flugzeug Alternativen zum Pkw sind.

Die Betrachtung von Aktionsradien bietet zudem eine neue Interpretation des Fernverkehrs, da hier der Fernverkehr in Abhängigkeit vom Wohnort definiert wird und nicht mehr über die zurückgelegte Distanz. Abbildung 4.8 zeigt zum Verständnis den Zusammenhang zwischen Aktionsradius und approximiertem Wohnort anhand des Fahrzeugs, welches bereits in Abbildung 4.3 dargestellt wurde. Für jedes Fahrzeug wird individuell von seinem approximierten Wohnort ausgehend der Aktionsradius bestimmt.

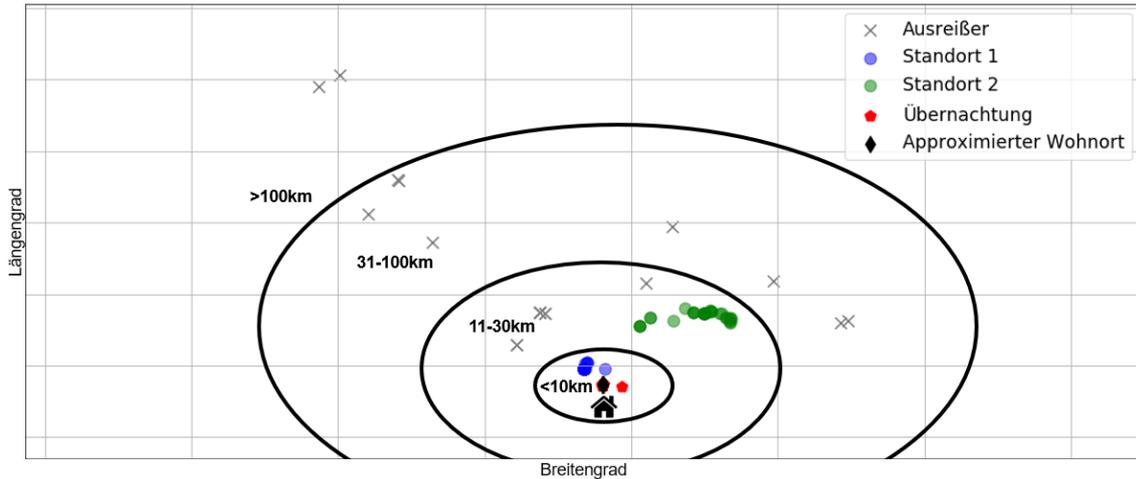


Abbildung 4.8: Schematische Darstellung des Aktionsradius eines Fahrzeugs

Die Dimension **Zeit** berücksichtigt verschiedene Variablen. Die Variable „Fahrten am Werktag“, wird verwendet, um die differenzierte Nutzung von Fahrzeugen an Werktagen und am Wochenende zu beobachten. Dieser Faktor kann Hinweise liefern, inwieweit Fahrzeuge für Pflichtfahrten eingesetzt werden. Der Anteil der „Fahrzeugübernachtungen zu Hause“ ist speziell für die Ladethematik relevant. Diese Variable liefert Rückschlüsse darauf, ob es möglich ist, ein Fahrzeug über Nacht zu Hause zu laden, vorausgesetzt Lademöglichkeiten sind am Wohnort vorhanden. Die Variable „Tage mit Fahrzeugbewegungen“ liefert Informationen darüber, wie intensiv ein Fahrzeug eingesetzt wird. Bei der Betrachtung der „Standzeiten nach Aktionsradius“ ist es möglich, den Bezug zum Wohnort eines Fahrzeugs über die jeweiligen Aktionsradien (1-10 km, 11-30 km, 31-100 km, >100 km) zu untersuchen. Steht ein Fahrzeug hauptsächlich nahe zum Wohnort, könnte es langfristig auch durch andere Verkehrsmittel ersetzt werden.

Die Dimension **Fahrtzweck** beinhaltet die Anteile der „Fahrten nach Fahrtzwecken“ (nach Hause, zur Arbeit, Einkaufen, Freizeit, Sonstiges). Um die Vergleichbarkeit mit HTS zu gewährleisten, wurden für die Clusteranalyse die gleichen Fahrtzwecke wie bereits bei der Validierung verwendet. Anhand der Fahrtzwecke ist es möglich zu untersuchen, ob Fahrzeuge eher speziell für bestimmte oder universal für jegliche Fahrtzwecke eingesetzt werden.

In der Dimension **Distanz** werden zum einen die Anteile der „Tagesdistanzen“ und zum anderen die „Fahrtdistanzen von zu Hause“ betrachtet. Durch die Berücksichtigung der Fahrtdistanzen kann die Zusammensetzung einer Tagesdistanz näher untersucht werden. Beispielsweise können hohe Tagesdistanzen über wenige Fahrten mit großen Fahrtdistanzen oder über viele Fahrten mit geringen Fahrtdistanzen von zu Hause erreicht werden. Durch die Betrachtung der Fahrtdistanzen von zu Hause werden wiederum die Abhängigkeit vom Wohnort und die Ladethematik thematisiert. Um die Übersicht über die vielen Variablen zu wahren, wurden auch

hier die definierten Distanzintervalle hinsichtlich Kurz-, Mittel- und Langstrecken (1–10 km, 11–30 km, 31–100 km, > 100 km) sowohl für die Fahrdistanzen als auch für die Tagesdistanzen angenommen.

Tabelle 4.5 stellt die fahrzeugspezifischen Variablen sowie die verwendeten 20 Variablen aus den Dimensionen Zeit, Fahrtzweck und Distanz für das Clustering dar. Zum Verständnis sind in der Tabelle Beispielwerte der einzelnen Inputvariablen ausgewiesen. Im Gegensatz zu HTS steht hier die Fahrzeugnutzung in Abhängigkeit vom Wohnort im Vordergrund, was eine neue Perspektive auf die Fahrzeugnutzung darstellt. Besonders für die Potenziale von Elektromobilität (Laden über Nacht zu Hause) sowie des Ersetzens des Pkw durch andere Verkehrsmittel nimmt der Aktionsradius eine Bedeutung ein. Die Auswahl der Variablen in den verschiedenen Dimensionen ist flexibel gestaltet. Je nach Bedarf und Anwendungsfall können die jeweiligen Intervalle und Fahrtzwecke anders definiert werden.

Tabelle 4.5: *Fahrzeugetabelle*

Dimension	Beschreibung	Beispielwert
Fahrzeug	Wohnort	Längengrad: 48,8508; Breitengrad: 10,4895
	Arbeitsort	Längengrad: 48,1771; Breitengrad: 11,5603
	Anzahl wichtiger Standorte	4
	Durchschnittliche Pendeldistanz	120 km
	Durchschnittliche Fahrten pro Tour	3,8
Zeit	Fahrten am Werktag	66 %
	Fahrzeugübernachtungen zu Hause	95 %
	Tage mit Fahrzeugbewegungen	90 %
	Standzeiten nach Aktionsradius	1–10 km: 70,5 %; 11–30 km: 4,1 % 31–100 km: 22,4 %; > 100 km: 3,0 %
Fahrtzweck	Fahrten nach Fahrtzwecken	nach Hause: 49,3 %, zur Arbeit: 18,4 % Einkaufen: 9,6 %; Freizeit: 14,7 % Sonstiges: 8,0 %
	Tagesdistanzen	1–10 km: 2,7 %; 11–30 km: 8,1 % 31–100 km: 16,2 %; > 100 km: 73,0 %
Distanz	Fahrdistanzen von zu Hause	1–10km: 54,7 %; 11–30 km: 3,1 % 31–100 km: 4,2 %; > 100 km: 38,0 %

In einem ersten Schritt werden die Korrelationen zwischen den Variablen betrachtet. Sollten starke Zusammenhänge zwischen vielen Variablen existieren, hätte dies einen negativen Effekt auf das Clustering. Korrelationen zwischen zwei Variablen können einen Wert zwischen -1 (negative Korrelation) und 1 (positive Korrelation) annehmen. Je näher der Wert an -1 bzw. 1 ist, desto stärker sind die Variablen voneinander abhängig. Die Korrelationen zwischen den Inputvariablen sind in Abbildung 4.9 dargestellt. Zur Vereinfachung wurden sie farblich visualisiert. Je größer ein Quadrat ist, desto stärker ist die Verbindung zwischen den Variablen. Blaue Quadrate suggerieren positive Korrelationen, während rote Quadrate negativen Korrelationen entsprechen. Von den möglichen Kombinationen haben acht Kombinationen einen Wert größer $0,5$. Die stärkste Korrelation ($-0,66$) existiert zwischen „Anteil Fahrtzweck nach Hause“ und „Anteil Übernachtungen zu Hause“. Die Zweitstärkste ($-0,61$) befindet sich zwischen „Anteil Tagesdistanz (> 100 km)“ und „Anteil Tagesdistanz (31–100 km)“. Ebenfalls herrschen negative Korrelationen zwischen den Fahrzeugübernachtungen und den Standzeiten des Fahrzeugs in einem Radius zwischen 1–10 km ($-0,59$) sowie dem Fahrtzweck „nach Hause“ und dem Fahrtzweck „Freizeit“ ($-0,53$). Die Variable „Standzeit im Radius von zu Hause“ (1–10 km) korreliert mit den anderen Standzeiten (11–30 km ($-0,50$), 31–100 km ($-0,55$), > 100 km ($-0,54$)) sowie dem Fahrtzweck „nach Hause“ ($0,56$).

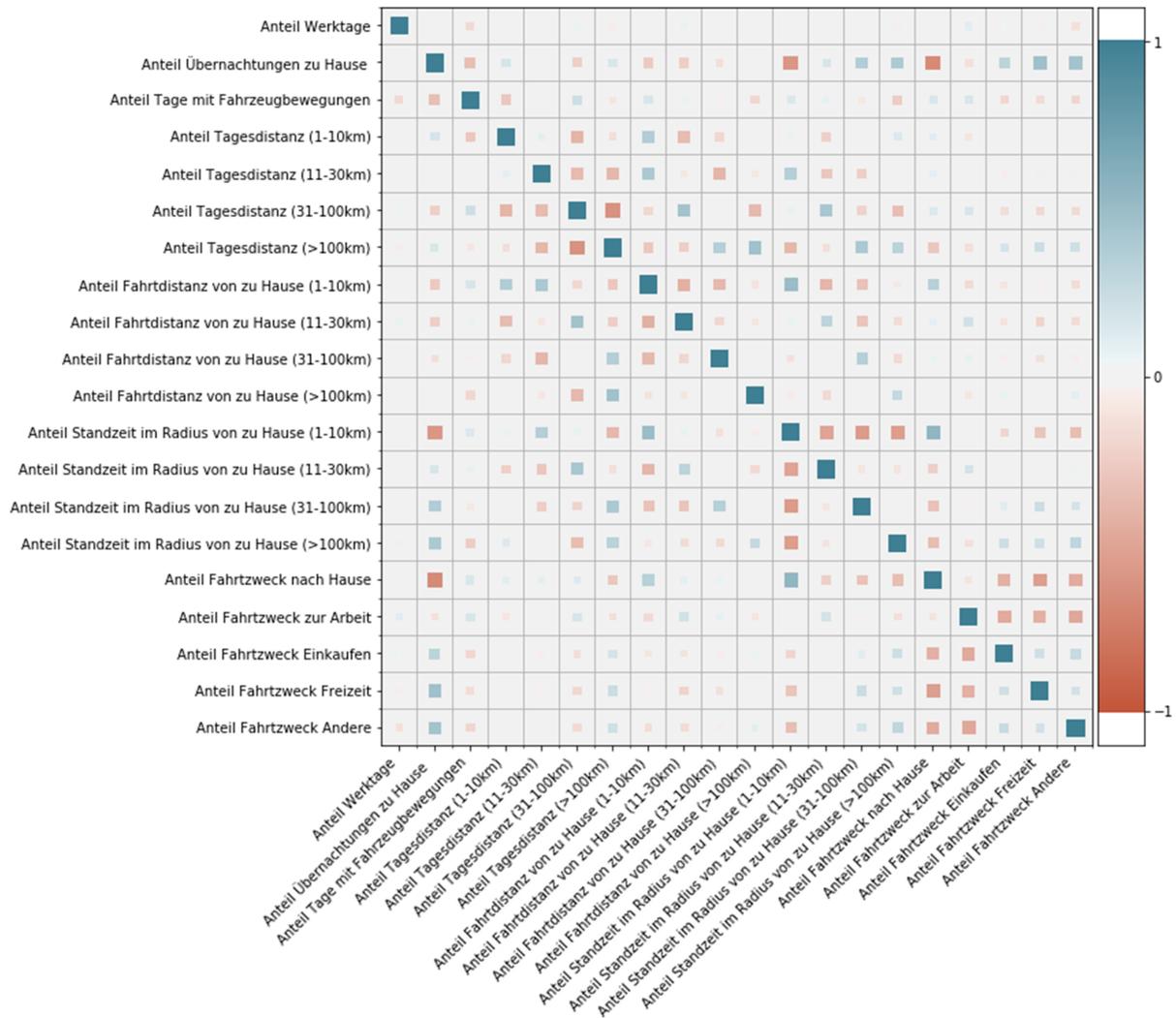


Abbildung 4.9: Korrelationen der Inputvariablen

4.2.3 Methodik

Der folgende Abschnitt stellt das Vorgehen zur Bestimmung von Zeit-Distanz-Fahrtzweck-FNPs vor. Im ersten Schritt wird die generelle Herangehensweise bei Clusteranalysen in fünf Schritten (Variablenauswahl, Dimensionsreduktion, Clustering, Bewertung, Interpretation) dargestellt. Danach werden die 20 Eingangsvariablen auf fünf Dimensionen reduziert und visualisiert. Hierbei wird ein Autoencoder, welcher auch mit nicht linearen Transformationen umgehen kann, angewandt. Das eigentliche Clustering wird über ein Dirichlet Process Gaussian Mixture Model (DPGMM) durchgeführt. Das Verfahren bringt gewisse Vorteile gegenüber anderen Methoden (z. B. k-Means-Verfahren) mit sich, welche im Laufe des Abschnitts noch näher diskutiert werden. Als Nächstes werden die resultierenden Cluster interpretiert und ihre Besonderheiten beschrieben. Die folgenden Berechnungen und Abbildungen wurden mit Python und den entsprechenden Paketen (scikit-learn, keras, scipy, seaborn) durchgeführt.

Generelles Vorgehen bei Clusteranalysen

Die Clusteranalyse wird im Bereich des maschinellen Lernens als unüberwachtes Lernen bezeichnet. Hierbei werden Objekte, in diesem Fall Fahrzeuge und deren Fahrzeugnutzungseigenschaften, in eindeutige Untergruppen eingeteilt. Dabei ist das Ziel, die Gruppen so zu wählen, dass zum einen Fahrzeuge innerhalb einer Gruppe möglichst homogen und zum anderen Fahrzeuge zwischen Gruppen möglichst heterogen sind. Im Gegensatz zum überwachten Lernen gibt es keine vorherige Definition der gebildeten Cluster. Wenn sich die vorliegenden Fahrzeuge gut voneinander trennen lassen, ist ein weiterer Schritt der Validierung die inhaltliche Interpretation der gebildeten Gruppen. Die Vorgehensweise beim Clustering kann grob in fünf iterative Einzelschritte unterteilt werden:

1. Variablenauswahl
2. Dimensionsreduktion
3. Clustering
4. Bewertung des Clusterings
5. Interpretation des Clusterings

(1) Im 1. Schritt werden die entsprechenden Variablen ausgewählt. Die Auswahl der Variablen ist für die Interpretation der Ergebnisse von zentraler Bedeutung. Je nachdem welche Informationen verwendet und wie diese kombiniert werden, hat dies einen Einfluss auf die Cluster. In Abschnitt 3.1.4 wurden vier Variablen (Tage0, Tage10, Tage100, VKT) zur Beschreibung der Distanz und eine Variable (AntWerktag) zur Beschreibung der Zeit verwendet. Um die Fahrzeugnutzung differenzierter zu betrachten, wurde für die Bildung dieser FNP der Fahrtzweck approximiert. Zudem wurden die Dimensionen Distanz und Zeit in Abhängigkeit von Aktionsradien dargestellt (Tabelle 4.5).

(2) Viele Clusteringverfahren verwenden zur Bestimmung von Ähnlichkeiten und Unterschieden Distanzmetriken (z. B. euklidische Distanz). Je mehr Variablen berücksichtigt werden, desto größer sind der Lösungsraum und damit auch die Distanzen zwischen den Datenpunkten. Hochdimensionale Daten können deshalb zu dünn besetzten Matrizen führen. In der Literatur wird dieser Zusammenhang als Fluch der Dimensionalität bezeichnet (Bishop, 2006; Géron, 2019; Hastie et al., 2009; Murphy, 2013). Um diesen Fluch der Dimensionalität zu überwinden, wird die Variablenmenge reduziert. Dabei können Linearkombinationen oder auch nicht lineare Funktionen gebildet werden, welche die ursprünglichen Variablen möglichst gut repräsentieren.

Beispielsweise verwendet Wittwer (2014) zur Reduzierung der Variablenmenge eine PCA. Die zentrale Idee der PCA besteht darin, die Dimensionalität eines Datensatzes, der aus einer großen Anzahl von Dimensionen besteht, zu reduzieren und gleichzeitig die im ursprünglichen Datensatz enthaltenen Informationen so gut wie möglich zu erhalten (Dunteman, 1989). Im Anhang A wird die PCA im Detail erklärt.

(3) Nach der Dimensionsreduktion wird das eigentliche Clustering durchgeführt und Gruppen (in unserem Fall FNPs) gebildet. Wie die Literaturübersicht bereits gezeigt hat, ist eines der gängigsten Verfahren zur Identifikation von Clustern das k-Means-Verfahren. Auch Wittwer (2014) verwendet, nachdem die Dimensionsreduktion über eine PCA durchgeführt worden ist, das k-Means-Verfahren zur Bestimmung von typischen Verhaltensmustern junger Erwachsener. Im Anhang B wird der k-Means-Algorithmus im Detail erklärt.

(4) Die Bewertung des Clusterings kann gleichzeitig auch ein Indikator für die Anzahl der Gruppen sein. Viele Algorithmen (z. B. k-Means-Algorithmus, Gaussian Mixture Models (GMM)) setzen die Anzahl der Cluster bzw. Nutzenfunktionen voraus. Beispielsweise wird beim k-Means-Algorithmus oder den GMM die Anzahl der Gruppen variiert und die berechneten Modelle gegenübergestellt. Bei einer latenten Klassenanalyse müssen andere Verfahren (z. B. Ward-Methode) vor dem eigentlichen Clustering vorgeschaltet werden, um die Anzahl der Cluster zu bestimmen. Im Anhang C werden zwei Metriken (Elbow-Kriterium und Silhouettenkoeffizient) zur Bewertung von Clusterlösungen präsentiert.

(5) Im letzten Schritt werden die berechneten Gruppen interpretiert, indem besondere Eigenschaften (z. B. sehr hohe oder niedrige Werte), welche sich in den ausgewählten Variablen widerspiegeln, identifiziert und beschrieben werden. Die inhaltliche Validierung erfordert immer eine Experteneinschätzung und kann gerade bei Segmentierungsansätzen nur schwer automatisch erfolgen.

Sind die Cluster nicht sinnvoll interpretierbar, muss der Prozess mit einer veränderten Variablenauswahl und veränderten Parametern wiederholt werden.

Dimensionsreduktion mit Autoencoder

Nachdem die Variablen für das Clustering in Tabelle 4.5 bereits beschrieben worden sind, findet im nächsten Schritt die Reduzierung der Variablenmenge und damit der Dimensionen statt. Die Anzahl der dimensionsreduzierten Variablen z wurde iterativ durch Erprobung verschiedener Lösungen festgelegt. Der Autoencoder ist ein gängiges Verfahren zur Dimensionsreduktion aus dem Bereich des Deep Learnings. Während beim maschinellen Lernen Funktionen zwischen Eingangs- und Ausgangsvariablen, also zwei Schichten, erlernt werden, kann beim Deep Learning die Anzahl der Schichten in Abhängigkeit von der Problemstellung variieren. Infolgedessen kann die Komplexität einer Funktion zur Lösung von nicht linearen Problemen beliebig ausgestaltet werden (Goodfellow, Bengio & Courville, 2016). Dies ist auch ein Vorteil

gegenüber einer PCA, welche nur lineare Zusammenhänge abbilden kann. Deep-Learning-Verfahren werden oftmals in sogenannten Netzarchitekturen skizziert, um den Aufbau des Modells vereinfacht darzustellen (Abbildung 4.10). Die Kreise entsprechen den einzelnen Variablen des Modells. Die Vertikale von Variablen ist eine sogenannte Schicht. Die erste und letzte Schicht sind jene der Ein- und Ausgangsvariablen. Die Linien sind das Bindeglied zwischen den Variablen, welche durch Gewichte θ miteinander in Verbindung stehen. Eine Variable (z. B. $x_1^{(1)}$) aus der ersten Schicht ist mit allen Variablen ($x_1^{(2)}, x_2^{(2)}, \dots, x_{10}^{(2)}$) aus der zweiten Schicht über Gewichte ($w_1^{(12)}, w_2^{(12)}, \dots, w_n^{(12)}$) verbunden. Dadurch ergibt sich eine große Menge von Gewichten $\theta^{(12)} = \{w_1^{(12)}, w_2^{(12)}, \dots, w_n^{(12)}\}$ zwischen den Schichten. Je größer die Anzahl von Variablen und Schichten, desto mehr Gewichte müssen optimiert werden. Da alle Variablen einer Schicht mit allen Variablen der Nachbarschicht verbunden sind, ist es nicht möglich, Linearkombinationen zu bilden. Dadurch repräsentieren die dimensionsreduzierten Variablen lediglich die zugrundeliegende Struktur der Originalvariablen in einem Raum mit weniger Dimensionen.

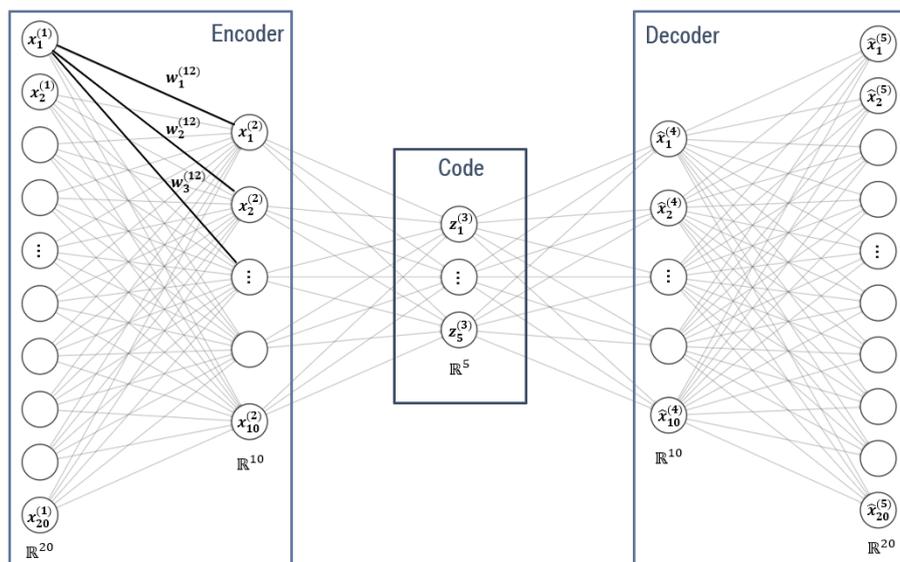


Abbildung 4.10: Schematische Netzarchitektur des verwendeten Autoencoders

Die Architektur des Autoencoders kann in drei Bereiche (Encoder, Code und Decoder) eingeteilt werden. Im Bereich „Encoder“ wird die Eingangsschicht $x^{(1)} = \{x_1^{(1)}, x_2^{(1)}, \dots, x_{20}^{(1)}\}$ mit einer verdeckten zweiten Schicht $x^{(2)} = \{x_1^{(2)}, x_2^{(2)}, \dots, x_{10}^{(2)}\}$ über Gewichte $\theta^{12} = \{w_1^{(12)}, w_2^{(12)}, \dots, w_{210}^{(12)}\}$ verbunden:

$$x^{(2)} = f(\theta^{(12)}x^{(1)} + b^{(1)}) \quad (4.6)$$

Die zusätzliche Variable $b^{(1)}$ entspricht einem sogenannten Verzerrungsterm, welcher standardmäßig bei jeder Schicht mit integriert wird. Die Anzahl der Gewichte ($210 = (20 + 1) \cdot 10$) berechnet sich über die Anzahl der Variablen in der Eingangsschicht (20) sowie in der zweiten Schicht (10) und den Verzerrungsterm (1). Ähnlich wie beim Vorgehen der Softmax-Regression (SR) fließen die Variablen und ihre Gewichte von einer Schicht zur nächsten durch sogenannte Aktivierungsfunktionen $f(\cdot)$, welche unterschiedliche Eigenschaften erfüllen können (Géron, 2019). Beispielsweise ist es auch möglich, eine Softmax-Funktion als Aktivierungsfunktion zu verwenden. Ist die verwendete Aktivierungsfunktion linear, entspricht der Autoencoder einer herkömmlichen PCA. Für diese Anwendung wurde die sogenannte Scaled-Exponential-Linear-Units (SELU)-Aktivierungsfunktion gewählt. Details zur verwendeten Aktivierungsfunktion liefert die Arbeit von Klambauer, Unterthiner, Mayr und Hochreiter (2017).

Analog zur Verbindung der Eingangsschicht mit der zweiten Schicht erfolgt die Verbindung zwischen der zweiten und der dritten Schicht, wodurch $\theta_e = \{\theta^{(12)}, \theta^{(23)}\}$ Gewichte im Encoder-Bereich berücksichtigt werden:

$$z = f(\theta^{(23)}x^{(2)} + b^{(2)}) \quad (4.7)$$

Der Bereich „Code“ stellt den dimensionsreduzierten Vektor $z = \{z_1^{(3)}, z_2^{(3)}, \dots, z_5^{(3)}\}$ dar. Der Vektor z wird, ähnlich wie die Hauptkomponenten bei einer PCA, später als Eingangsvektor für das Clustering verwendet. Die Besonderheit bei einem Autoencoder ist die Symmetrie in der Netzarchitektur, welche sich um den Code-Bereich spiegelt. Im „Decoder“ wird symmetrisch zum „Encoder“ die Funktion mit den Gewichten $\theta_d = \{\theta^{(34)}, \theta^{(45)}\}$ hinsichtlich $\hat{x}^{(4)} = \{\hat{x}_1^{(4)}, \hat{x}_2^{(4)}, \dots, \hat{x}_{10}^{(4)}\}$ und $\hat{x}^{(5)} = \{\hat{x}_1^{(5)}, \hat{x}_2^{(5)}, \dots, \hat{x}_{20}^{(5)}\}$ angepasst:

$$\hat{x}^{(4)} = f(\theta^{(34)}z + b^{(3)}) \quad (4.8)$$

$$\hat{x}^{(5)} = f(\theta^{(45)}\hat{x}^{(4)} + b^{(3)}) \quad (4.9)$$

Der Nachteil eines Autoencoders ist, dass durch die tiefen Kettenstrukturen keine Linearkombinationen aus den ursprünglichen Variablen gebildet werden können, wodurch der Vektor z nicht mehr interpretierbar ist.⁹

$$J(\theta) = \frac{1}{n} \sum_{i=1}^N (x_i^{(1)} - \hat{x}_i^{(5)})^2 \quad (4.10)$$

Ziel des Optimierungsverfahrens ist es, die Gewichte $\theta = \{\theta_e, \theta_d\}$ anzupassen, sodass die Abweichung zwischen den Variablen der Eingangsschicht $x^{(1)}$ und der Ausgangsschicht $\hat{x}^{(5)}$ und damit auch die Verlustfunktion $J(\theta)$ minimal sind (Géron, 2019; Goodfellow et al.,

⁹Zur Vereinfachung wird der generierte Vektor z für das Clustering im Folgenden wieder als x bezeichnet.

2016). Dadurch wird aus einem unüberwachten Verfahren ein überwachtes Verfahren, da die Ausgangsvariablen den Eingangsvariablen möglichst ähnlich sein müssen. Die Anpassung der Gewichte θ erfolgt über verschiedene Optimierungsverfahren. Bei diesem Autoencoder wurde das am meisten genutzte Verfahren, das sogenannte stochastische Gradientenabstiegsverfahren (SGD), angewandt:

$$\theta_{t+1} = \theta_t - \eta \nabla J(\theta_t) \quad (4.11)$$

Hierbei entspricht θ_t den Gewichten zum Zeitpunkt t . Die Gewichte θ_{t+1} stimmen mit θ_t abzüglich der 1. Ableitung der Verlustfunktion nach den Gewichten zum Zeitpunkt t ($\nabla J(\theta_t)$) und einer Lernrate η überein. Die Lernrate beträgt 0,03, was einer Standardeinstellung gleichkommt. Die mathematisch effiziente Berechnung der 1. Ableitung der Verlustfunktion erfolgt über die Kettenregel und wird im Bereich des Deep Learnings als Backpropagation-Algorithmus bezeichnet (Rumelhart, Hinton & Williams, 1986). Die Anzahl der Epochen (eine Epoche entspricht einer Anpassung der Gewichte) wurde auf 500 festgelegt. Die Anzahl wird über die Verlustfunktion bestimmt; sobald diese gegen einen bestimmten Wert konvergiert, hat die Hinzunahme einer weiteren Epoche keinen Mehrwert mehr. Die Verlustfunktion liefert zudem Hinweise auf die optimale Anzahl von Variablen in den verdeckten Schichten sowie die generelle Anzahl von Schichten. Je komplexer das Problem ist, desto mehr Variablen und Schichten sind nötig. Gleichzeitig steigt mit der Größe des Netzwerkes auch der Rechenaufwand. Da 20 Eingangsvariablen keine hohe Komplexität für ein neuronales Netzwerk darstellen (beispielsweise werden Autoencoder auch intensiv in der Bild- und Videoverarbeitung eingesetzt), wurde die Architektur möglichst klein gehalten. Zudem steht in dieser Arbeit auch die generelle Methodik neuronaler Netze zur Anwendung im Verkehrsbereich im Vordergrund. Die Einheiten in der Ein- und Ausgangsschicht des verwendeten Netzwerkes entsprechen jeweils der Anzahl der generierten Variablen (20). Der Code-Bereich (z) hat fünf Variablen. Die zweite und vierte Schicht zwischen Encoder und Code bzw. zwischen Code und Decoder haben jeweils zehn Einheiten. Die Anzahl der zu optimierenden Parameter hängt von der Anzahl der Variablen und Schichten ab. Insgesamt wurden 545 $((20 + 1) \cdot 10 + (10 + 1) \cdot 5 + (5 + 1) \cdot 10 + (10 + 1) \cdot 20)$ Parameter während des Trainings optimiert. Ein weiterer Vorteil von Autoencodern gegenüber einer PCA ist der hohe Grad an Flexibilität. Durch die vielen Einstellungen (Anzahl Schichten, Anzahl Variablen in den verdeckten Schichten, Aktivierungsfunktion, Optimierungsverfahren, Verlustfunktion) kann das Netzwerk flexibel auf die Problemstellung angepasst werden. Für nähere Informationen zu Optimierungsverfahren bzw. neuronalen Netzen wird das Standardwerk von Goodfellow et al. (2016) empfohlen. Abbildung 4.11 stellt einen sogenannten Paarplot mit den fünf für das Clustering verwendeten dimensionsreduzierten z -Variablen dar.

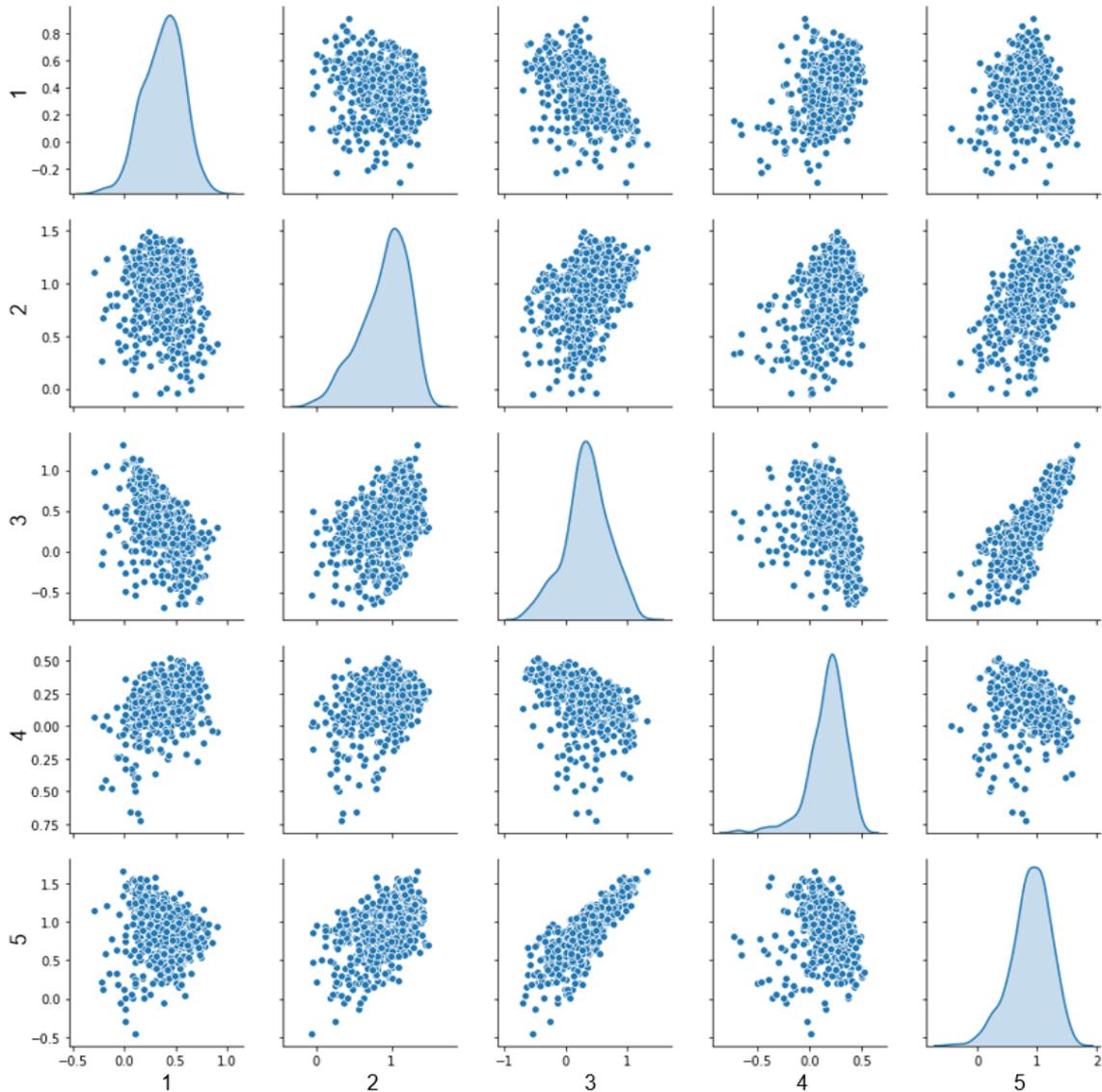


Abbildung 4.11: Paarplot mit den fünf z -Variablen

Der Paarplot zeigt in den Diagonalen die Verteilungen der fünf z -Variablen und außerhalb der Diagonalen die Korrelationen zwischen den Variablen anhand von Punktwolken. Die fünf Variablen folgen alle einer Gauß-Verteilung mit unterschiedlicher Streuung. Je spitzer die Verteilung ist, desto geringer ist die Streuung. Auch zwischen den Variablen zeigen sich keine starken Korrelationen. Diese Darstellung dient zur Validierung, um zu prüfen ob es zwischen den Variablen starke Zusammenhänge gibt. Wäre dies der Fall, würde sich das negativ auf das Clustering auswirken.

Visualisierung hochdimensionaler Daten

Nachdem die Variablenmenge reduziert worden ist, werden im nächsten Schritt die zugrundeliegenden fünf Variablen zweidimensional dargestellt. Mittels t-Distributed-Stochastic-Neighbor-Embedding (t-SNE)-Verfahren können hochdimensionale Daten ($d > 3$) in zwei- bzw. dreidimensionale Daten konvertiert und visualisiert werden. Details zur Durchführung des Verfahrens liefert die Arbeit von Maaten und Hinton (2008). Abbildung 4.12 zeigt die dimensionsreduzierten Daten anhand eines Kerndichteschätzer (KDE)-Plots. Der KDE-Plot eignet sich vor allem, um Dichten von Daten darzustellen. Je dunkler die Farbe, desto dichter sind die Daten in diesem Bereich. Anhand der Visualisierung lassen sich schon fünf bis sechs Bereiche identifizieren, an denen eine hohe Dichte an Datenpunkten vorhanden ist.

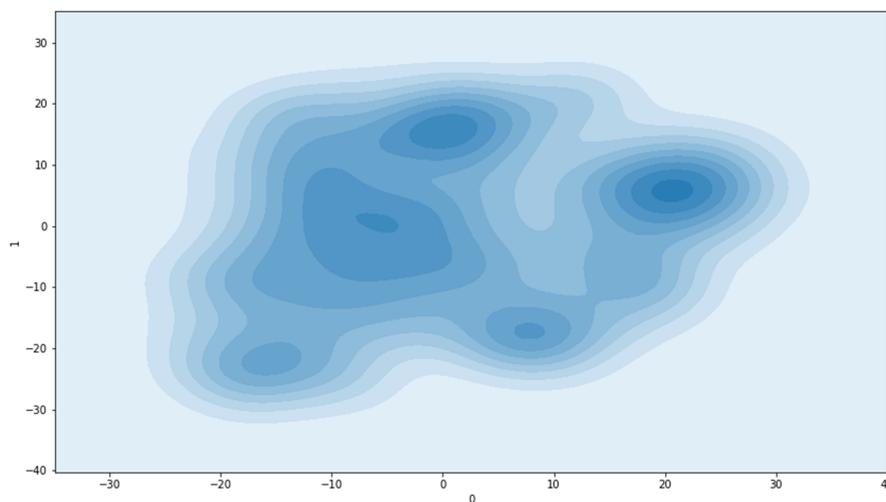


Abbildung 4.12: *t-SNE-KDE-Plot*

Clustering mit einem Dirichlet Process Gaussian Mixture Model

Das Clustering mit einem Dirichlet Process Gaussian Mixture Model (DPGMM) setzt sich aus einem sogenannten Dirichlet-Prozess und einem herkömmlichen GMM zusammen. Da das angewandte Verfahren kein gängiges Verfahren aus dem Verkehrsbereich ist, wird es im Detail beschrieben und die Unterschiede zu gängigeren Verfahren (k-Means-Algorithmus) erarbeitet. Bei einem GMM handelt es sich um ein probabilistisches Modell, welches annimmt, dass die zugrundeliegenden Daten aus k multivariaten Gauß-Verteilungen (Normalverteilungen) stammen. Eine multivariate Gauß-Verteilung ist die verbreitetste Verteilung in der Statistik, deren Wahrscheinlichkeitsdichtefunktion sich wie folgt zusammensetzt:

$$\mathcal{N}(\mathbf{x}|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right], \quad (4.12)$$

wobei μ den Mittelwerten und Σ den Kovarianzen von K Verteilungen entsprechen. $\sqrt{(2\pi)^{\frac{D}{2}} \Sigma^{\frac{1}{2}}}$ dient zur Normalisierung, damit die Dichte auf 1 integriert werden kann.

Der große Unterschied zum k-Means-Clusteringverfahren ist, dass dieses für die Bestimmung der Cluster lediglich den Mittelwert \bar{x}_k eines Clusters anpasst, indem die Streuungsquadratsumme (Differenz zwischen allen Mittelwerten eines Clusters k und den zugehörigen Datenpunkten x_{ik}) minimiert wird (Murphy, 2013). Die Varianz hingegen ist bei der Optimierung konstant. Diese Annahme kann bei bestimmten Verteilungen zu Fehlklassifikationen führen. Bei einem GMM hingegen wird die Varianz bei der Optimierung berücksichtigt. Der zweite Unterschied zwischen einem GMM und dem k-Means-Verfahren ist, dass es sich bei einem GMM um ein sogenanntes „weiches Clustering“ handelt. Somit werden, wie bereits bei der Anwendung der SR in Unterkapitel 3.2, Wahrscheinlichkeiten für die Clusterzugehörigkeit berechnet.

Ziel des verwendeten GMM ist es, die bedingte Wahrscheinlichkeit zu berechnen, sodass die Eigenschaften x_i eines Fahrzeugs i aus dem Cluster / der Gauß-Verteilung z_k stammen (Bishop, 2006; Géron, 2019; Murphy, 2013):

$$p(z_k = 1|x_i) \quad (4.13)$$

Mit Hilfe des Satzes von Bayes (im Anhang D wird die Herleitung des Satzes von Bayes im Detail erklärt) kann die Gleichung wie folgt dargestellt werden:

$$p(z_k = 1|x_i) = \frac{p(x_i|z_k = 1)p(z_k = 1)}{p(x_i)} \quad (4.14)$$

- $p(z_k = 1)$ ist die Gesamtwahrscheinlichkeit, dass die Beobachtungen aus der Gauß-Verteilung k stammen. Diese Wahrscheinlichkeit wird im Folgenden als π_k bezeichnet, wobei $\sum_{k=1}^K \pi_k = 1$ und $\pi_k \geq 0$. Unter der Annahme, dass jedes z voneinander unabhängig ist, gilt: $p(\mathbf{z}) = p(z_1 = 1)^{z_1} p(z_2 = 1)^{z_2} \dots p(z_K = 1)^{z_K} = \prod_{k=1}^K \pi_k^{z_k}$.
- $p(x_i|z_k = 1)$ beschreibt die Wahrscheinlichkeit, dass, unter der Annahme von k Gauß-Verteilungen, x_i Daten beobachtet werden. Dieser Zusammenhang ist nichts anderes als die Wahrscheinlichkeitsdichtefunktion der Gauß-Verteilung selbst: $p(x_i|z_k = 1) = \mathcal{N}(x_i|\mu_k, \Sigma_k)$. Unter der Annahme, dass alle z unabhängig voneinander sind, gilt: $p(x_i|\mathbf{z}) = \prod_{k=1}^K \mathcal{N}(x_i|\mu_k, \Sigma_k)^{z_k}$.
- $p(x_i)$ wird mit Hilfe der berechneten Wahrscheinlichkeiten $p(\mathbf{z})$ und $p(x_i|\mathbf{z})$ bestimmt. Die Kettenregel für Wahrscheinlichkeiten besagt, dass die multivariate Verteilung zweier Zufallsvariablen $p(x_i, \mathbf{z}) = p(x_i|\mathbf{z})p(\mathbf{z})$ entspricht. Um nun die Randverteilung $p(x_i)$ zu erhalten, müssen lediglich alle \mathbf{z} aufsummiert werden: $p(x_i) = \sum_{k=1}^K p(x_i|z_k)p(z_k)$.

Somit kann Gleichung 4.14 wie folgt umgestellt werden:

$$p(z_k = 1|x_i) = \frac{p(x_i|z_k = 1)p(z_k = 1)}{\sum_{j=1}^K p(x_i|z_j = 1)p(z_j = 1)} \quad (4.15)$$

Durch Einsetzen der aufgeführten Variablen in Gleichung 4.15 wird die Formel zur Bestimmung der Wahrscheinlichkeit generiert, sodass die Eigenschaften x_i für das Fahrzeug i aus dem FNP / der Gauß-Verteilung k stammen:

$$p(z_k = 1|x_i) = \frac{\mathcal{N}(x_i|\mu_k, \Sigma_k)\pi_k}{\sum_{j=1}^K \mathcal{N}(x_i|\mu_j, \Sigma_j)\pi_j} = \gamma(z_{ik}) \quad (4.16)$$

Wie aus der Gleichung 4.16 abzulesen ist, müssen die Parameter $\theta_k = \{\pi_k, \mu_k, \Sigma_k\}$ für k Verteilungen geschätzt werden. Die Berechnung der Parameter erfolgt mit Hilfe des Erwartungs-Maximierungs (EM)-Verfahrens (Dempster, Laird & Rubin, 1977). Ähnlich wie bei der Minimierung der Streuungsquadratsumme beim k-Means-Verfahren, handelt es sich hier ebenfalls um ein iteratives Verfahren. Die Parameter θ_k werden initialisiert und es wird mit k zufälligen Gauß-Verteilungen gestartet. Die Werte für die Parameter $\hat{\mu}_1, \dots, \hat{\mu}_K, \hat{\sigma}_1, \dots, \hat{\sigma}_K$ und $\hat{\pi}_1, \dots, \hat{\pi}_K$ werden mit Hilfe des Dirichlet-Prozesses bestimmt.

Im Erwartungsschritt sind die Werte $\hat{\theta} = \{\hat{\theta}_1, \dots, \hat{\theta}_K\}$ fix und die Wahrscheinlichkeiten der Zugehörigkeit $\hat{\gamma}(z_{ik})$ zu k Clustern/Gauß-Verteilungen werden berechnet:

$$\hat{\gamma}(z_{ik}) = \frac{\mathcal{N}(x_i|\mu_k, \Sigma_k)\hat{\pi}_k}{\sum_{j=1}^K \mathcal{N}(x_i|\mu_j, \Sigma_j)\hat{\pi}_j}, \quad (4.17)$$

wobei $\hat{\gamma}(z_{ik}) = p(z_k|x_i, \hat{\pi}, \hat{\mu}, \hat{\sigma})$ die Wahrscheinlichkeit von i hinsichtlich der Zugehörigkeit zu Cluster z_k , unter der Annahme der fixen Parameter $\hat{\theta} = \{\hat{\pi}, \hat{\mu}, \hat{\sigma}\}$, darstellt.

Im Maximierungsschritt werden anhand der berechneten Wahrscheinlichkeiten $\hat{\gamma}(z_{ik})$ die Parameter $\hat{\theta}$ angepasst:

$$\begin{aligned} \hat{\pi}_k &= \frac{\sum_{i=1}^N \hat{\gamma}(z_{ik})}{N} \\ \hat{\mu}_k &= \frac{\sum_{i=1}^N \hat{\gamma}(z_{ik})x_i}{\sum_{i=1}^N \hat{\gamma}(z_{ik})} \\ \hat{\Sigma}_k &= \frac{\sum_{i=1}^N \hat{\gamma}(z_{ik})(x_i - \mu_k)(x_i - \mu_k)^T}{\sum_{i=1}^N \hat{\gamma}(z_{ik})} \end{aligned} \quad (4.18)$$

Diese beiden Schritte werden so lange wiederholt, bis sich die Parameter nicht mehr wesentlich verändern und konvergieren. Im Endergebnis können dann die Wahrscheinlichkeiten $p(x_i|\mu, \Sigma)$ berechnet werden, sodass ein Fahrzeug i mit einer gewissen Wahrscheinlichkeit der k -ten Gauß-Verteilung angehört:

$$p(x_i | \mu_k, \Sigma_k) = \sum_{k=1}^K \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k) \quad (4.19)$$

Ähnlich wie bei der Anpassung des Mittelwertes beim k-Means-Verfahren, kann es zu lokalen Optima kommen, weshalb die Modellparameter θ für das Clustering 200-mal initialisiert werden. Die Herausforderung beim Clustering ist, die Anzahl der k multivariaten Gauß-Verteilungen zu identifizieren. Hierzu kann der Algorithmus, wie bei der Beschreibung des Vorgehens bei Clusteranalysen bereits erläutert, mit unterschiedlichen Annahmen zu k durchlaufen, um im Nachgang, mit verschiedenen Metriken (z. B. AIC, BIC, Silhouettenkoeffizient, Elbow-Methode), die optimale Anzahl von Cluster festzulegen. Nachteil dabei ist, dass das Verfahren mehrmals mit unterschiedlichen k das Optimierungsproblem lösen muss bzw. andere Verfahren zur Bestimmung der optimalen Anzahl der Cluster vorgeschaltet werden müssen.

Durch den Dirichlet-Prozess kann dieses Effizienzproblem gelöst werden, indem die Anzahl von k (Gauß-)Verteilungen vorab bestimmt wird. Der Dirichlet-Prozess stammt aus der nicht parametrischen bayesianischen Statistik. Dabei wird angenommen, dass die Anzahl der möglichen Verteilungen gegen unendlich geht ($k \rightarrow \infty$). Somit ist k eine dynamische Variable, welche vom Anstieg der Stichprobe und einem Hyperparameter α abhängt. Dies ist speziell bei dynamischen Fahrzeugdaten relevant, da dort täglich neue Fahrzeuge mit unterschiedlicher Nutzung hinzukommen können. Die austauschbare Anzahl von k Gauß-Verteilungen und die damit verbundene Veränderung von $\pi = \{\pi_k\}_{k=1}^{\infty}$ kann über verschiedene stochastische Prozesse (z. B. Chinese Restaurant Process (CRP) (Aldous, 1985) oder Stick-Breaking Process (Paisley, 2010)) vorhergesagt werden. Der CRP ist ein metaphorisches Beispiel für die Identifikation von k und π , anhand dessen sich der Dirichlet-Prozess einfach beschreiben lässt. Die grundlegende Idee ist, dass Kunden in einem Restaurant die Wahl haben, sich zu einem bereits besetzten Tisch zu setzen oder einen neuen Tisch zu eröffnen. Je mehr Personen an einem Tisch sitzen, desto wahrscheinlicher fällt die Wahl auf diesen Tisch. Die k Tische im Restaurant stehen metaphorisch für k (Gauß-)Verteilungen:

$$\begin{aligned} \pi_k &= \frac{n_k}{\alpha + n - 1} \\ \pi_{-k} &= \frac{\alpha}{\alpha + n - 1} \end{aligned} \quad (4.20)$$

Dabei ist π_k die Wahrscheinlichkeit, dass ein neues Fahrzeug einem bestehenden FNP / Cluster k zugewiesen wird. π_{-k} ist die Wahrscheinlichkeit, dass durch ein weiteres Fahrzeug ein neues FNP/Cluster entsteht. Anhand der Formel ist zu erkennen, dass π_k und π_{-k} von der Stichprobengröße n und dem Konzentrationsparameter α abhängen. Darüber hinaus ist Folgendes ersichtlich: Je größer die Anzahl der Fahrzeuge n_k in einem Cluster k ist, desto

eher wird das FNP / Cluster von einem hinzukommenden Fahrzeug besetzt. Die Variable n ist die Gesamtzahl der Fahrzeuge über alle Cluster $\sum_{k=1}^K n_k$. α ist der Konzentrationsparameter. Je größer α ist, desto eher wird ein neues Cluster eröffnet (Heller, 2008). α ist in diesem Modell 1, was einer Standardeinstellung entspricht. Ist die Anzahl k (Gauß-)Verteilungen für die Stichprobengröße n bestimmt, können die Parameter $\{\theta_k\}_{k=1}^{\infty}$ für k (Gauß-)Verteilungen bestimmt werden:

$$\mathbf{G} \sim DP(\alpha, G_0) \quad (4.21)$$

Der Dirichlet-Prozess an sich ist eine Verteilung über mehrere (Dirichlet-)Verteilungen (Görür & Rasmussen, 2010; Rasmussen, 2000). Mit dem Konzentrationsparameter α und der Basisverteilung G_0 können unendliche Verteilungen \mathbf{G} generiert werden. Dabei können über den Hyperparameter α , welcher bereits beim stochastischen Prozess (CRP oder Stick-Breaking Process) verwendet wurde, k Gauß-Verteilungen $\mathbf{G} = \{G_k\}_k^{\infty}$ mit den Parametern $\{\mu_k, \Sigma_k\}_{k=1}^{\infty}$ generiert werden. Somit sind alle notwendigen Parameter ($\sum_{k=1}^{\infty} \theta_k$) bestimmt und können in die Gleichung 4.17 eingesetzt werden. Danach findet, wie bereits erläutert, iterativ die Anpassung der Parameter über das EM-Verfahren bis zur Konvergenz statt (Blei, Jordan et al., 2006; Liu, Beyca, Rao, Kong & Bukkapatnam, 2016). Die maximale Anzahl von Iterationen wurde auf 250 festgelegt. Auch dieser Wert entspricht einer Standardeinstellung. Als weiteren Validierungsschritt konnte über eine aus dem verwendeten Softwarepaket implementierte Funktion geprüft und bestätigt werden, dass die Parameter zu einem Wert konvergiert sind. Durch die Annahme von unendlich vielen Gauß-Verteilungen kann eine große Anzahl von k Gauß-Verteilungen mit geringen Wahrscheinlichkeiten π_k resultieren. Gerade bei dynamischen Daten, wie Sensordaten eines Fahrzeugs, ist die Variabilität in der Anzahl der Cluster sinnvoll, da mit dem Hinzukommen neuer Fahrzeuge auch neue FNPs entstehen können. Darüber hinaus ist für ein automatisiertes Clustering die Annahme unendlich vieler Gauß-Verteilungen sinnvoll und praktisch. Für das hier vorliegende Clustering wird aufgrund der relativ kleinen Stichprobe ein Schwellenwert für die Gesamtwahrscheinlichkeit (π_k) der Zuordnung eines Fahrzeugs zu einem FNP k von größer 5% (0,05) festgelegt. Dadurch reduziert sich die Anzahl von Cluster und es entstehen keine Cluster mit nur einem Fahrzeug. Zukünftige Arbeiten sollten sich genau damit beschäftigen, solch kleine Cluster zu untersuchen. Für die 625 Fahrzeuge und ihre fünf dimensionsreduzierten Variablen wurden fünf Cluster auf der Basis der verwendeten Methodik und der verwendeten Einstellungen identifiziert.

Zusammenfassend wird durch die Verwendung von DPGMM, im Gegensatz zum k-Means-Verfahren, die Streuung von Clustern in der Optimierung berücksichtigt. Des Weiteren ist es über einen Dirichlet-Prozess möglich, die Anzahl der Cluster zu bestimmen, ohne dass verschiedene Lösungen im Vorfeld getestet oder die Anzahl über ein vorgeschaltetes Verfahren identifiziert werden muss.

Visualisierung der Clusterlösung

Wie bereits bei der formalen Definition der Methodik erläutert, entspricht π_k der durchschnittlichen Zuordnungswahrscheinlichkeit zum/zur k -ten FNP/Gauß-Verteilung. Abbildung 4.13 zeigt die unterschiedlichen Zuordnungswahrscheinlichkeiten der fünf identifizierten FNPs. Cluster 1 (32 %) und 3 (31 %) haben mit 0,325 bzw. 0,299 die höchste Gewichtung und sind damit auch die größten Cluster. Die durchschnittliche Zuordnungswahrscheinlichkeit bei Cluster 2 (16 %) entspricht 0,159, während Cluster 4 (14 %) eine Gewichtung von 0,153 besitzt. In Cluster 5 (6 %) sind die wenigsten Fahrzeuge enthalten und es hat damit auch die geringste Zuordnungswahrscheinlichkeit mit 0,063.

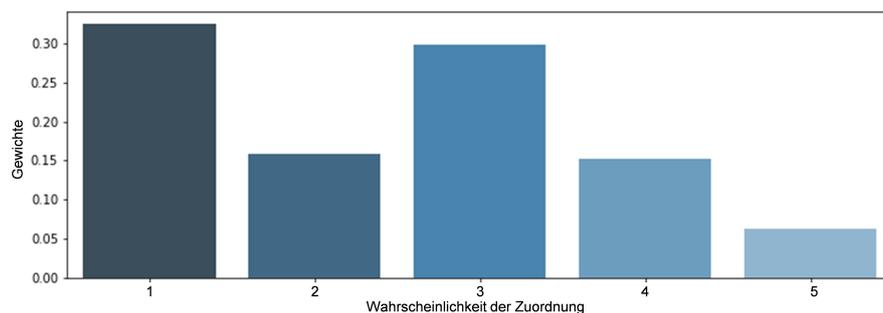


Abbildung 4.13: Berechnete Gesamtwahrscheinlichkeit π_k der fünf Gauß-Verteilungen

Abbildung 4.14 zeigt den t-SNE-KDE-Plot mit den farblich zugeordneten Clustern. Zu erkennen ist hier, dass sich die größten Cluster (1 und 3) zentral positioniert haben, während sich die kleineren Cluster eher am Rand befinden und möglicherweise „extreme“ FNPs repräsentieren.

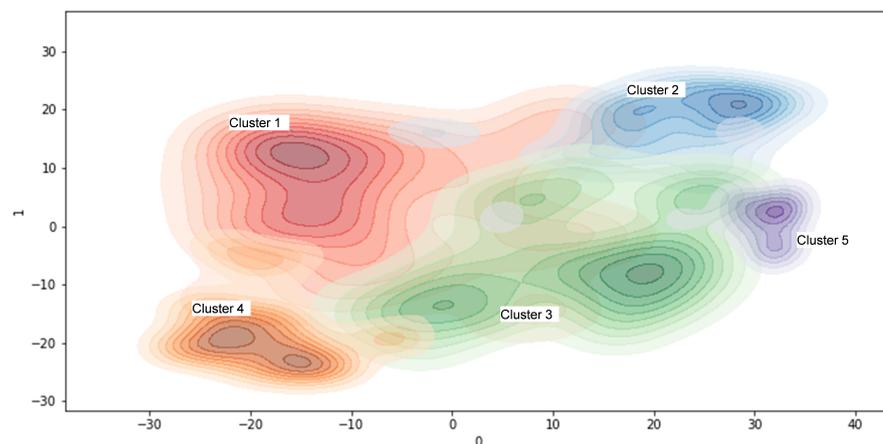


Abbildung 4.14: t-SNE-KDE-Plot in Abhängigkeit von den fünf Clustern

Abbildung 4.15 stellt mittels Paarplot die Korrelationen und Verteilungen der dimensionsreduzierten fünf Variablen in Abhängigkeit von den fünf bestimmten FNPs/Clustern dar. Hier zeigt sich, dass die jeweiligen Clusterpunkte entlang der fünf Variablen alle eng nebeneinander

verortet und nicht zufällig verteilt sind. Lediglich das kleinste Cluster 5 (lila) weist eine gewisse Streuung auf, was sich möglicherweise durch die niedrige Wahrscheinlichkeit der Zuordnung (π_5) und die damit verbundene niedrige Clustergröße begründen lässt. Darüber hinaus können auch Ähnlichkeiten und Unterschiede entlang der fünf Gauß-Verteilungen identifiziert werden. Beispielsweise besitzen Cluster 2 (blau) und Cluster 3 (grün) bei der dimensionsreduzierten Variable „2“ eine sehr ähnliche Verteilung. Bei Variable „3“ sind sich wiederum Cluster 1 (rot) und Cluster 3 (grün) sehr ähnlich. Cluster 5 ist, im Vergleich zu den anderen Clustern, am unterschiedlichsten. Fast über alle fünf Variablen weicht die Verteilung von Cluster 5 von den restlichen ab. Dies konnte bereits in Abbildung 4.14 beobachtet werden und unterstreicht die Annahme des extremen FNP.

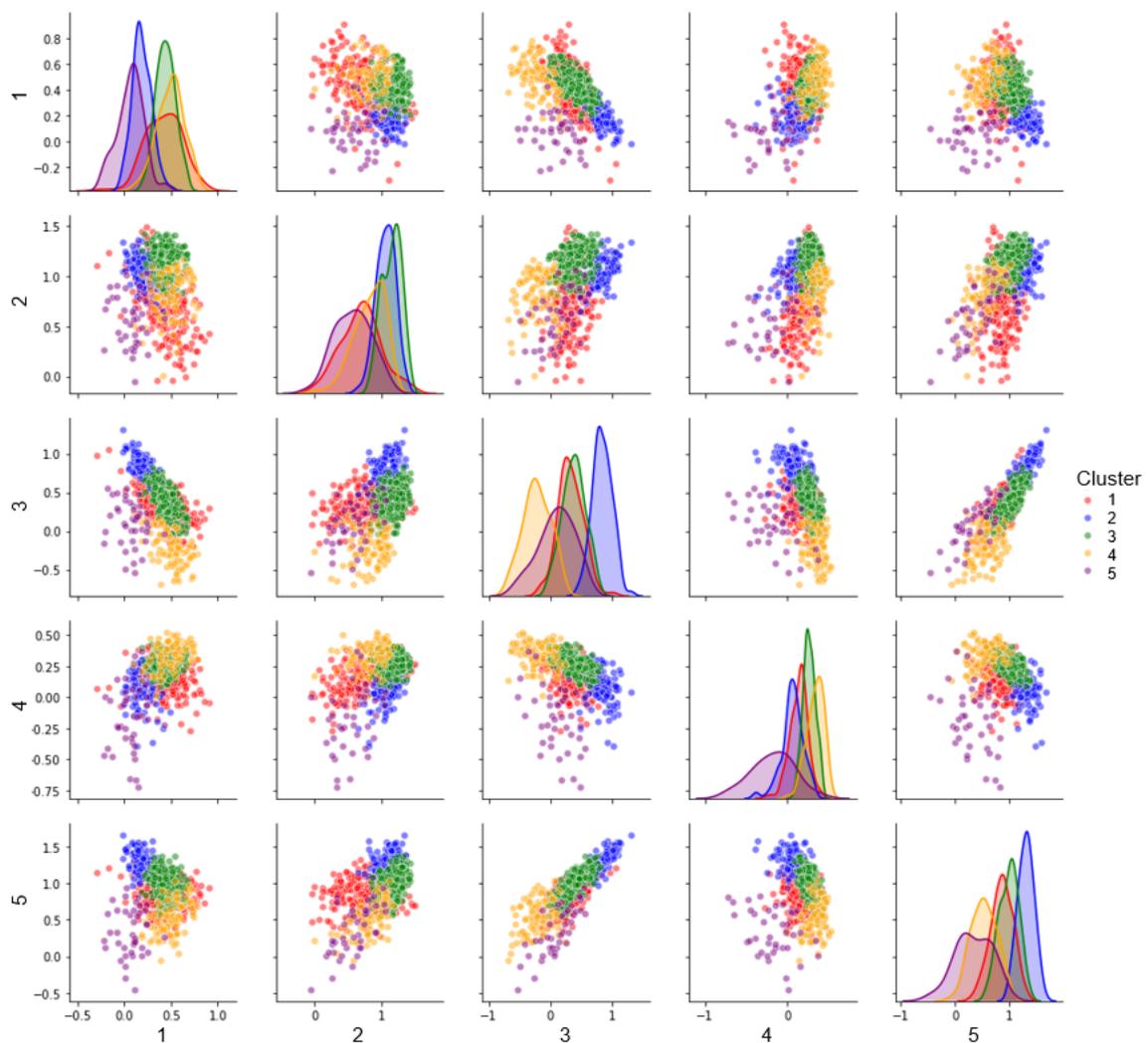


Abbildung 4.15: Paarplot in Abhängigkeit von den fünf Clustern

4.2.4 Interpretation der Cluster

Allgemein gilt: Auch wenn Clusterlösungen statistisch richtig bestimmt wurden, kommen sie nur bei einer inhaltlichen Interpretierbarkeit zur Anwendung. Deshalb werden in diesem Abschnitt die fünf identifizierten Cluster inhaltlich beschrieben und die besonderen Merkmale in der Fahrzeugnutzung diskutiert. Tabelle 4.6 zeigt die fünf Clusterlösungen und die durchschnittlichen Werte der ursprünglich verwendeten 20 Inputvariablen über die drei Dimensionen (Zeit, Fahrtzweck und Distanz). Zur besseren Lesbarkeit wurden vom Autor besondere Eigenschaften der jeweiligen FNPs fett markiert. Die Bezeichnung der FNPs ergibt sich aus den beschriebenen Besonderheiten.

Tabelle 4.6: Merkmalsausprägungen der Originalvariablen der DPGMM-Clusterlösung

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Clustergröße	32 %	16 %	31 %	14 %	6 %
Zeit					
<i>AntStand</i> [1 – 10 km]	70,9 %	70,2 %	89,7 %	77,8 %	29,5 %
<i>AntStand</i> [11 – 30 km]	9,5 %	24,7 %	4,6 %	3,0 %	25,1 %
<i>AntStand</i> [31 – 100 km]	11,8 %	2,4 %	2,6 %	6,2 %	23,7 %
<i>AntStand</i> [> 100 km]	7,8 %	2,7 %	3,0 %	13,1 %	21,8 %
<i>AntWerktag</i>	80,3 %	80,8 %	79,4 %	80,0 %	78,8 %
<i>AntÜbernachtung</i>	94,6 %	97,3 %	97,6 %	90,4 %	73,0 %
<i>AntBewegung</i>	88,0 %	90,0 %	92,6 %	75,6 %	78,8 %
Fahrtzweck					
<i>AntZweck</i> [nach Hause]	35,6 %	38,2 %	44,5 %	37,8 %	3,7 %
<i>AntZweck</i> [zur Arbeit]	12,8 %	25,4 %	13,8 %	8,5 %	10,0 %
<i>AntZweck</i> [Einkaufen]	16,0 %	12,0 %	12,8 %	16,9 %	24,8 %
<i>AntZweck</i> [Freizeit]	19,4 %	13,5 %	16,0 %	18,6 %	31,8 %
<i>AntZweck</i> [Sonstiges]	16,2 %	10,8 %	12,8 %	18,2 %	29,7 %
Distanz					
<i>AntDaydis</i> [1 – 10 km]	6,2 %	3,2 %	7,4 %	26,7 %	11,7 %
<i>AntDaydis</i> [11 – 30 km]	12,6 %	9,3 %	29,5 %	22,8 %	16,8 %
<i>AntDaydis</i> [31 – 100 km]	37,8 %	67,4 %	46,5 %	21,7 %	34,6 %
<i>AntDaydis</i> [> 100 km]	43,4 %	20,0 %	16,7 %	28,8 %	37,0 %
<i>AntHomeDistance</i> [1 – 10 km]	43,4 %	30,0 %	65,9 %	72,9 %	11,1 %
<i>AntHomeDistance</i> [11 – 30 km]	20,2 %	55,9 %	24,7 %	10,1 %	5,7 %
<i>AntHomeDistance</i> [31 – 100 km]	28,4 %	12,3 %	7,2 %	11,2 %	2,0 %
<i>AntHomeDistance</i> [> 100 km]	8,0 %	1,8 %	2,2 %	5,9 %	1,1 %

Cluster 1 – „universales Langstreckenfahrzeug“ (n = 199)

Beim ersten Cluster ist die Verteilung der Fahrtzwecke annähernd (bis auf „nach Hause“) gleich verteilt ($AntZweck[zur\ Arbeit]: 12,8\%$; $AntZweck[Einkaufen]: 16,0\%$; $AntZweck[Freizeit]: 19,4\%$). Des Weiteren sind die hohen Tagesdistanzen gegenüber den restlichen Clustern auffällig. An 43,4% der beobachteten Tage wurden Tagesdistanzen über 100 km zurückgelegt. Dieser sehr hohe Anteil könnte ein Indiz für Fahrzeuge von Vertretern oder Außendienstmitarbeitern sein. Bei 8,0% der Einzelfahrten von zu Hause wurden Distanzen größer 100 km beobachtet. Darüber hinaus werden bei 28,4% der Einzelfahrten von zu Hause eine Distanz von 31–100 km zurückgelegt. Die hohen Distanzen wirken sich auch auf die Standzeiten nach Entfernung vom Wohnort ($AntStand$) aus. Knapp 20% der Standzeit finden außerhalb eines Radius von 30 km vom Wohnort statt. Die hohen Tagesdistanzen spiegeln sich zudem noch in der Komplexität der Fahrzeugnutzung wider. Durchschnittlich werden 4,42 Fahrten pro Tour absolviert. Auch die durchschnittliche Pendeldistanz (Fahrdistanz zwischen approximiertem Wohn- und Arbeitsort) in Höhe von 35 km ist im Vergleich zu den restlichen Clustern am größten.

Cluster 2 – „Mittelstreckenpendelfahrzeug“ (n = 103)

Beim zweiten Cluster dominiert, im Gegensatz zu Cluster 1, der Fahrtzweck „zur Arbeit“. 25,4% der Fahrten entsprechen den Kriterien der Fahrt zur Arbeit. Bei der Dimension Distanz ist das Fahrzeug primär in der Mittelstrecke (31–100 km) unterwegs. An 67,4% der beobachteten Tage werden Tagesdistanzen zwischen 31 km und 100 km zurückgelegt. Die durchschnittliche Pendeldistanz beträgt 18 km. Bei über der Hälfte (55,9%) der Einzelfahrten, welche von zu Hause stattgefunden haben, wurden Distanzen zwischen 11 km und 30 km zurückgelegt. Der Fernverkehr (Fahrten außerhalb von 100 km vom Wohnort) findet hier so gut wie nicht statt. Lediglich 2,7% der Standzeit ist außerhalb des 100-km-Radius. Interessant ist hier noch, dass zwar hauptsächlich Tagesdistanzen zwischen 31 km und 100 km absolviert werden, sich die hauptsächlich Standzeiten ($AntStand[1 - 10km]: 70,2\%$, $AntStand[11 - 30km]: 24,7\%$) allerdings nahe am Wohnort befinden. Möglicherweise handelt es sich bei den Fahrten nahe am Wohnort unter anderem auch um Hol- und Bringdienste.

Cluster 3 – „universales Kurzstreckenfahrzeug“ (n = 196)

Das dritte Cluster hat, ähnlich wie bei Cluster 1, keinen dominierenden Fahrtzweck ($AntZweck[zur\ Arbeit]: 13,8\%$; $AntZweck[Einkaufen]: 12,8\%$; $AntZweck[Freizeit]: 16,0\%$). Im Gegensatz zu Cluster 1 werden die Fahrzeuge in Cluster 3 vor allem für geringe Distanzen eingesetzt. Durchschnittlich wurden an 36,9% ($AntDaydis[1 - 10km]: 7,4\%$; $AntDaydis[11 - 30km]: 29,5\%$) der beobachteten Tage lediglich Tagesdistanzen zwischen 1 km und 30 km absolviert. Auch bei diesem Cluster spielt der Fernverkehr nahezu keine Rolle

in der Fahrzeugnutzung. 89,7 % der Standzeit ($AntStand[1 - 10km]$) findet in einem Radius von 1–10 km vom Wohnort statt. Die starke Fokussierung auf den Wohnort spiegelt sich auch im hohen Anteil (97,6 %) der Fahrzeugübernachtungen zu Hause ($AntÜbernachtung$) wider. Auch die durchschnittlichen Fahrten pro Tour (2,44) und die durchschnittliche Pendeldistanz (10,3 km) weisen die geringsten Werte im Vergleich zu den restlichen Clustern auf.

Cluster 4 – „distanzunabhängiges Freizeitfahrzeug“ (n = 87)

Das vierte Cluster hat den geringsten Anteil von Tagen mit Fahrzeugbewegungen. Durchschnittlich werden die Fahrzeuge in Cluster 4 an nur 75,6 % der beobachteten Tage genutzt (siehe Variable $AntBewegung$). Auch der Anteil der Fahrzeugübernachtungen am Wohnort ($AntÜbernachtung$) ist im Vergleich zu den Clustern 1 bis 3 am geringsten. Des Weiteren ist auffällig, dass die Nutzung von Fahrzeugen in Cluster 4 nur wenig den Kriterien des Fahrtzwecks „zur Arbeit“ erfüllt. Lediglich 8,5 % der Fahrten wurden mit dem Fahrtzweck „zur Arbeit“ deklariert. Des Weiteren ist bei diesem Cluster kein dominantes Tagesdistanzintervall ($AntDaydis[1 - 10km]$: 26,7 %; $AntDaydis[11 - 30km]$: 22,8 %; $AntDaydis[31 - 100km]$: 21,7 %; $AntDaydis[> 100km]$: 28,8 %) vorhanden. Dies könnte bedeuten, dass Fahrzeuge in diesem Cluster sehr variabel, unabhängig von der Distanz, eingesetzt werden. Möglicherweise handelt es sich bei diesen Fahrzeugen um viele Zweitwagen. Trotz der nahezu gleich verteilten Tagesdistanzen sind die Standzeiten sehr auf den Wohnort forciert. 77,8 % ($AntStand[1 - 10km]$) der Standzeit finden in einem Radius von 1–10 km um den Wohnort statt. Wenn ein Fahrzeug von Cluster 4 sich nicht nahe am Wohnort befindet, dann steht es sehr weit entfernt. 13,1 % ($AntStand[> 100km]$) der Standzeit finden außerhalb eines Radius von 100 km statt. Diese Erkenntnis bestätigt zudem den nicht linearen Zusammenhang zwischen Standzeit und Distanz, was die Relevanz der separaten Betrachtung untermauert.

Cluster 5 – „ortsunabhängiges Langstreckenfahrzeug“ (n = 40)

Das fünfte Cluster hat einen auffällig geringen Anteil von Fahrzeugübernachtungen zu Hause. An lediglich 73,0 % ($AntÜbernachtung$) der beobachteten Tage übernachtet das Fahrzeug an seinem identifizierten Wohnort. Zum einen könnte dies daran liegen, dass das Fahrzeug mehrere regelmäßige Standorte hat, an denen das Fahrzeug übernachtet (z. B. Zweitwohnsitz, Übernachtung beim Lebenspartner etc.). Dies hätte zur Folge, dass der Beobachtungszeitraum möglicherweise nicht ausreicht, um den Wohnort des Fahrzeugs eindeutig zu bestimmen. Die verschiedenen Standorte hätten auch Auswirkungen auf die Berechnung des Wohnortes (Median über alle GPS-Koordinaten mit Fahrzeugübernachtungen). Dies würde auch die geringen Anteile von Fahrten von zu Hause ($AntHomeDistance$) erklären. Zum anderen könnte es sich auch um Geschäftsfahrzeuge handeln, welche von unterschiedlichen Haushalten

genutzt werden, was zu wechselnden Übernachtungsorten führen würde. All diese Punkte spiegeln sich sowohl in den nahezu gleich verteilten Anteilen der Standzeiten ($AntStand[1 - 10km]$: 29,5 %; $AntStand[11 - 30km]$: 25,1 %; $AntStand[31 - 100km]$: 23,7 %; $AntStand[> 100km]$: 21,8 %), als auch in den hohen Tagesdistanzen ($AntDaydis[31 - 100km]$: 34,6 %; $AntDaydis[> 100km]$: 37,0 %) wider.

4.3 Zwischenfazit

In diesem Kapitel stand die Erweiterung der in Kapitel 3 erstellten Zeit-Distanz-FNPs um die Dimension des Fahrtzwecks im Mittelpunkt. Deshalb wurde mit Hilfe raumbezogener Sensordaten eine Methodik entwickelt, um Fahrtzwecke zu approximieren. Die Validierung des Ansatzes erfolgte mit fahrzeugbezogenen Umfragedaten (inkl. Fahrtzweck). Die entwickelte Methodik liefert eine einheitliche und flexible Definition von Fahrtzwecken. Allerdings kann die Approximation von nicht regelmäßigen Fahrtzwecken (Einkaufen, Freizeit), gerade in hochverdichteten Räumen mit vielen POIs, Probleme mit sich bringen. Auf der Basis von Zeit, Distanz und Fahrtzweck konnten anschließend fünf FNPs identifiziert werden. Das angewandte Verfahren bringt den Vorteil mit sich, dass die Anzahl der FNPs mit der Anzahl der Fahrzeuge dynamisch steigen kann. Darüber hinaus wurden die methodischen und inhaltlichen Veränderungen der Zeit-Distanz-FNP aus Unterkapitel 3.2 bei der Auswahl des Verfahrens berücksichtigt. Die zugrundeliegenden Fahrzeugdaten mit Raumbezug haben im Vergleich zu den Fahrzeugdaten ohne Raumbezug aus Kapitel 3 eine kleine Stichprobe. Aus diesem Grund können die hier betrachteten Fahrzeuge eine spezielle Nutzung und möglicherweise nicht die generelle Nutzung von Fahrzeugen widerspiegeln. Dies liegt auch daran, dass der ausgewählte Beobachtungszeitraum und die verwendeten Fahrzeuge zufällig für den konzeptionellen Beweis herangezogen wurden. Das Vorgehen zur Approximation von Fahrtzwecken sowie die Identifikation von FNPs unter Berücksichtigung von dynamischen Fahrzeugdaten sind aber wichtige Bausteine für die Ausrollung auf eine größere Fahrzeugflotte. In einem weiteren Schritt ist es nun von Interesse, inwieweit die identifizierten FNPs von bestimmten Faktoren wie Kontext und Fahrzeug (Abbildung 2.1) beeinflusst werden. Aus diesem Grund findet im folgenden Kapitel eine nähere Untersuchung der Einflussfaktoren auf die Fahrzeugnutzung statt.

5 Einflussfaktoren auf die Fahrzeugnutzung

In Abbildung 2.1 wurden über eine Literaturrecherche und eine Anforderungsanalyse die drei Dimensionen der Fahrzeugnutzung und deren Einflussfaktoren dargestellt. Auf der Basis dieser drei Dimensionen und der zugrundeliegenden Daten konnten insgesamt fünf FNP's identifiziert werden. Im nächsten Schritt ist es nun von Interesse, inwieweit bestimmte Faktoren die FNP's beeinflussen. Deshalb werden in Unterkapitel 5.1 die Verknüpfungsmöglichkeiten zwischen der Fahrzeugnutzung und den definierten Einflussfaktoren schematisch vorgestellt. Von den gesamthaften Zusammenhängen ausgehend werden in Unterkapitel 5.2 kontextbezogene Raumstrukturen genauer untersucht und eine länderübergreifende Logik entwickelt, um Urbanität auf Postleitzahl (PLZ)-Ebene zu vergleichen. Die PLZ als räumliche Einheit liefert, speziell im Industriekontext, eine gute Ebene, um unterschiedliche Datenquellen miteinander zu verknüpfen, da viele Informationen (z. B. Wohnort des Kunden) auf einer PLZ-Ebene vorliegen. Abschließend werden in Unterkapitel 5.3 die Zusammenhänge genauer untersucht.

5.1 Überblick über die Einflussfaktoren auf die Fahrzeugnutzung

Anhand der generierten Informationen (z. B. Approximation des Wohnortes) und Informationen direkt vom Fahrzeug (z. B. Fahrzeugtyp) können Verknüpfungen zu den Einflussfaktoren hergestellt werden. Abbildung 5.1 stellt die Zusammenhänge und Verbindungspunkte zwischen der Fahrzeugnutzung im Mobilitätskontext und den Einflussfaktoren (Kontext, Kunde und Fahrzeug) schematisch dar. Darüber hinaus lassen sich Verbindungen zu den in Kapitel 2 erwähnten Kunden- und Fahrzeugsegmentierungen aufbauen. Über den Fahrzeugtyp ist es möglich, auf die Fahrzeugsegmentierung und damit auf das Konzept und Segment des Fahrzeugs zu schließen. Die Fahrzeug-ID, welche für jedes Fahrzeug eindeutig ist, kann als Schlüssel zu Kundeninformationen und Fahrzeugalter fungieren. Über die Kundeninformationen könnte perspektivisch auch das soziale Milieu der jeweiligen Fahrzeugnutzer bestimmt werden. Beispielsweise könnten über Befragungen im Auto Eigenschaften zur Werteorientierung und zum sozialen Status abgefragt werden. Über die Information des approximierten Wohnortes des Fahrzeugs erfolgt die Verknüpfung zur Raumstruktur (z. B. Anzahl der POIs, Zugang zu ÖPNV, Bevölkerungsdichte etc.), Soziodemografie (z. B. Haushaltsgröße) oder dem Wetter.

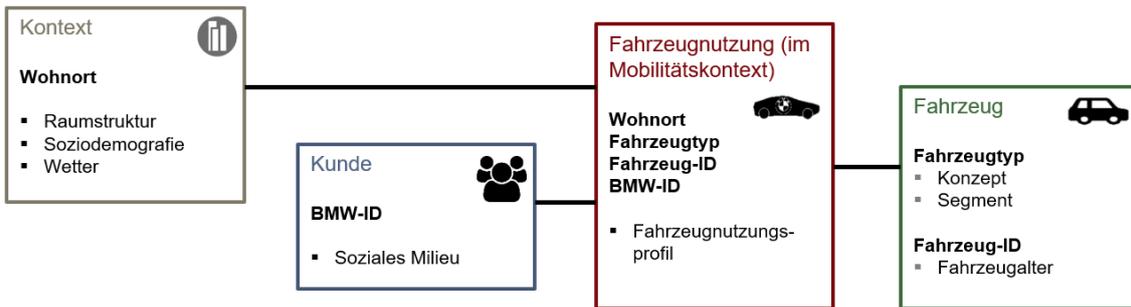


Abbildung 5.1: Schematischer Zusammenhang zwischen den Einflussfaktoren und die Fahrzeugnutzung

Gerade der Einfluss raumstruktureller Gegebenheiten auf die Fahrzeugnutzung findet in der Literatur große Aufmerksamkeit (Dieleman et al., 2002; Feng, Dijst, Prillwitz & Wissink, 2013; Giuliano & Narayan, 2003; Pan et al., 2009). Oftmals beschränkt sich die Betrachtung der Zusammenhänge zwischen Raumstruktur und Fahrzeugnutzung auf einen begrenzten Bereich (z. B. Stadt). Ein flächendeckender und länderübergreifender Vergleich der Raumstruktur auf der Basis einheitlicher Kriterien ist in der Literatur nicht existent. Aus diesem Grund wird im nächsten Unterkapitel eine Methodik entwickelt, um Urbanität im Kontext der Fahrzeugnutzung flächendeckend vergleichbar zu machen. Mit diesem Urbanitätsindex ist es anschließend möglich, die Fahrzeugnutzung im ländlichen und urbanen Raum auf internationaler Ebene zu vergleichen und zu quantifizieren.

5.2 Quantifizierung von Urbanität

Dieses Unterkapitel basiert großteils auf der Veröffentlichung „Spatial Factor – Using a Random Forest Classification Model to Measure an Internationally Comparable Urbanity Index“

(Niklas, von Behren, Soylu et al., 2020).

5.2.1 Motivation

Bereits viele Studien zeigen auf, wie wichtig die räumliche Differenzierung für die Fahrzeugnutzung oder das Mobilitätsverhalten im Allgemeinen ist. Giuliano und Narayan (2003) verdeutlichen in ihrer Untersuchung, dass die Häufigkeit und Variabilität in der Nutzung verschiedener Verkehrsmittel eine Funktion aus individuellen (z. B. Einstellung) und haushaltsspezifischen (z. B. Haushaltsgröße) Merkmalen, den verfügbaren Verkehrsmitteln (z. B. Verfügbarkeit von Fahrzeugen, Führerscheinbesitz), Transportpreisen (z. B. Kosten für Monatsticket), Angebot (z. B. ÖPNV-Qualität) und der Landnutzung (z. B. POIs) sind. Letztere ist die physische Umgebung, in der Menschen leben und ihre Zeit verbringen. Die Nutzung und Abhängigkeit

vom Fahrzeug sowie die Bereitschaft, andere Verkehrsmittel zu verwenden, unterscheiden sich mit der Verfügbarkeit und Qualität verschiedener Faktoren (z. B. Zugang zum ÖPNV, POI etc.).

Der Zusammenhang zwischen Urbanität und Mobilitätsverhalten wird bereits intensiv in der Literatur diskutiert, weil die räumlichen Gegebenheiten einen Einfluss auf die Nutzung verschiedener Verkehrsmittel (z. B. Fahrzeug) haben. Da sich der räumliche Kontext damit auch auf die individuelle Abhängigkeit vom Fahrzeug auswirkt, untersuchen Pan et al. (2009) den Einfluss urbaner Strukturen auf den Fahrzeugbesitz, die Fahrdistanzen und die Verkehrsmittelwahl. Dabei können sie Unterschiede zwischen den vier ausgewählten Nachbarschaften in Schanghai nachweisen. Auch Dieleman et al. (2002) können Zusammenhänge zwischen dem Mobilitätsverhalten und der Wohnumgebung in den Niederlanden identifizieren. Feng et al. (2013) erweitern die Perspektive von einer nationalen zu einer internationalen Ebene, indem das Mobilitätsverhalten zweier Regionen mit unterschiedlichen Raumstrukturen in China (Nanjing) und den Niederlanden (Randstad) verglichen wird. Auch Giuliano und Narayan (2003) untersuchen den Einfluss von niedrig- und hochverdichteten Räumen auf das Mobilitätsverhalten in den Vereinigten Staaten von Amerika (USA) und im Vereinigten Königreich (UK). Hier stellt sich heraus, dass in hochverdichteten Räumen weniger Fahrten im Vergleich zu niedrigverdichteten Räumen gemacht werden. Des Weiteren gibt es einen Zusammenhang zwischen der Qualität des ÖPNV und der Bereitschaft, das Fahrzeug abzuschaffen. Newman, Wells, Donovan, Nieuwenhuis und Davies (2014) untersuchen die Nutzbarkeit von BEVs in bestimmten Raumtypen und zeigen auf, dass vor allem in ländlichen und suburbanen Gebieten das BEV zukünftig eine wichtige Rolle spielen könnte. Speziell in ländlicheren Gebieten ist es eher wahrscheinlich, dass Haushalte die nötige Infrastruktur für privates Laden haben. Aultman-Hall, Sears, Dowds und Hines (2012) untersuchen die Nutzung von Fahrzeugen mit Hybridantrieb (PHEV)s sowie reinen BEVs. Hierbei bestätigt sich, dass vor allem in ländlichen Gebieten PHEVs einen Vorteil haben, da dort auch längere Distanzen zurückgelegt werden müssen. Feng et al. (2013) bieten zudem einen Überblick über Studien, welche sich mit dem Zusammenhang von Raumstruktur und Mobilitätsverhalten auf nationaler und internationaler Ebene beschäftigen. Allerdings sind die aufgezeigten Studien und ihre Untersuchungsgebiete sehr selektiv (z. B. Stadtebene) und können international nur eingeschränkt verglichen werden. Zudem ist oftmals die Übertragbarkeit der angewandten Methoden schwierig, da nur lokal verfügbare Daten genutzt werden. Das bietet im Einzelfall die Möglichkeit sehr detailliert zu differenzieren, schafft aber das Problem der Übertragbarkeit. Bereits Giuliano und Narayan (2003) identifizieren eine Forschungslücke hinsichtlich der einheitlichen Messung des Zusammenhangs zwischen Mobilitätsverhalten und Raumstruktur über verschiedene Länder. Ziel dieses Unterkapitels ist es, diese Lücke zu schließen und einen Urbanitätsindex (UI) zu entwickeln, damit Flächen in Abhängigkeit von ihren räumlichen Gegebenheiten bewertet

werden können. In Konsequenz ist es dann möglich, die Fahrzeugnutzung differenziert nach Urbanität einheitlich zu betrachten. Basierend auf existierender Literatur (Abschnitt 5.2.2) wurden vier Anforderungen an die Entwicklung des UI definiert:

1. Berücksichtigung von Einflussfaktoren auf das Mobilitätsverhalten
2. Internationale Vergleichbarkeit
3. Reproduzierbarkeit und praktische Anwendbarkeit im industriellen Kontext
4. Skalierbare Methodik

Um die internationale Vergleichbarkeit zu garantieren, wurde bereits zu Entwicklungsbeginn darauf geachtet, dass für die ausgewählten Länder möglichst gleiche Daten und einheitliche Kriterien verwendet werden.

Für die Gewährleistung der praktischen Anwendbarkeit unserer Daten und deren Erkenntnisse wurden als geografische Grenzen die PLZs verwendet. Die Aggregation der Daten auf PLZ-Ebene ist ein guter Kompromiss zwischen Detaillierungsgrad und der Möglichkeit, verschiedene Datenquellen (z. B. bestehende Mobilitäts- und Marketingstudien) miteinander zu verknüpfen. Existierende Studien berücksichtigen oftmals kleinräumigere Gebietseinteilungen. Auf der einen Seite kann dadurch die Heterogenität innerhalb einer PLZ besser abgebildet werden, auf der anderen Seite ist es schwierig, solch kleinräumige Gebietseinteilungen in der Praxis zu nutzen und mit anderen Datenquellen zu verknüpfen. Darüber hinaus existieren in der Praxis viele Informationen auf PLZ-Ebene, weshalb diese als universaler Verknüpfungsschlüssel geeignet ist. Um die Reproduzierbarkeit und Konsistenz des Ansatzes zu gewährleisten, wurden hauptsächlich frei zugängliche Daten verwendet.

Damit die Skalierbarkeit des Ansatzes berücksichtigt wird, wurde die Beziehung zwischen den geografischen Daten und den vier Raumtypen mit Hilfe einer überwachten Lernmethode, eines sogenannten Random-Forest-Modells, erlernt. Gianotti, Getson, Hutyra und Kittredge (2016) nutzen bereits ein Random-Forest-Modell, um Urbanität für bestimmte Gebiete zu quantifizieren. Ein Random-Forest-Modell hat verschiedene Vorteile gegenüber anderen überwachten Lernmethoden. Zum einen müssen die vorliegenden Daten kaum für die Modellierung angepasst werden, da keinerlei Standardisierungen oder Verteilungsannahmen notwendig sind. Zum anderen ist es möglich, den Einfluss verschiedener Variablen auf die Kategorisierung zu beobachten und die Variablen mit dem höchsten Erklärungsgehalt zu bestimmen. Das trainierte Random-Forest-Modell wird dann verwendet, um nicht kategorisierte PLZs in einen der vier Raumtypen einzuteilen. Nach der Erstellung des UI ist es möglich Deutschland und Frankreich auf der Basis einer einheitlichen Urbanitätsdefinition miteinander zu vergleichen und weitere Informationen über die PLZs zu verknüpfen.

Zur Generierung der notwendigen geografischen Daten für die Entwicklung des UI wurden Informationen hinsichtlich Erreichbarkeit, Dichte und Variabilität aus unterschiedlichen Datenquellen kombiniert und auf PLZ-Ebene aggregiert. Anschließend fand eine Einteilung bekannter PLZs in einen der vier Raumtypen superurban, urban, suburban/kleinstädtisch, ländlich durch Experten (Raumplaner, Mobilitätsforscher, Mobilitätsanalysten etc.) statt. Die grundlegende Annahme des UI ist, dass sich die Fahrzeugnutzung bzw. das Mobilitätsverhalten über die vier Raumtypen unterscheidet. Existiert beispielsweise in ländlichen Regionen nahezu kein ÖPNV, sind Individuen auf ihren Pkw angewiesen und damit auch abhängig von diesem. Im Kontrast dazu gibt es in urbanen Gebieten Alternativen zum Pkw und viele POIs, um die täglichen Bedürfnisse zu befriedigen. Mit dem UI ist es nun möglich, GPS-basierte Fahrzeugnutzungsdaten zu verknüpfen und die Unterschiede zwischen den Raumtypen zu untersuchen. Aufgrund der internationalen Vergleichbarkeit ist die Quantifizierung auch über mehrere Länder möglich.

Abbildung 5.2 stellt den Prozess zur Entwicklung des UI von der Literaturübersicht, Datensammlung und Klassifikation von PLZs über die angewandte Methode bis hin zu den Ergebnissen dar.

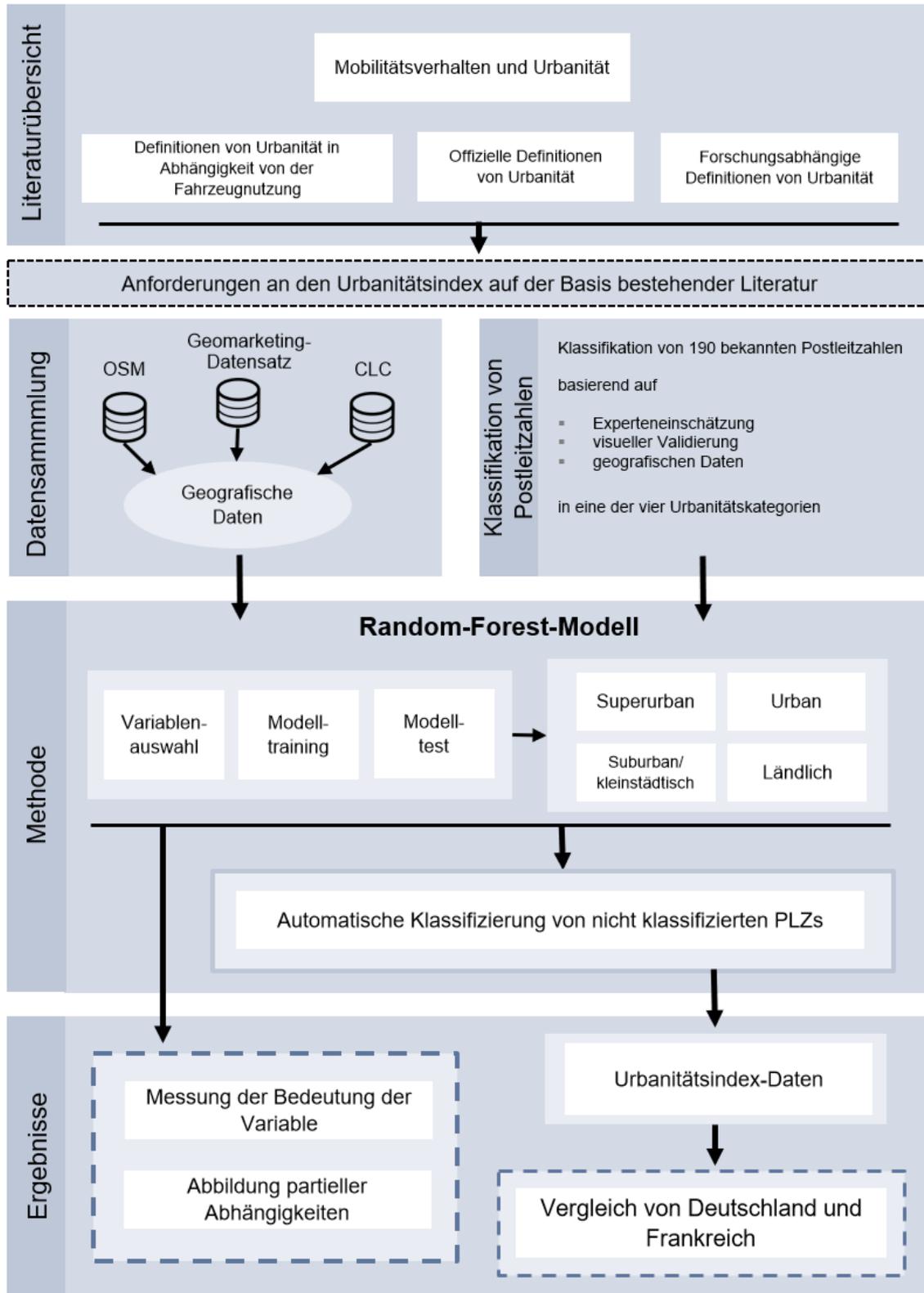


Abbildung 5.2: Prozess der Entwicklung des Urbanitätsindex

5.2.2 Literaturübersicht

Für die Entwicklung eines Urbanitätsindex mit dem Ziel, die Fahrzeugnutzung räumlich differenziert zu betrachten, ist es in einem ersten Schritt wichtig zu verstehen, wie Urbanität in existierenden Arbeiten definiert wird. Hierbei werden nicht nur die verschiedenen Definitionen im Kontext des Pkw-Mobilitätsverhaltens, sondern auch forschungsabhängige (z. B. für Bewirtschaftung natürlicher Ressourcen) oder offizielle (z. B. von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD)) Auffassungen betrachtet. Dadurch kann ein umfangreicher Überblick zur Abbildung von Urbanität generiert werden. Tabelle 5.1 fasst die Ergebnisse der Literaturrecherche zusammen. Dabei werden das entsprechende Untersuchungsgebiet, die räumliche Einheit sowie die Kriterien zur Definition von Urbanität aufgezeigt. Mit Hilfe dieser Recherche können die Anforderungen an die Entwicklung eines Urbanitätsindex bestimmt werden.

Giuliano und Narayan (2003) weisen bereits in ihrer Arbeit darauf hin, dass eine länder- oder städteübergreifende Messung der Zusammenhänge zwischen Urbanität und Mobilitätsverhalten schwierig ist. Dieses Problem ist primär durch Dateneinschränkungen (z. B. lokale Verfügbarkeit) zu erklären. In Konsequenz ist die Definition von Urbanität zwischen Städten oder Ländern nicht konsistent. Das kann wiederum zu Verzerrungen in der Messung von Urbanität und damit zu ungenauen Ergebnissen führen. Motte-Baumvol, Massot und Byrd (2010) untersuchen die räumliche Autoabhängigkeit in den Vororten von Paris (Frankreich) und definieren drei Raumtypen über POIs. POIs beschreiben hier die Anzahl lokaler Angebote, Geschäfte oder Dienstleistungen. Mit diesen Informationen wird auf Ebene von Gemeinden die räumliche Autoabhängigkeit bestimmt. Damit können die Autoren Gemeinden in die vier Kategorien der Autoabhängigkeit einteilen: schwach, medium, stark und sehr stark. Fehlen bestimmte Angebote, sind die Bewohner dazu gezwungen, in benachbarte Gemeinden zu fahren, um die Angebote zu erwerben. Gebiete mit hoher Autoabhängigkeit zeichnen sich durch eine hohe Motorisierungsrate und eine hohe Verkehrsleistung aus (Motte-Baumvol et al., 2010). Auch Siedentop, Roos und Fina (2013) definieren die räumliche Autoabhängigkeit in Regionen um Stuttgart (Deutschland), basierend auf fehlenden Alternativen zum Pkw. Dabei wird ein Konzept entwickelt, welches für jedes Wohngebäude die Qualität des ÖPNV sowie die Erreichbarkeit von POIs mit nicht motorisierten Verkehrsmitteln misst. Ein Algorithmus wurde verwendet, um die Distanz von jedem Wohngebäude zum nächsten POI (ÖPNV-Haltestelle, Schule, Arzt oder Apotheke) zu berechnen. Die Autoren unterteilen dabei die Autoabhängigkeit in vier Kategorien: niedrig, niedrig-mittel, mittelhoch und sehr hoch. Sowohl Motte-Baumvol et al. (2010) als auch Siedentop et al. (2013) untersuchen die Beziehungen zwischen Raumstrukturen und Mobilitätsmustern für spezifische Regionen und verwenden gleichermaßen vier Kategorien zur Beschreibung der Autoabhängigkeit.

Über die Definitionen von Urbanität im Kontext des Pkw-Mobilitätsverhaltens hinaus gibt es Bestrebungen, Urbanität für bestimmte Bereiche zu definieren. Die unterschiedlichen Definitionen sind auch erforderlich, denn je nach Anwendungsfall werden verschiedene Aspekte von Urbanität betrachtet (Gianotti et al., 2016; Ögdül, 2010). Beispielsweise untersucht Ögdül (2010) Urbanität auf der räumlichen Bezugseinheit „Nomenclature des unités territoriales statistiques (NUTS) 4“ für die Türkei. Dabei verwendet er eine zweistufige Clusteranalyse und unterteilt die Türkei in (1) überwiegend ländliche Bereiche, (2) überwiegend urbane Bereiche und sogenannte (3) Übergangsbirke. Soylu (2017) analysiert die Instabilität von Urbanität über die Zeit und identifiziert sieben verschiedene Formen von Urbanität in Istanbul. Ein gänzlich anderer Ansatz, um die Veränderungen der Landschaft in China zu beobachten, stammt von Arellano und Roca (2018). Hierbei charakterisiert sich Urbanität über die Anzahl der Nachtbeleuchtungen in einem bestimmten Gebiet. Ein Vorteil dieses Ansatzes besteht in der Verwendung von weitestgehend frei verfügbaren Daten (U.S. Air Force Defense Meteorological Satellite Program Operational Linescan System), wodurch die Ergebnisse reproduzierbar sind. Gianotti et al. (2016) untersuchen Urbanität in Boston (USA) aus einer ökologischen Perspektive, um die Bewirtschaftung natürlicher Ressourcen zu verbessern. Dabei findet eine Umfrage via E-Mail statt, in welcher 314 Landbesitzer befragt werden, ob sie ihren Besitz als „urban“, „suburban“ oder „ländlich“ einschätzen. Mit Hilfe eines Entscheidungsbaums wurde die Beziehung zwischen der Einschätzung der Landbesitzer und den geografischen Informationen erlernt. Danach konnten automatisch weitere Gebiete jeweils einem Raumtyp zugeordnet werden.

Abseits forschungsabhängiger Definitionen von Urbanität gibt es eine Vielzahl von Ansätzen von offiziellen Institutionen. Hierbei liegt der Fokus vor allem auf der Betrachtung von Urbanität auf nationaler Ebene. Primär wird dabei Urbanität über Bevölkerungskriterien definiert. Eine weitverbreitete Definition stammt von der OECD (Brezzi, Dijkstra & Ruiz, 2011). Dabei wurden Gebiete in drei Kategorien (überwiegend städtisch, mittelmäßig ländlich und überwiegend ländlich) eingeteilt. Dijkstra und Poelman (2014) liefern eine harmonisierte Definition von Urbanität basierend auf Bevölkerungskriterien über 31 Länder in einer Auflösung von einem Quadratkilometer. Die UN (2018) schätzen in ihrem jährlichen Bericht die Veränderung der ländlichen und städtischen Bevölkerung von über 200 Ländern. Aufgrund länderspezifischer Definitionen von Urbanität und inkonsistenter Datenquellen ist ein direkter Vergleich zwischen den Ländern aber nicht möglich (Arellano & Roca, 2018).

Zusammenfassend kann gesagt werden, dass ein Interesse an Untersuchungen zur Urbanität und der Fahrzeugnutzung oder dem Mobilitätsverhalten gegeben ist. Allerdings mangelt es an einer konsistenten Definition von Urbanität im Kontext des Mobilitätsverhaltens, speziell auf internationaler Ebene. Einige Studien adressieren dieses Problem, allerdings nur auf regionaler Ebene. Ein Ansatz zur internationalen Vergleichbarkeit würde darüber hinaus auch

praktische Vorteile für Mobilitätsanbieter oder Automobilhersteller mit sich bringen, um den Zusammenhang zwischen Fahrzeugnutzung und Raumstrukturen besser zu verstehen. Spezifische Definitionen anderer Forschungsbereiche sind oftmals sehr detailliert und deshalb nicht skalierbar für eine Anwendung auf internationaler Ebene. Offizielle Definitionen sind zwar auf internationaler Ebene verfügbar, allerdings wird hier Urbanität lediglich über die Bevölkerung definiert. Im Kontext der Fahrzeugnutzung ist es aber auch wichtig, Faktoren wie den Zugang zum ÖPNV, POIs sowie die Landnutzung zu berücksichtigen.

Tabelle 5.1: Literaturübersicht über Definitionen von Urbanität

Autor (Jahr)	Untersuchungsgebiet	Räumliche Einheit	Kriterien
Pkw-mobilitätsabhängige Definitionen			
Motte-Baumvol et al. (2010)	Paris (FR)	Kommune	POI
Siedentop et al. (2013)	Stuttgart (DE)	Siedlungsgebiet	ÖPNV/POI
Forschungsabhängige Definitionen			
Gianotti et al. (2016)	Boston (USA)	1–2,5 km ²	Bevölkerung, Landnutzung, Landqualität
Ögdül (2010)	Türkei	NUTS4	Bevölkerung, Landschaft, Ökonomie
Soylu (2017)	Istanbul (TR)	Nachbarschaft	Soziodemografie
Arellano und Roca (2018)	China	30 m/pixel	Nachtbeleuchtung, Industriestruktur
Offizielle Definitionen			
Brezzi et al. (2011)	30 Länder	Land	Bevölkerung
Dijkstra und Poelman (2014)	31 Länder	1 km ²	Bevölkerung
UN (2018)	200 Länder	Land	Bevölkerung

Deshalb ist es nun das Ziel, ein Verständnis der Abhängigkeiten des Mobilitätsverhaltens und der Urbanität zu generieren. Gleichzeitig werden eine einheitliche Definition sowie die Verwendung konsistenter Datenquellen berücksichtigt, um einen direkten Ländervergleich zu ermöglichen. Zudem werden Einflusskriterien hinsichtlich der Fahrzeugnutzung bei der Charakterisierung von Urbanität verwendet. Bei der Generierung der relevanten Daten und

um Reproduzierbarkeit zu garantieren, werden die Ausführungen von Gianotti et al. (2016) und Xu, Ye und Xu (2017) aufgegriffen und hauptsächlich frei zugängliche Daten (z. B. OSM-Daten) genutzt.

5.2.3 Daten

Wie bereits in der Literaturübersicht dargestellt, gibt es für die Erhebung und Definition von räumlichen Strukturen unterschiedliche Betrachtungsweisen und Ansätze. In Konsequenz gibt es auch keine vergleichenden Analysen, welche Unterschiede hinsichtlich räumlicher Strukturen über verschiedene Länder in einer feinen Auflösung (z. B. PLZ) aufzeigen. Ziel in dieses Abschnitts ist es, einen methodischen Ansatz zu entwickeln, welcher es erlaubt, räumliche Strukturen auf PLZ-Ebene zu identifizieren und länderübergreifend miteinander zu vergleichen.

Datensammlung und -bearbeitung

Zu den größten Herausforderungen bei Studien zur Urbanität zählen die Definition, Messung und Quantifizierung von Urbanität unter Verwendung von konsistenten, international vergleichbaren Datensätzen. Die Auswahl der entsprechenden Datensätze erfordert umfangreiche Recherchen sowie Qualitätsprüfungen. Im Ergebnis wurden für die Entwicklung des UI drei Datenquellen verwendet: OSM-Daten, Geomarketing-Datensatz sowie Corine-Land-Cover (CLC) Daten (Abbildung 5.2).

Über OSM ist es möglich, kostenlos für die ausgewählten Länder von Mai bis September 2018 sogenannte Shapefiles (geografisches Datenformat) über den Anbieter Geofabrik herunterzuladen. OSM-Daten stellen die physischen Elemente einer Landschaft, welche durch Knoten, Linien und Flächen repräsentiert werden, dar. Knoten definieren sich über den Längen- und Breitengrad im Raum (z. B. POI). Eine Linie ist die geometrische Darstellung des geografischen Verlaufs im Gelände (z. B. Straßen). Flächen sind ausgefüllte Polygone, welche sich durch die Fläche auf der Basis mehrerer Linien definieren (z. B. die Fläche eines Gebäudes). Jedes physische Element ist noch mit verschiedenen Informationen gekennzeichnet (Beispiel: POI ist ein Restaurant). Für die Entwicklung des UI wurden Knotenelemente aus dem OSM-Datensatz verwendet.

Die digitalen Grenzen, speziell für PLZs, sind nicht für alle Länder auf der Welt frei verfügbar. Der Geomarketing-Datensatz stellt die digitalen Grenzen der PLZ-Flächen in Shapefiles bereit. Zusätzlich sind noch soziodemografische Informationen, wie Bevölkerung, Kaufkraft sowie Haushaltsgröße, enthalten. Die administrativen Grenzen der PLZs besitzen, abhängig vom betrachteten Land, zwischen zwei und sieben Ziffern. Beispielsweise hat in Deutschland jede PLZ fünf Ziffern (z. B. 86720). Die erste Ziffer steht für die geografische Zone (z. B. Bayern) und die zweite Ziffer für die Region (z. B. Donau-Ries). Innerhalb der Region erhält jede

Gemeinde eine Zahlenkombination. Die gleiche Logik nutzt auch Frankreich. Im Gegensatz dazu werden beispielsweise in UK fünf bis sieben alphanumerische Zeichen (Zahlen und Buchstaben wie z. B. AA9A 9AA) verwendet.

Der aktuellste gesamteuropäische Datensatz zur Landnutzung/Flächenbedeckung (LU/LC) wurde von CLC 2018 zur Verfügung gestellt. Arsanjani, Zipf, Mooney und Helbich (2015) definieren die bereitgestellten Karten wie folgt: „LU maps illustrate human activities, such as artificial surface construction, that represent the usage of land; whilst LC maps display the physical and biological cover over the land surface regardless of the purpose for which they are used“. Darüber hinaus betonen sie die wichtige Rolle der LU-/LC-Datensätze für verschiedenste Bereiche (z. B. Stadt- und Regionalplanung sowie Politikgestaltung), was ein weiterer Grund für die Verwendung dieses Datensatzes ist. Der CLC-2018-Datensatz ist in Zellen mit einer Auflösung von 100 m² eingeteilt. Jedes Element enthält Informationen zur Flächenbedeckung, welche in fünf Klassen unterteilt wird: (1) künstliche Flächen, (2) landwirtschaftliche Flächen, (3) wald- und naturnahe Gebiete, (4) Feuchtgebiete sowie (5) Gewässer (Kosztra, Büttner, Hazeu & Arnold, 2017).

Nachdem die Daten validiert, konsolidiert und vorverarbeitet worden sind, findet im nächsten Schritt, zur weiteren Verarbeitung, die Integration in die GIS-Umgebung (ArcGIS 10.6) statt. Insgesamt beinhalten die geografischen Daten für Deutschland und Frankreich ca. 14 250 Polygone (PLZ-Flächen) mit ca. 1,25 Millionen Knoten.

Bestimmung der Variablen zur Messung von Urbanität

Basierend auf den drei integrierten Datensätzen kann nun eine Reihe von Variablen ausgewählt werden, um Urbanität zu definieren. In der vorliegenden Studie wird Urbanität über verschiedene Arten von bebauten und unbebauten Flächen (z. B. Gebäude, Verkehrsinfrastruktur, Grünflächen) definiert. Je nachdem wie bebaut bzw. unbebaut eine Fläche ist, liefert dies einen ersten Hinweis auf den Grad der Urbanität. Darüber hinaus sind die Art (Industrie- bzw. Wohnfläche) und Nutzung (POIs) der bebauten Fläche für die Bestimmung von Urbanität von Relevanz. Das angewandte Verfahren ermöglicht es zusätzlich, die Variablen mit dem größten Einfluss auf die Klassifizierung zu identifizieren (Abschnitt 5.2.4). Die verschiedenen Stufen von Urbanität im Kontext der Fahrzeugnutzung definieren sich über die Erreichbarkeit des ÖPNV sowie über die Dichte (bebaute Fläche, Industriefläche, Bevölkerung) und Variabilität (POIs) einer Fläche. Der dargestellte Ansatz orientiert sich an den Arbeiten von Cervero und Kockelman (1997). Dabei definiert sich die bebaute Umwelt über Dichte, Diversität und Design (3D). Die Annahme bei dem hier dargestellten Ansatz ist, dass das Mobilitätsverhalten durch die Verfügbarkeit von ÖPNV beeinflusst wird: Je besser die Erreichbarkeit und Dichte, desto höher ist der Grad der Urbanität.

Um diese Faktoren zu berücksichtigen, wurden für jede PLZ im Untersuchungsgebiet die entsprechenden Variablen berechnet (Tabelle 5.2). Die **Erreichbarkeit** einer PLZ wird durch die ÖPNV-Qualität dargestellt, welche sich durch die Anzahl von Bushaltestellen und Haltestellen im Schienenverkehr definiert. Zur Normalisierung werden die Haltestellen durch die PLZ-Fläche dividiert. Haltestellen im Schienenverkehr in einer PLZ werden mit einem Faktor von 20 gewichtet. Sind keine Haltestellen im Schienenverkehr in der PLZ vorhanden, werden die Bushaltestellen mit dem Faktor 3 multipliziert. Die stärkere Gewichtung der Schienen stimmt auch mit der Literatur überein. Beispielsweise bestätigen Scherer und Dziekan (2012) über persönlich geführte Interviews, dass bei gleicher Fahrtzeit der Schienenverkehr gegenüber dem Busverkehr bevorzugt wird. Die generierten Informationen stammen aus den OSM-Daten und dem Geomarketing-Datensatz. Die **Dichte** wird durch den Anteil der bebauten Fläche an der Gesamtfläche, den Anteil der Industriefläche an der bebauten Fläche sowie über die Bevölkerungsdichte auf der bebauten Fläche charakterisiert. Die Daten stammen aus dem CLC-2018-Datensatz. Die **Variabilität** einer PLZ wird durch die POIs auf die bebaute Fläche beschrieben. Diese werden nach den für die Studie relevanten POIs gefiltert (z. B. Restaurants, Schulen etc.). Die Generierung der POIs für das Untersuchungsgebiet findet über OSM-Daten statt.

Tabelle 5.2: Variablen zur Definition von Urbanität

Variable	Definition/Berechnung	Datenquelle
Erreichbarkeit		
<i>PLZ-Fläche</i>	Gesamtfläche (km ²)	Geomarketing (2016)
<i>Bus</i>	Anzahl der Bushaltestellen	OSM (2018)
		Geomarketing (2016)
<i>Schiene</i>	Anzahl der Schienenhaltestellen	OSM (2018)
		Geomarketing (2016)
ÖPNV-Qualität	$(20 \cdot \text{Schiene} + \text{Bus}) / \text{PLZ-Fläche},$ wenn <i>Schiene</i> > 0 $(3 \cdot \text{Bus}) / \text{PLZ-Fläche},$ wenn <i>Schiene</i> = 0	
Dichte		
<i>Wohnfläche</i>	Wohnfläche (km ²)	CLC-2018
<i>Industriefläche</i>	Industriefläche (km ²)	CLC-2018
<i>Bebaute Fläche</i>	<i>Wohnfläche</i> + <i>Industriefläche</i>	
Anteil bebaute Fläche	<i>bebaute Fläche</i> / <i>PLZ-Fläche</i>	
Anteil Industriefläche	<i>Industriefläche</i> / <i>Bebaute Fläche</i>	
Bevölkerungsdichte bezogen auf die bebaute Fläche	<i>Bevölkerung</i> / <i>bebaute Fläche</i>	
Variabilität		
<i>POI</i>	Anzahl der POIs gefiltert nach relevanten Kategorien	OSM (2018)
POIs bezogen auf die bebaute Fläche	<i>Anzahl POIs</i> / <i>bebaute Fläche</i>	

Die Größe der PLZ-Flächen kann sich zwischen den PLZs unterscheiden. Je urbaner der Raum, desto kleiner werden die PLZ-Flächen. Im Umkehrschluss sind ländliche PLZs großflächiger. Zum einen ist das ein Vorteil, da oftmals urbane Gebiete im Fokus politischer Diskussionen stehen und dafür eine granulare Untersuchung möglich ist, zum anderen ein Nachteil, da es die Vergleichbarkeit erschwert. Um dem entgegenzuwirken, werden die Flächen und Dichten standardisiert, indem sie durch die (bebaute) Fläche dividiert werden.

Nach der Berechnung der Variablen auf PLZ-Ebene aus den verschiedenen Datenquellen wird im nächsten Schritt die Zielvariable generiert und PLZs durch Experten in vier Raumtypen (superurban, urban, suburban/kleinstädtisch, ländlich) eingeteilt.

5.2.4 Methodik

In diesem Abschnitt wird als Erstes der PLZ-Klassifizierungsprozess beschrieben. Im Gegensatz zur bisherigen methodischen Herangehensweise werden für die Entwicklung des Index bekannte PLZs durch Experteneinschätzungen nach ihrer Urbanität bewertet, um darauf aufbauend ein Random-Forest-Modell für die automatisierte Zuteilung weiterer PLZs zu generieren. Durch dieses Vorgehen kann zum einen die subjektive Wahrnehmung von Urbanität durch die Experten genutzt und unter Verwendung der geografischen Informationen je PLZ objektiviert werden. Zum anderen können die besonderen Eigenschaften des verwendeten Random-Forest-Modells angewandt werden, um die Abhängigkeiten und die Bedeutung der einzelnen Variablen für die Raumtypen darzustellen.

PLZ-Klassifizierung in vier Raumtypen

Bei den offiziellen Definitionen von Urbanität (Abschnitt 5.2.2) findet die Trennung urbanerer Räume von weniger urbanen Räumen über Bevölkerungskriterien statt. Durch die niedrigdimensionale Betrachtung ist eine lineare Trennung durchführbar. Der UI berücksichtigt aber fünf Faktoren, wodurch eine lineare Trennung nicht mehr ohne Weiteres möglich ist. Beispielsweise könnten PLZs mit einer geringen Bevölkerungsdichte, aber mit vielen POIs als genauso urban eingestuft werden wie PLZs mit einer hohen Bevölkerungsdichte und wenigen POIs. Darüber hinaus ist eine manuelle Einteilung aller PLZs im Sinne der Skalierbarkeit ineffizient. Aus diesem Grund findet die Zuordnung der PLZ in die vier Raumtypen über die Einschätzung von Experten statt. Dieser Ansatz wird auch in Studien aus anderen Fachbereichen zur Klassifizierung genutzt. Beispielsweise nutzen Gorges, Öztürk und Liebich (2019) einen ähnlichen Ansatz für die Erkennung von Sonderereignissen (z. B. Bordsteinüberfahrt, Schlaglochdurchfahrt) im Kontext der Betriebsfestigkeit von Motorrädern. Dabei teilt ein Testfahrer Fahrmanöver in zwei Kategorien – sanft und stark – ein.

Basierend auf den Ansätzen von Gorges et al. (2019) und Gianotti et al. (2016) haben für die vorliegende Studie Experten ihnen bekannte PLZs in die vier Kategorien (superurban, urban, suburban/kleinstädtisch und ländlich) eingeteilt. Die Anzahl an Kategorien wird durch existierende Literatur festgelegt. Siedentop et al. (2013) und Motte-Baumvol et al. (2010) verwenden vier Kategorien, um Unterschiede in der Autoabhängigkeit zu definieren. Gianotti et al. (2016) nutzen lediglich drei Kategorien, wobei hier nicht zwischen superurban und urban

unterschieden wird. Gleichzeitig werden auch verschiedene Verfahren (z. B. hierarchisches Clustering) zur Bestimmung der optimalen Anzahl an Kategorien angewandt. Auch hier stellt sich die optimale Anzahl von Kategorien auf drei bis vier Kategorien ein.

Die Einteilung von PLZs in die vier Kategorien wurde während eines Eintagesworkshops (27.09.2018) abgewickelt. Beim Workshop nahmen zehn Personen teil. Die vertretenen Disziplinen waren ein Raumplaner einer deutschen Universität und zwei Mobilitätsforscher des MOP. Darüber hinaus gehörten zwei Mobilitätsanalysten, drei Experten, welche sich mit dem Thema Mobilität als Dienstleistung intensiv beschäftigen, sowie zwei Experten aus dem Bereich des maschinellen Lernens eines deutschen Automobilherstellers zum Teilnehmerkreis. Die Zusammensetzung der Beteiligten garantiert sowohl die Erfüllung des wissenschaftlichen Anspruchs als auch den Transfer in die Praxis. Der Workshop setzte sich aus drei Bausteinen zusammen. Im ersten Teil wurden verschiedene Ansätze zur Definition von Urbanität mit Blick auf die bestehende Literatur und die vorhandenen Daten präsentiert. Der zweite Teil bestand aus der Darstellung verschiedener Methodiken (regelbasierter Ansatz, unüberwachtes Lernen, überwachtes Lernen durch Experteneinschätzung) zur Messung von Urbanität sowie der Darstellung deren Vor- und Nachteile. Basierend auf diesen Diskussionen und den vier definierten Anforderungen (Berücksichtigung von Einflussfaktoren auf das Mobilitätsverhalten, internationale Vergleichbarkeit, Reproduzierbarkeit und praktische Anwendbarkeit im industriellen Kontext, Skalierbarkeit) wurde der Ansatz des überwachten Lernens mit Experteneinschätzungen gewählt. Als Erstes fand die Auswahl der zu klassifizierenden PLZs statt. Im nächsten Schritt wurden die PLZs nach ihrer Urbanität in absteigender Reihenfolge sortiert und den Raumtypen zugewiesen. Da die Experten aus Deutschland stammen, werden in Deutschland 160 PLZs nach ihrer Urbanität bewertet. Außerhalb Deutschlands werden insgesamt 30 bekannte PLZs in Spanien (8), Frankreich (4), dem UK (14) sowie Italien (4) den Kategorien zugeordnet. Um die Subjektivität der Einschätzung von Experten zu reduzieren, findet zusätzlich eine Validierung der Einteilung statt. Zum einen werden über Google Maps die Einschätzungen visuell überprüft und zum anderen wird eine datenbasierte Verifizierung der gekennzeichneten PLZs mit den gesammelten Daten von OSM und CLC-2018 durchgeführt.

Schätzung mit einem Random-Forest-Modell

Zur Klassifikation von PLZs in die genannten Raumtypen wurde eine überwachte Lernmethode namens Random-Forest-Modell verwendet. Ein großer Vorteil von überwachten Lernmethoden ist die Skalierbarkeit, d. h., es können automatisch weitere (unbekannte) PLZs auf der Basis der erlernten Entscheidungsfunktion in einen der vier Raumtypen eingeteilt werden. Hierbei fungieren die Einschätzungen der Experten als Zielvariablen und die generierten geografischen Daten als erklärende Variablen. Das Random-Forest-Modell erlernt dann den Zusammenhang zwischen den erklärenden Variablen und der Zielvariable. Ein Random Forest besteht aus einer

endlichen Anzahl von Entscheidungsbäumen. Jeder Entscheidungsbaum stellt an sogenannten Knotenpunkten Kriterien auf, um die Kategorien (hier Raumtypen) bestmöglich vorherzusagen. Jeder Knotenpunkt repräsentiert eine Bedingung, wie Daten mittels eines Schwellenwertes einer erklärenden Variable aufgeteilt werden. Ziel dabei ist es, möglichst homogene Daten in einem Bereich zu sammeln. Da eine große Anzahl von (unkorrelierten) Entscheidungsbäumen aus zufällig gewählten Teilstichproben parallel Kriterien für die Klassifizierung bildet, wird der individuelle Fehler einzelner Entscheidungsbäume minimiert. Die aufgestellten Kriterien können direkt abgelesen und interpretiert werden, weshalb der Random Forest im Vergleich zu anderen überwachten Lernverfahren einen hohen Erklärungsanteil hat und den Einfluss von Variablen auf die Zielvariable gut abbilden kann. Beispielsweise ist die Darstellung der Relevanz einzelner Variablen für die Klassifizierung möglich. Immer wenn eine Variable an einem Knotenpunkt verwendet wird, um die übrigen Daten in zwei Bereiche zu trennen, kann die Entropie berechnet werden (James et al., 2013):

$$E(X) = \sum_{i=1}^c -p_i \log_2 p_i \quad (5.1)$$

Die Entropie $E(X)$ ist ein Maß für die Unordnung bzw. Ordnung in den Daten. p_i ist hierbei die Wahrscheinlichkeit, dass Objekte mit der Klasse i in dem beobachteten Bereich vorkommen. Die Entropie kann einen Wert zwischen 0 und 1 annehmen. Je niedriger der Wert, desto größer ist die Ordnung in diesem Bereich. Um den Einfluss einer Variable auf die Zielvariable zu messen, wird die Veränderung der Entropie von einem Knotenpunkt zum anderen betrachtet. Dies entspricht dem sogenannten Informationsgewinn (IG), welcher sich wie folgt darstellt:

$$IG(X, Y) = E(Y) - E(Y|X) \quad (5.2)$$

$E(Y)$ entspricht dabei der Entropie der Zielvariable (hier Raumtyp) und $E(Y|X)$ setzt sich aus den fünf verwendeten Variablen zur Definition von Urbanität X bedingt der Zielvariable Y zusammen. Je größer der IG, desto wirkungsvoller ist der Einfluss der Variable am Knotenpunkt für die Ordnung des Bereiches. Mit diesem Vorgehen ist es auch möglich, die Rangfolge der Variablen an den jeweiligen Knotenpunkten zu bestimmen, indem an jedem Knotenpunkt der IG einer jeden Variable berechnet wird. Die Variable mit dem größten IG wird dann am Knotenpunkt verwendet. Bei kontinuierlichen Variablen werden die Werte geteilt (z. B. in 0,5er-Schritten) und in kategoriale Variablen transformiert. Für diese wird dann wiederum der IG berechnet, um daraus den optimalen Schwellenwert festzulegen. Da der IG abhängig von der Anzahl der Knotenpunkte ist und mit zunehmender Tiefe des Baumes sinkt, spiegelt er keinen realen Informationsgehalt wider und kann nur innerhalb eines Random-Forest-Modells zwischen

den eingesetzten Variablen verglichen werden. Mit Hilfe des IG können auch irrelevante Variablen aus dem Klassifikationsmodell ausgeschlossen und die einflussreichsten Variablen identifiziert werden.

Die in Tabelle 5.2 beschriebenen Variablen wurden als Variablen mit dem größten Einfluss auf die Zielvariable identifiziert (Abbildung 5.3). Den höchsten Einfluss auf die Klassifikation hat die Variable „Anteil bebauter Fläche“. Die zweitwichtigste Variable ist die „ÖPNV-Qualität“, welche die vorhandenen Alternativen zum Auto berücksichtigt und damit die Abhängigkeit vom Auto darstellt. Die drittwichtigste Variable ist die „Anzahl POIs bezogen auf die bebauten Fläche“. Diese Variable stellt die Entwicklung eines Gebietes dar und gibt an, inwieweit Individuen ihr Gebiet verlassen müssen, um Dinge wie Einkaufen oder Freizeitaktivitäten zu erledigen. An vierter Stelle ist die Bevölkerungsdichte, welche in vielen Studien als Kriterium für die Definition von Urbanität verwendet wird. Dies zeigt, dass der Einfluss von Bevölkerungskriterien zur Definition von Urbanität möglicherweise überschätzt wird. Die fünftwichtigste Variable ist der „Anteil der Industriefläche“ an der bebauten Fläche.

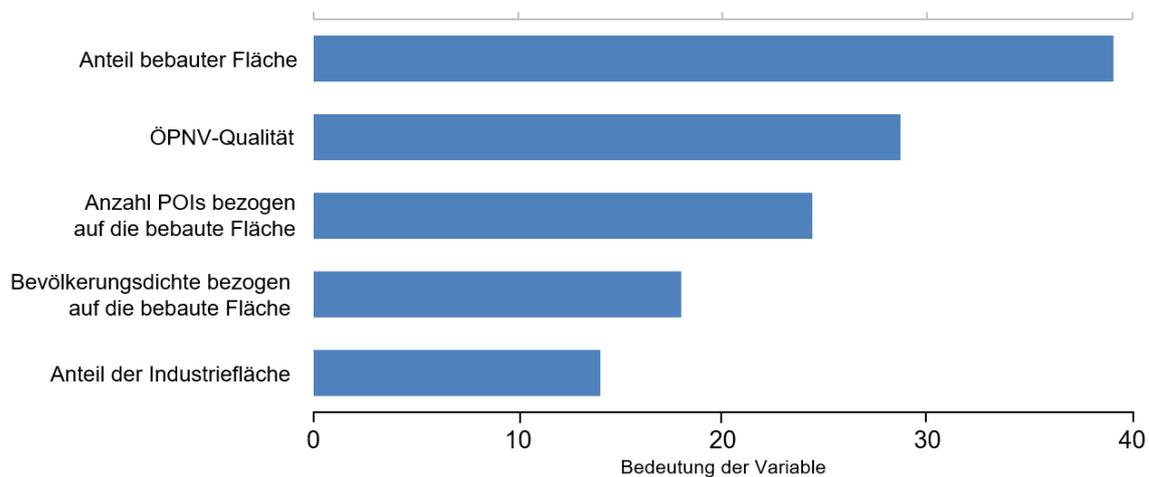


Abbildung 5.3: Variablen mit dem größten Einfluss auf die Klassifizierung

Eine weitere Möglichkeit, den Einfluss der Variablen auf die Klassifikation zu untersuchen, ist die Visualisierung der partiellen Abhängigkeiten (Abbildung 5.4) (Friedman, 2001; Zhao & Hastie, 2019). Dabei wird aufgezeigt, wie der Einfluss der Variablen entlang der Raumtypen variiert. Ist der „Anteil bebauter Fläche“ gering, dominiert die Variable vor allem in ländlichen PLZs. Je größer der Anteil wird, desto eher hat die Variable einen Einfluss auf die Klassifikation von suburbanen/kleinstädtischen und urbanen PLZs. Im Vergleich dazu sind die „ÖPNV-Qualität“ und die „Anzahl POIs bezogen auf die bebauten Fläche“ vor allem urbane Phänomene. Lediglich bei POI auf die bebauten Fläche dominieren am Anfang suburbane/kleinstädtische PLZs und spielen auch bei steigender Anzahl noch eine Rolle. Die „Bevölkerungsdichte bezogen

auf die bebaute Fläche“ hat ab einer Bevölkerungsdichte von 20 000 einen Einfluss auf die Klassifizierung von superurbanen PLZs. Somit charakterisieren sich superurbane PLZs vor allem über extrem hohe Bevölkerungsdichten.

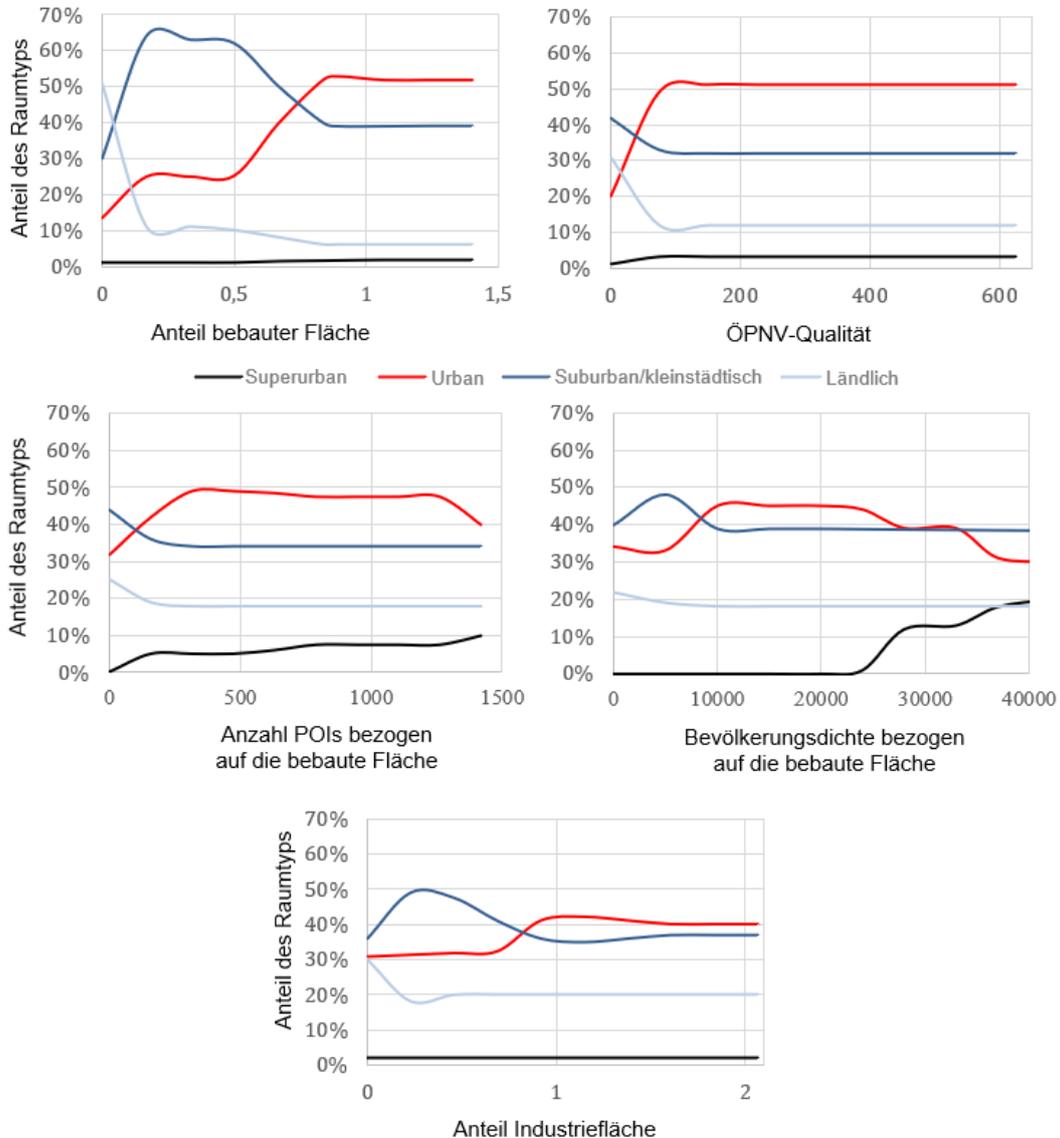


Abbildung 5.4: Partielle Abhängigkeiten der Variablen entlang der Raumtypen

Der durchschnittliche Fehler wird innerhalb der Entscheidungsbäume geschätzt, da jeder Entscheidungsbaum mit Hilfe verschiedener Teilstichproben bestimmt wird (Yan, 2019). Um eine Aussage zur Präzision des Modells zu treffen, wurde der sogenannte „Out-of-Bag-Fehler“ berechnet. Der „Out-of-Bag-Fehler“ ist ähnlich zur n-fachen Kreuzvalidierung und gibt die durchschnittliche Anzahl an falsch klassifizierter PLZs über alle Teilstichproben an. Das finale Modell hat einen Out-of-Bag-Fehler von 16,32 %.

Für die Bildung des Modells wurde das Softwarepaket ranger verwendet, was eine schnelle Random-Forest-Implementierung auf C++ und R darstellt. Die Visualisierung von partiellen Abhängigkeiten und die Berechnung des Einflusses der Variablen auf die Klassifizierung werden bereits in hinterlegten Funktionen zur Verfügung gestellt (Wright & Ziegler, 2015). Für das finale Modell wurde die Anzahl der Entscheidungsbäume auf 850 gesetzt.

5.2.5 Ergebnisse

Im Folgenden wird der UI mit anderen Raumtypologien verglichen, visualisiert und mit verschiedenen Informationen verknüpft. Zum einen kann dadurch nochmals eine Validierung der Einteilung der PLZs in die vier Raumtypen stattfinden. Zum anderen kann mit den zusätzlich verknüpften Informationen ein besseres Verständnis der Unterschiede zwischen den Raumtypen entwickelt werden. Dies ist vor allem für die Beleuchtung der Hintergründe für einer unterschiedlichen Fahrzeugnutzung notwendig.

Bei einem Vergleich des UI mit anderen Raumtypologien am Beispiel von Berlin werden die Vorteile des UI deutlich. Das Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR) beobachtet die siedlungsstrukturellen Zustände in Deutschland. Im Ergebnis zeigt das BBSR die Veränderungen der Lebensverhältnisse zwischen den Regionen für verschiedene Faktoren (Verkehr, Landnutzung etc.) auf. Damit wird unter anderem die gerechte Verteilung finanzieller Ressourcen gewährleistet. Für diesen Forschungszweck ist die Nutzung der Klassifizierung des BBSR allerdings nur begrenzt möglich. Zum einen ist die kleinste räumliche Einheit beim BBSR die Gemeindeebene, wodurch keine innerstädtischen Strukturen beobachtet werden können. Das BBSR deklariert beispielsweise Berlin als einen großen städtischen Raum. Beim UI hingegen setzt sich Berlin aus insgesamt 191 PLZs zusammen, welche superurbane, urbane, suburbane/kleinstädtische sowie ländliche PLZs enthalten (Abbildung 5.5). Des Weiteren unterscheidet das BBSR lediglich zwischen zwei Raumtypen (dem ländlichen und dem städtischen Raum), während der UI Gebiete in vier Raumtypen einteilt. Als Letztes ist die Raumtypisierung des BBSR lediglich für Deutschland verfügbar und deshalb nicht im internationalen Vergleich anwendbar. Zur Validierung werden die Flächen- und Bevölkerungsverteilung zwischen dem BBSR und dem UI verglichen. Unter dem BBSR sind 69 % der Fläche und 32 % der Bevölkerung in Deutschland als ländlich deklariert, während der UI 68 % der Fläche und 28 % der Bevölkerung als ländlich einstuft. Die Ergebnisse sind somit trotz der unterschiedlichen Ansätze durchaus konsistent.

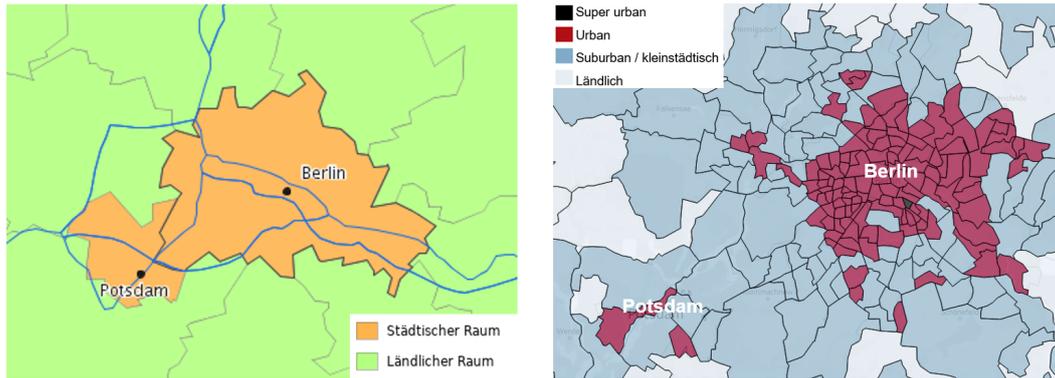


Abbildung 5.5: Vergleich der Raumtypisierung des BBSR mit der Raumtypisierung anhand des UI am Beispiel von Berlin

Anhand des UI werden nun Unterschiede hinsichtlich Bevölkerung, Kaufkraft, Motorisierungsrate und Affinität zum Vielfahren entlang der vier Raumtypen für Deutschland und Frankreich untersucht. Bereits die Visualisierung liefert erste Hinweise auf die Unterschiede hinsichtlich Urbanität zwischen den Ländern (Abbildung 5.6). Die schwarzen Polygone repräsentieren superurbane PLZs. In Deutschland gibt es lediglich eine PLZ (12047 Berlin-Neukölln), welche als superurban deklariert wird. In Frankreich werden sieben Arrondissements als superurban ausgewiesen. Die rot markierten urbanen PLZs sind in Deutschland über das ganze Land verstreut. In Frankreich hingegen sind urbane PLZs vor allem um die Hauptstadt Paris angesiedelt. 105 der 189 urbanen PLZs befinden sich in Paris und Umgebung. Dies führt auch dazu, dass es in Frankreich großflächige ländliche Gebiete gibt. In Deutschland hingegen dominiert ein großer Anteil an suburbaner/kleinstädtischer Gebiete die Landkarte.

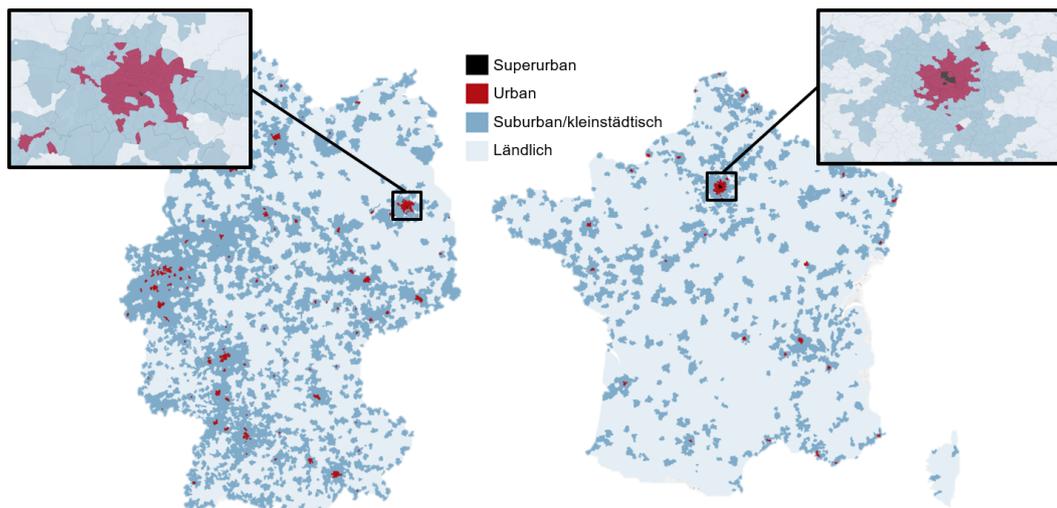


Abbildung 5.6: Vergleich von Deutschland (Berlin) und Frankreich (Paris)

Die visuellen Unterschiede zwischen Deutschland und Frankreich bestätigen sich auch bei Betrachtung der erklärenden Variablen (ÖPNV-Qualität (1), Anteil bebauter Fläche (2), Anteil Industriefläche (3), Bevölkerungsdichte bezogen auf die bebaute Fläche (4), Anzahl POI bezogen auf die bebaute Fläche (5)) (Tabelle 5.3). Die Werte in den Klammern stellen die Standardabweichung dar. Bei Betrachtung der Minimum-/Maximumwerte der Variablen wird der Vorteil durch die Nutzung von nichtlinearen Strukturen sichtbar. Der minimale und der maximale Wert der Bevölkerungsdichte bezogen auf die bebaute Fläche liegt in Deutschland im urbanen Gebieten bei 0 bzw. 33 036 sowie in suburbanen/kleinstädtischen Gebieten bei 160 bzw. 16 073. Somit können auch Gebiete als urban eingestuft werden, deren Bevölkerungsdichte niedrig ist, sofern andere Variablen einen hohen Wert ausweisen. Die Bevölkerungsdichte bezogen auf die bebaute Fläche (4) beträgt in Deutschland durchschnittlich 3043 (2542) und in Frankreich 2496 (2687). Wie zu erwarten, steigt die Bevölkerungsdichte mit zunehmender Urbanität stark an. Während in Frankreich lediglich ein leichter Anstieg zwischen ländlichen und suburbanen/kleinstädtischen Gebieten zu verzeichnen ($\Delta 407$) ist, ist dieser in Deutschland deutlicher erkennbar ($\Delta 796$). Dies liegt vor allem an der Siedlungsstruktur in Deutschland, welche das Prinzip der zentralen Ortstheorie verfolgt, was gewisse Standards für die Bildung von Kleinstädten definiert. Dadurch werden Kleinstädte automatisch attraktiver und damit auch dichter besiedelt. Das spiegelt sich vor allem bei der Betrachtung der Anzahl von POIs bezogen auf die bebaute Fläche (5) wider. Während sich in Frankreich die Anzahl für ländliche und suburbane/kleinstädtische Gebiete kaum unterscheidet ($\Delta 0,2$), ist die Differenz in Deutschland ($\Delta 10,3$) wesentlich wahrnehmbarer. Weniger POIs in einem Gebiet bedeuten auch, dass Individuen sich in anderen Gebieten aufhalten müssen, um die alltäglichen Dinge (z. B. Einkaufen) zu erledigen. In Konsequenz müssen Individuen mobiler sein und größere Tagesdistanzen zurücklegen. Dieses Szenario bestätigen auch Motte-Baumvol et al. (2010) in ihrer Arbeit. Der Anteil der bebauten Fläche (2) an der Gesamtfläche steigt auch mit zunehmender Urbanität. Im Gegensatz zu den anderen Variablen ist hier der Anstieg in beiden Ländern ähnlich. Der Anteil der Industriefläche (3) an der bebauten Fläche ist vor allem in ländlichen Bereichen (DE: 0,08; FR: 0,08) sehr gering. Dort wird die bebaute Fläche primär als Wohnfläche genutzt. In suburbanen/kleinstädtischen und urbanen Gebieten ist das Mischverhältnis zwischen Industrie- und Wohnfläche wesentlich größer. In urbanen Gebieten in Deutschland ist der Anteil sogar fast ausgeglichen (0,44). Die Betrachtung von superurbanen Gebieten kann hier vernachlässigt werden, da nur eine sehr geringe Anzahl von PLZs als superurban deklariert wurden. Große Unterschiede gibt es bei der ÖPNV-Qualität. Durchschnittlich ist die ÖPNV-Qualität in Deutschland (9,8) mehr als doppelt so hoch wie in Frankreich (3,8), dadurch können Deutsche im Vergleich zu Franzosen eher auf Alternativen zum Pkw zurückgreifen. Diesen Zusammenhang zeigen auch Siedentop et al. (2013) in ihrer Untersuchung auf. Ein Grund dafür ist das gut ausgebaute Schienennetz in Deutschland. Auch

in urbanen Gebieten ist die Divergenz zwischen Deutschland (83,6) und Frankreich (50,2) noch sehr groß. Im Rückschluss bedeutet das, dass in Frankreich urban lebende Menschen noch von ihrem Pkw abhängig sind. Das Ergebnis der Untersuchung der Variablen mit dem größten Einfluss auf die Klassifizierung (Abbildung 5.3) hat die Relevanz des ÖPNV bereits aufgezeigt und wird hier noch einmal bestätigt. In ländlichen Gebieten in Frankreich existiert nahezu kein ÖPNV-Angebot, während in Deutschland ein Basisangebot an ÖPNV (z. B. Bus) vorhanden ist.

Tabelle 5.3: Erklärende Variablen nach Raumtyp und Land

Erklärende Variablen zur Definition von Urbanität					
Raumtyp	1	2	3	4	5
Deutschland					
Superurban	49,4 (–)	1,0 (–)	0,00 (–)	28 747 (–)	480,0 (–)
Urban	83,6 (72,5)	0,8 (0,2)	0,44 (1,95)	9268 (4902)	203,9 (753,0)
Suburban/ kleinstädtisch	8,0 (10,5)	0,3 (0,2)	0,28 (0,53)	3037 (1363)	25,1 (61,9)
Ländlich	1,4 (1,2)	0,1 (0,0)	0,08 (0,21)	2241 (1239)	14,7 (20,9)
Durchschnitt	9,8 (29,1)	0,2 (0,2)	0,18	3043 (2542)	32,2 (210,7)
Frankreich					
Superurban	122,8 (41,7)	0,9 (0,1)	0,01 (0,02)	33 229 (5905)	526,6 (213,0)
Urban	50,2 (33,2)	0,8 (0,1)	0,39 (1,18)	11 524 (6777)	110,3 (156,0)
Suburban/ kleinstädtisch	6,5 (8,3)	0,3 (0,2)	0,30 (0,39)	2465 (1356)	11,0 (11,8)
Ländlich	0,4 (0,8)	0,1 (0,1)	0,08 (0,21)	2058 (1558)	11,2 (19,0)
Durchschnitt	3,8 (12,05)	0,2 (0,2)	0,15	2496 (2687)	14,7 (40,5)

Durch die Nutzung der PLZ als räumliche Einheit können nun beliebige Informationen mit dem UI verknüpft und in Abhängigkeit von den Raumtypen ausgewertet werden. Tabelle 5.4 zeigt die Verteilungen der Bevölkerung, der Kaufkraft, der durchschnittlichen Haushaltsgröße, der Motorisierungsrate und der Affinität zu häufigem Fahren entlang der Raumtypen. In Deutschland leben fast 60 % der Bevölkerung in suburbanen/kleinstädtischen Gebieten, in Frankreich hingegen sind es nur 46 %. Diese Differenz verschiebt sich hauptsächlich auf die ländlichen Gebiete. Während in Deutschland 28 % auf dem Land wohnen, sind es in Frankreich 37 %. Dieser hohe Anteil in Frankreich deutete sich bereits bei der visuellen Darstellung (Abbildung 5.6) an und wird durch die Verteilung der Bevölkerung bestätigt. In Frankreich leben 16 %

(1% + 15%) der Bevölkerung in superurbanen/urbanen Gebieten. In Deutschland hingegen sind es nur 13%. Auch die Betrachtung der Kaufkraft zeigt Unterschiede zwischen den Ländern auf. In Frankreich ist die Kaufkraft im Durchschnitt größer (DE: € 21 025; FR: € 22 258) und steigt mit zunehmender Urbanität (ländlich: € 17 856, suburban/kleinstädtisch: € 20 060, urban: € 22 128, superurban: € 28 987) an, während sich in Deutschland die Kaufkraft auf die Raumtypen ähnlicher verteilt (ländlich: € 21 388, suburban/kleinstädtisch: € 22 512, urban: € 22 235, superurban: € 17 967). Dies könnte bedeuten, dass sich die Spreizung zwischen Arm und Reich in Deutschland nicht nach Raumtypen aufteilt, während in Frankreich durchaus Unterschiede entlang der Raumtypen erkennbar sind. Wie zu erwarten, nimmt die Haushaltsgröße mit zunehmender Urbanität (superurban: 1,67, urban, 1,72, suburban/kleinstädtisch: 2,05, ländlich: 2,25) in Deutschland ab. Einpersonenhaushalte sind in urbaneren Gebieten gängiger, während auf dem Land möglicherweise mehrere Generationen in einem Haus wohnen. In Frankreich ist die größte Haushaltsgröße in suburbanen/kleinstädtischen Gebieten (2,37) vorhanden, während die Haushaltsgröße auf dem Land leicht darunter liegt (2,33). Eine weitere Bestätigung für die Wirksamkeit des UI, Unterschiede des Mobilitätsverhaltens entlang der Raumtypen abzubilden, zeigt sich in der Betrachtung der durchschnittlichen Motorisierungsrate pro 1000 Einwohner. Generell ist in Deutschland die Motorisierungsrate mit 604 Fahrzeugen pro 1000 Einwohner größer als in Frankreich (543 Fahrzeuge pro 1000 Einwohner). Darüber hinaus zeigt sich, dass in Deutschland (superurban: 169, urban: 562, suburban/kleinstädtisch: 569, ländlich: 633), analog zur Kaufkraft, die Motorisierungsrate ähnlicher verteilt ist als in Frankreich (superurban: 222, urban: 390, suburban/kleinstädtisch: 530, ländlich: 555). Die Differenz der Motorisierungsrate zwischen urbanen und suburbanen/kleinstädtischen Gebieten ist in Deutschland ($\Delta 7$) wesentlich geringer als in Frankreich ($\Delta 140$). Somit ist das Interesse für den Fahrzeugbesitz in urbanen Gebieten, trotz der Existenz von Alternativen zum Pkw, vorhanden. Der Faktor „Affinität zu häufigem Fahren“ basiert auf einem Rating, welches die zurückgelegten Pkw-Kilometer in einem bestimmten Zeitraum berücksichtigt. Ein Wert von 107,51 bedeutet, dass die Affinität hinsichtlich häufigen Fahrens 7,51-Prozentpunkte über dem Landesdurchschnitt liegt. Dieser Wert ist somit nicht zwischen den Ländern vergleichbar. In Deutschland (superurban: 81,33, urban: 88,59, suburban/kleinstädtisch: 101,65, ländlich: 107,51) nimmt die Affinität zu häufigem Fahren mit zunehmender Urbanität ab. In Frankreich hingegen ist sie in suburbanen/kleinstädtischen Gebieten (108,25) überdurchschnittlich. Dafür könnte das Fehlen einer zentralen Ortsfunktion verantwortlich sein, wodurch die Menschen gezwungen sind, in andere Gebiete zu fahren. Die Befragten aus ländlichen Gebieten repräsentieren den Landesdurchschnitt (99,83).

Tabelle 5.4: Verteilung verschiedener Faktoren nach Raumtyp und Land

Verknüpfte Informationen zur Beschreibung von Urbanität					
Raumtyp	Bevölkerung	Kaufkraft	Haushaltsgröße	Motorisierungsrate	Affinität zu häufigem Fahren
Deutschland					
Superurban	0%	€ 17 967	1,67	169	81,33
Urban	13%	€ 22 235	1,72	562	88,59
Suburban/ kleinstädtisch	59%	€ 22 512	2,05	569	101,65
Ländlich	28%	€ 21 388	2,25	633	107,51
Durchschnitt	–	€ 21 025	1,92	604	100
Frankreich					
Superurban	1%	€ 28 987	1,84	222	82,59
Urban	15%	€ 22 128	2,20	390	88,08
Suburban/ kleinstädtisch	46%	€ 20 060	2,37	530	108,25
Ländlich	37%	€ 17 856	2,33	555	99,83
Durchschnitt	–	€ 22 258	2,19	543	100

Ausweitung des Urbanitätsindex

Aufgrund der Skalierbarkeit konnten analog zum beschriebenen Vorgehen 13 weitere Länder (Österreich (AT), Belgien (BE), Dänemark (DK), Spanien (ES), Finnland (FI), Italien (IT), Japan (JP), Niederlande (NL), Norwegen (NO), Schweden (SE), Südkorea (SK), Vereinigtes Königreich (UK) sowie USA (US)) mit dem UI erschlossen werden. Zusätzlich zu den bereits gelabelten 190 PLZs (Deutschland: 160, Frankreich: 4, Spanien: 8, UK: 14, Italien: 4) werden noch 80 weitere PLZs in den jeweiligen Ländern klassifiziert (Österreich: 11, Belgien: 12, Dänemark: 17, Finnland: 9, Japan: 14, Niederlande: 12, Norwegen: 9, Schweden: 8, Südkorea: 10, US: 2), um mögliche Besonderheiten der Länder zu berücksichtigen. Die automatische Klassifizierung erfolgt über das neu trainierte Random-Forest-Modell. Durch die Verwendung der PLZ als räumliche Einheit unterscheiden sich die durchschnittlichen PLZ-Flächen, vor allem in skandinavischen Ländern (DK, FI, SE), stark von den restlichen. Beispielsweise repräsentieren einige PLZs lediglich ein Straßensegment. Um die kleinteiligen PLZs an das Niveau der restlichen PLZs anzugleichen, werden sie aggregiert, indem die letzte Ziffer der

PLZ abgeschnitten wird. Nach der automatisierten Klassifizierung der abgeschnittenen PLZs werden diese wieder in ihre ursprüngliche Größe transformiert. Nachdem alle PLZs in den 15 betrachteten Ländern nach ihrer Urbanität bewertet worden sind, können im nächsten Schritt die PLZs als Schlüssel verwendet werden, um beispielsweise die Bevölkerungsverteilung über die Länder und Raumtypen zu betrachten (Abbildung 5.7). Dänemark, Finnland und Norwegen sind sich in ihrer Bevölkerungsverteilung relativ ähnlich. Im Vergleich zu den anderen Ländern leben sehr viele Menschen in ländlichen Gebieten (DK: 68 %, FI: 69 %, NO: 76 %). Der Rest verteilt sich auf suburbane/kleinstädtische Gebiete (DK: 30 %, FI: 27 %, NO: 21 %). Nur ein ganz geringer Anteil der Bevölkerung lebt in urbanen Gebieten (DK: 2 %, FI: 4 %, NO: 2 %). Die Niederlande haben über alle Länder den größten Bevölkerungsanteil in suburbanen/kleinstädtischen Gebieten (NL: 60 %).

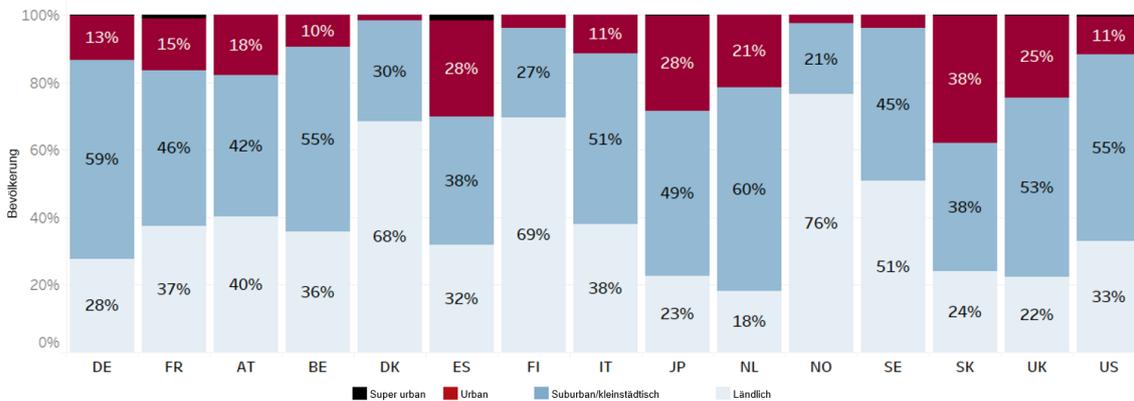


Abbildung 5.7: Bevölkerung nach Raumtypen und Ländern

Der größte Bevölkerungsanteil in urbanen Gebieten ist in Südkorea (SK: 38 %) zu finden. Diese verteilen sich vor allem um Seoul (Abbildung 5.8). Die urbanen Gebieten gehen vom Stadtkern aus und breiten sich immer weiter aus. Visuell sind die Verbindungen der umliegenden Städte (z. B. Gangnam-gu, Gwangmyeon, Bucheon, Uijeongbu und Suwon) untereinander und zu Seoul erkennbar.

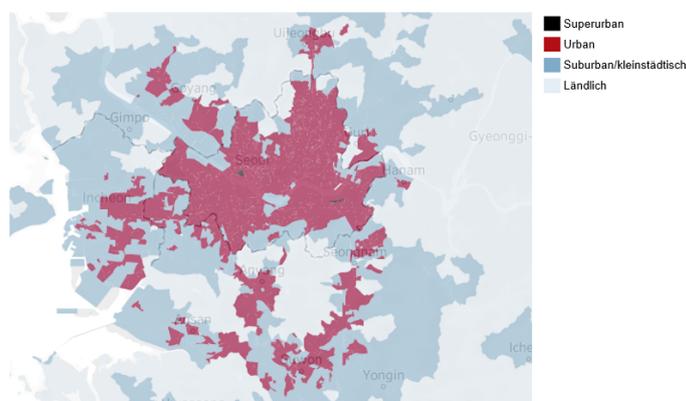


Abbildung 5.8: Darstellung von Seoul

In den USA leben 55 % bzw. 11 % der Bevölkerung in suburbanen/kleinstädtischen bzw. urbanen Gebieten, allerdings erfüllen 95 % der PLZ-Flächen die Kriterien für eine ländliche Klassifizierung (Abbildung 5.9). In Konsequenz sind großflächige Gebiete in den USA nur sehr dünn besiedelt. Dadurch haben die dort lebenden Menschen eine große Abhängigkeit vom Pkw. Gerade in den Bundesstaaten Montana, Wyoming und Nord-/Süddakota gibt es nahezu keine suburbanen/kleinstädtischen oder urbane Strukturen. New York hingegen ist vor allem in Manhattan extrem superurban. 32 PLZs in Manhattan, Queens und Brooklyn sind als superurban klassifiziert. In keiner anderen Stadt gibt es einen so hohen Grad an superurbanen Räumen.

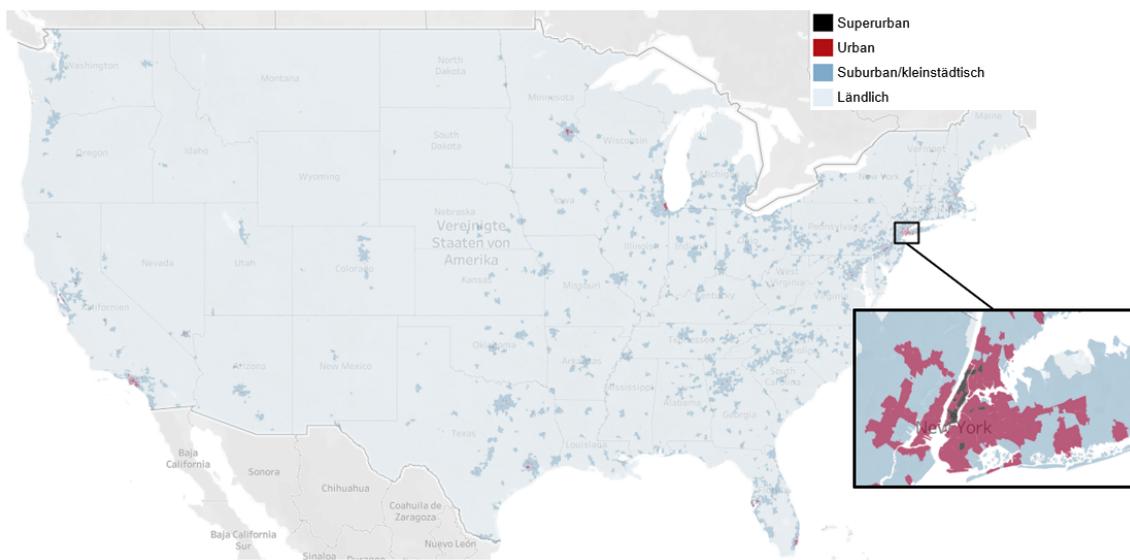


Abbildung 5.9: Darstellung USA und New York

Spanien ist das Land mit dem größten Bevölkerungsanteil in superurbanen Gebieten (ES: 2 %), welche sich primär in den Zentren der Großstädte (z. B. Barcelona: 8, Madrid: 5, Bilbao: 3) (Abbildung 5.10) finden lassen.

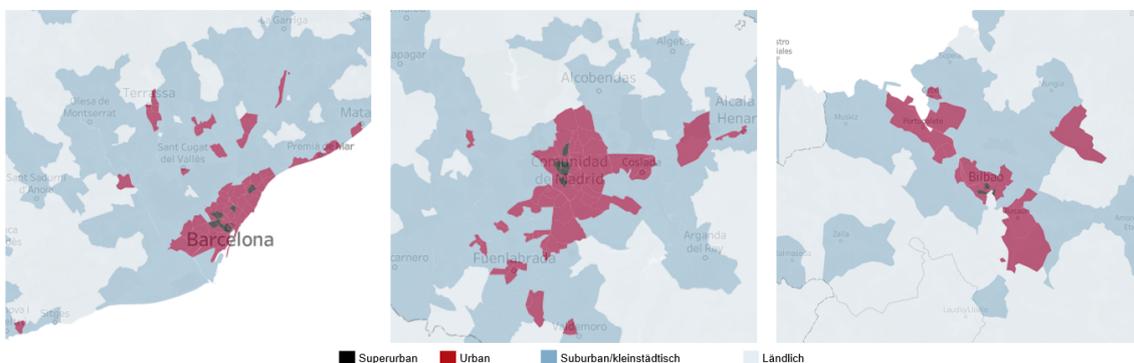


Abbildung 5.10: Darstellung Barcelona, Madrid und Bilbao

5.2.6 Diskussion

In diesem Unterkapitel wurde ein Ansatz entwickelt, um Urbanität im Kontext der Fahrzeugnutzung oder des Mobilitätsverhaltens zu untersuchen. Besondere Aufmerksamkeit wurde hierbei der (1) Berücksichtigung von Einflussfaktoren auf das Mobilitätsverhalten (Anzahl POIs sowie ÖPNV-Qualität in einer PLZ), der (2) internationalen Vergleichbarkeit (einheitliche Datenquelle und Definition von Urbanität), der (3) Reproduzierbarkeit und der (4) praktischen Anwendbarkeit im industriellen Kontext (frei verfügbare Daten und PLZ als räumliche Einheit) geschenkt. Durch die Nutzung eines Random-Forest-Modells und die Experteneinschätzung ist zudem die Skalierbarkeit durch die automatische Einteilung der PLZs in die Raumtypen gewährleistet.

Darüber hinaus konnten durch die besonderen Eigenschaften des Random-Forest-Modells die Relevanz einzelner Variablen und deren partielle Abhängigkeiten hinsichtlich der Klassifizierung untersucht werden. Beide Techniken haben zum Verständnis der Einteilung und zur Untersuchung von Urbanität beigetragen. Die Auswertung des UI unterstellt in Deutschland in suburbanen/kleinstädtischen Gebieten einen höheren Bevölkerungsanteil als in Frankreich. Die gute ÖPNV-Qualität und die hohe Anzahl von POIs in Deutschland wirken sich auch auf das Mobilitätsverhalten aus. Individuen haben dadurch die Möglichkeit zu multimodalem Mobilitätsverhalten, da sie flexibel zwischen öffentlichen und privaten Verkehrsmitteln wechseln können. Dies ist gerade in Zeiten des Mobilitätswandels (z. B. Elektromobilität) bedeutsam. Auf dem Land ist die Abhängigkeit vom Pkw mangels Alternativen größer als in der Stadt. Da in Frankreich ein großer Anteil der Menschen in ländlichen Gebieten lebt, könnte gerade dort eine Mobilitätstransformation schwieriger umsetzbar sein als beispielsweise in Deutschland. Auch die Bevölkerungsverteilung zeigt Unterschiede auf, was zukünftig näher untersucht werden sollte. Durch die Nutzung der PLZ als räumliche Einheit können beliebige Informationen mit dem UI verknüpft und entlang der Raumtypen analysiert werden. Die Wahl dieser räumlichen Einheit bringt allerdings auch Limitationen mit sich:

- PLZs stellen eine Kombination aus Zahlen und Buchstaben dar, um Postsendungen zu verorten. Diese Logik ist vor allem im europäischen Raum stark verbreitet. In China hingegen werden Postsendungen über Straßen und Hausnummern verortet. Dies hat zur Folge, dass in diesen Ländern eine Ausrollung des UI schwieriger ist.
- Die erforderlichen Daten für die Erstellung des UI wurden hauptsächlich über OSM und CLC generiert. Diese Datenquellen können sich aber in Qualität und Aktualität von Land zu Land unterscheiden. Das macht einen internationalen Vergleich schwierig, vor allem wenn ein Land sich in großen strukturellen Veränderungen befindet (z. B. China).

Die gegenwärtige Ausgestaltung des UI bietet auch Raum für Verbesserungen und für die Erarbeitung neuer Erkenntnisse:

- Die verwendeten Datenquellen liefern lediglich Informationen über die Anzahl der Bus- bzw. Schienenhaltestellen, es gibt allerdings keine Informationen darüber, mit welcher Taktung diese angefahren werden. Die Frequenz hat aber durchaus Auswirkungen auf die Qualität der Haltestelle. Darüber hinaus ist auch nicht bekannt, wie viele Zielhaltestellen angesteuert werden können und wie weit diese von der Starthaltestelle entfernt sind. Anhand dieser Informationen könnte die ÖPNV-Qualität noch besser abgebildet werden.
- Um Unterschiede zwischen den Raumtypen zu identifizieren, ist es möglich, Pendelbewegungen von Fahrzeugen mit dem UI zu verknüpfen. Gerade für Mobilitätslösungen (z. B. ODM) als Alternative zum Pkw-Besitz und deren optimale Ausgestaltung sind diese Analysen hilfreich.
- Die Einstellung von Individuen könnte entlang der Raumtypen zur Identifikation von Abweichungen hinsichtlich der Aufgeschlossenheit für neue Mobilitätslösungen oder Antriebstechnologien untersucht werden.

Zusammenfassend kann gesagt werden, dass der UI die Grundlage für viele raumspezifische Untersuchungen ist. Durch die einheitliche Definition und die skalierbare Methodik sind auch internationale Vergleiche auf PLZ-Ebene möglich. Die Berücksichtigung von Urbanität im Kontext des Mobilitätsverhaltens für die Ausgestaltung neuer Mobilitätslösungen und deren länder- und städtespezifischen Besonderheiten ist ebenfalls denkbar. Durch den UI könnten politische Maßnahmen und deren Effekte auch auf andere Länder übertragen werden, da nun ein internationaler Vergleich von gleichen Raumtypen möglich ist. Dies hätte ein großes Einsparungspotenzial, da nicht jede politische Maßnahme in jedem Land pilotiert werden müsste. Des Weiteren wird durch die Untersuchung der Unterschiede zwischen den Raumtypen mit den verknüpften Informationen ein größeres Verständnis von Urbanität generiert. Dies ist notwendig, wenn im nächsten Unterkapitel die Fahrzeugnutzung entlang der Raumtypen beleuchtet wird und die Hintergründe für diese differenzierte Nutzung erörtert werden.

5.3 Analyse verschiedener Einflussfaktoren auf die Fahrzeugnutzung

In Kapitel 4 wurden raumbezogene Fahrzeugvariablen (Tabelle 4.5) bestimmt, um die Fahrzeugnutzung in Relation zum Kontext darzustellen. Dadurch können beispielsweise Unterschiede in der Fahrzeugnutzung in Abhängigkeit von Urbanität (Verknüpfung mit dem UI) untersucht werden, indem der approximierte Wohnort, bestehend aus Längen- und Breitengrad, auf die jeweilige PLZ aggregiert wird. Durch die internationale Vergleichbarkeit ist es zukünftig

auch möglich, identische Analysen in unterschiedlichen Ländern durchzuführen und Vergleiche zu ziehen. Des Weiteren kann mit Hilfe der Fahrzeuginformationen (z. B. Fahrzeugtyp) die Fahrzeugnutzung in Abhängigkeit vom Konzept und Segment analysiert werden.

Tabelle 5.5 stellt die Quelle-Ziel-Matrizen entlang der Raumtypen der approximierten Wohnorte dar.¹⁰ Um die Übersichtlichkeit der Tabellen zu erhöhen, wurden die Anteile der Fahrten, welche im selben Raumtyp beginnen und enden, fett markiert. 40,7 % der Fahrten von Fahrzeugen deren Wohnort im urbanem Raumtyp ist, starten im urbanen Raumtyp und enden im urbanen Raumtyp. Generell lässt sich sagen, dass der Raumtyp des Wohnortes auch gleichzeitig der Raumtyp ist, in dem am meisten Fahrten beginnen und enden (urban: urban–urban: 40,7 %; suburban/kleinstädtisch: suburban/kleinstädtisch–suburban/kleinstädtisch: 69,8 %; ländlich: ländlich–ländlich: 47,6 %). Vor allem in suburbanen/kleinstädtischen Gebieten herrscht ein großer Bezug zum Raumtyp. Lediglich 7,3 % bzw. 4,9 % der Fahrten, welche in suburbanen/kleinstädtischen Gebieten starten, enden in einem anderen Raumtyp. Hierbei sollte aber auch berücksichtigt werden, dass 59 % der Bevölkerung (Abbildung 5.7) auch in suburbanen/kleinstädtischen Gebieten leben. Deshalb ist schon allein aufgrund der Grundverteilung der Bevölkerung im Vergleich zu den anderen Raumtypen die Wahrscheinlichkeit größer, dass die Fahrt in suburbanen/kleinstädtischen Gebieten stattfindet. Bei Fahrzeugen, deren Wohnort als urban bzw. ländlich eingestuft ist, enden 17,9 % bzw. 14,5 % der betrachteten Fahrten in suburbanen/kleinstädtischen Gebieten und starten im jeweiligen Raumtyp des Wohnortes. Über alle Wohnorte ist die Symmetrie in den Quell-Ziel-Matrizen zu erkennen. Unterscheidet sich mit Fahrtbeginn und -ende der Raumtyp (Werte abseits der Diagonalen), findet sich dies auch in umgekehrter Kombination in der Matrix wieder. (urban: 17,4 %–17,9 %, 1,8 %–1,9 %, 1,6 %–1,6 %; suburban/kleinstädtisch: 7,3 %–7,4 %, 0,3 %–0,3 %, 4,8 %–4,9 %; ländlich: 1,5 %–1,7 %, 3,1 %–2,9 %, 14,5 %–14,7 %).

¹⁰Da in Deutschland nur eine PLZ die Kriterien für einen superurbanen Raum erfüllt, wird auf die Ausweisung des superurbanen Raumtyps für diese Analyse verzichtet.

Tabelle 5.5: Quelle-Ziel-Raumtypmatrizen: Anteil der Fahrten nach Raumtyp des approximierten Wohnortes

Wohnort	Raumtyp (Quelle)	Raumtyp (Ziel)		
		Urban	Suburban/ kleinstädtisch	Ländlich
Urban	Urban	40,7 %	17,9 %	1,9 %
	Suburban/kleinstädtisch	17,4 %	15,0 %	1,6 %
	Ländlich	1,8 %	1,6 %	2,2 %
Suburban/ kleinstädtisch	Urban	2,4 %	7,4 %	0,3 %
	Suburban/kleinstädtisch	7,3 %	69,8 %	4,9 %
	Ländlich	0,3 %	4,8 %	2,7 %
Ländlich	Urban	1,8 %	1,7 %	2,9 %
	Suburban/kleinstädtisch	1,5 %	12,3 %	14,7 %
	Ländlich	3,1 %	14,5 %	47,6 %

Im nächsten Schritt kann nun die Verteilung des Wohnortes entlang der Raumtypen in Abhängigkeit von den FNPs (universales Langstreckenfahrzeug (1), Mittelstreckenpendelfahrzeug (2), universales Kurzstreckenfahrzeug (3), distanzunabhängiges Freizeitfahrzeug (4), ortsunabhängiges Langstreckenfahrzeug (5)) aufgezeigt werden (Tabelle 5.6). Des Weiteren können die erklärenden Variablen zur Definition von Urbanität (Tabelle 5.3) sowie die verknüpften Informationen (Tabelle 5.4) genutzt werden, um ein besseres Verständnis der Gründe der unterschiedlichen Fahrzeugnutzung entlang der Raumtypen zu entwickeln.

Im gewichteten Durchschnitt ist die Verteilung der betrachteten Fahrzeuge nach Raumtypen (urban: 11 %, suburban/kleinstädtisch: 62 %, ländlich: 27 %) ähnlich zur generellen Bevölkerungsverteilung (urban: 13 %, suburban/kleinstädtisch: 59 %, ländlich: 29 %) in Deutschland. Universale Langstreckenfahrzeuge (1) sind gleichartig zum Durchschnitt verteilt mit einer marginalen Verschiebung zu ländlichen Gebieten. Möglicherweise repräsentieren universale Langstreckenfahrzeuge die durchschnittliche Nutzung im Datensatz, welcher über alle Raumtypen vorhanden ist.

Anteilmäßig ist in ländlichen Gebieten das Mittelstreckenpendelfahrzeug (2) mit 34,0 % am stärksten vertreten. In ländlichen Gebieten gibt es nur wenige POIs (14,7), weshalb Anreize existieren, sich in andere, urbanere Gebiete zu bewegen (Motte-Baumvol et al., 2010). Möglicherweise befindet sich der Arbeitsort in suburbanen/kleinstädtischen bzw. urbanen Gebieten, wodurch Fahrzeuge zum Pendeln eingesetzt werden. Darüber hinaus existiert in ländlichen Gebieten eine sehr hohe Motorisierungsrate (633 Fahrzeuge auf 1000 Einwohner).

Möglicherweise handelt es sich hier oftmals um Fahrzeuge aus einem Haushalt mit mehreren Fahrzeugen, wodurch sich die Pkw-Mobilität auf die vorhandenen Fahrzeuge verteilt und diese jeweils für bestimmte Zwecke eingesetzt werden.

70,4 % der universalen Kurzstreckenfahrzeuge (3) haben ihren Wohnort in suburbanen/kleinstädtischen PLZs. Wie bereits bei den Quelle-Ziel-Matrizen aus Tabelle 5.5 identifiziert, ist es in diesen Gebieten möglich, viele Aktivitäten (Arbeit, Restaurantbesuche, Einkaufen) innerhalb des Raumtyps zu erledigen. Aus diesem Grund ist die Nutzung des Fahrzeugs für größere Distanzen nicht zwangsweise notwendig, was den hohen Kurzstreckenanteil erklären würde. Darüber hinaus ist in suburbanen/kleinstädtischen Gebieten die ÖPNV-Qualität (8,0) wesentlich geringer als in urbanen Gebieten (83,6), weshalb das Fahrzeug noch sehr universell für jegliche Fahrtzwecke eingesetzt werden muss.

23 % der distanzunabhängigen Freizeitfahrzeuge (4) haben ihren Wohnort in urbanen Gebieten. Dieser Wert ist doppelt so hoch wie der gewichtete Durchschnitt. Vor allem in urbanen Gebieten ist eine Substitution des Pkw durch die Bahn für bestimmte Fahrtzwecke (z. B. zur Arbeit) aufgrund der sehr guten ÖPNV-Qualität (83,6) denkbar (von Behren et al., 2018). Dadurch wird das Fahrzeug nur für bestimmte Aktivitäten eingesetzt und könnte beispielsweise nur noch für Fahrten in solche Gebiete benötigt werden, in denen keine ausreichende ÖPNV-Qualität vorhanden ist. Der Einsatz des Fahrzeugs ist dadurch sehr variabel gestaltet, was sich möglicherweise in der Distanzunabhängigkeit widerspiegelt.

Bei 5 % der ortsunabhängigen Langstreckenfahrzeuge (5) entspricht der Wohnort den Kriterien eines urbanen Gebietes. Möglicherweise handelt es sich hierbei um Fahrzeuge mit mehreren Wohnorten. Der Hauptwohnsitz ist in ländlichen (32,5 %) bzw. suburbanen/kleinstädtischen (62,5 %) Gebieten und der Zweitwohnsitz in urbanen Gebieten. Diese Konstellation ist vor allem für Berufspendler in größeren Firmen mit hohem Verdienst und dem damit verbundenen finanziellen Rahmen zur Finanzierung von zwei Wohnorten (Haus auf dem Land, Wohnung in der Stadt) durchaus verbreitet.

Tabelle 5.6: Raumtyp des approximierten Wohnortes nach Fahrzeugnutzungsprofilen

Wohnort	Fahrzeugnutzungsprofil					Durchschnitt
	1	2	3	4	5	
Urban	11,6 %	8,7 %	8,7 %	23,0 %	5,0 %	11,4 %
Suburban/kleinstädtisch	56,7 %	57,3 %	70,4 %	57,5 %	62,5 %	61,6 %
Ländlich	31,7 %	34,0 %	20,9 %	19,5 %	32,5 %	27,0 %

Tabelle 5.7 zeigt die Quelle-Ziel-Matrizen entlang der identifizierten FNP (universales Langstreckenfahrzeug (1), Mittelstreckenpendelfahrzeug (2), universales Kurzstreckenfahrzeug (3), distanzunabhängiges Freizeitfahrzeug (4), ortsunabhängiges Langstreckenfahrzeug (5)). In

dieser Darstellung sind die jeweiligen Anteile der Fahrten und die mittleren Fahrdistanzen (in Klammern) zwischen Quelle-Ziel-Raumtypen enthalten. Über alle FNPs sind die Fahrdistanzen umso größer, je größer die Differenz der Urbanität zwischen den Quelle-Ziel-Matrizen ist. Wenn bei einem universalen Langstreckenfahrzeug (1) eine Fahrt im urbanen Raum beginnt und im ländlichen Raum endet, werden hier im Mittel 48 km zurückgelegt. Im suburbanen/kleinstädtischen Raum hingegen sind es nur 20 km. Endet eine Fahrt im gleichen (urbanen) Raumtyp, beträgt die Fahrdistanz im Mittel 5 km.

Zudem verhalten sich die FNPs 1–4, wie auch bei den wohnortsabhängigen Quelle-Ziel-Matrizen (Tabelle 5.5), relativ symmetrisch. Die Anteile der Fahrten und die mittleren Fahrdistanzen zwischen den Quelle-Ziel-Raumtypen spiegeln sich um die Diagonale. Das ortsunabhängige Langstreckenfahrzeug (5) wird im Hinblick auf die Fahrdistanzen leicht asymmetrisch genutzt. Beispielsweise entspricht die mittlere Fahrdistanz bei Fahrten, welche in urbanen Räumen beginnen und in ländlichen Räumen enden, 96 km, während die mittleren Fahrdistanzen bei Fahrten von ländlichen Räumen in urbane Räume lediglich 66 km betragen. Diese mittleren Fahrdistanzen sollten aber mit einer gewissen Sorgfalt betrachtet werden, da lediglich 1,1 % der Fahrten von ländlichen in urbane bzw. von urbanen in ländliche Gebiete absolviert wurden. Diese Asymmetrie ist möglicherweise auf die Ortsunabhängigkeit zurückzuführen, da dieser Nutzungstyp nicht einen Wohnort hat, bei dem er immer übernachtet.

Die Länge der mittleren Fahrdistanzen zwischen den Raumtypen unterscheidet sich aber zwischen den FNPs. Über 60 km legen distanzunabhängige Freizeitfahrzeuge (4) und ortsunabhängige Langstreckenfahrzeuge (5) bei Fahrten zwischen urbanen und ländlichen Gebieten zurück. Gerade diese Unabhängigkeiten hinsichtlich der Distanzen und Örtlichkeiten führen dazu, dass der Aktionsradius dementsprechend groß ist. Beide FNPs haben auch die größten Standzeiten außerhalb eines Radius von 100 km um den Wohnort (Tabelle 4.6: 4: 13,1 %; 5: 21,8 %). Universale Langstreckenfahrzeuge (1) bzw. Kurzstreckenfahrzeuge (3) legen ca. 45 km bzw. 30 km zwischen urbanen und ländlichen Gebieten zurück. Das Mittelstreckenpendelfahrzeug (2) legt im Mittel nur 26 km bei Fahrten von urbanen in ländliche Räume zurück. Sogar das universale Kurzstreckenfahrzeug (3) legt mit 30 km größere Fahrdistanzen zurück. Allerdings finden nur 0,6 % der Fahrten von urbanen in ländliche Gebiete statt. Knapp 75 % der Fahrten universaler Kurzstreckenfahrzeuge (4) (urban–urban: 5,8 %; suburban/kleinstädtisch–suburban/kleinstädtisch: 54,4 %; ländlich–ländlich: 14,6 %) beginnen und enden im selben Raumtyp. Dies kann auch über den Wohnort erklärt werden. 70,4 % (Tabelle 5.6) der universalen Kurzstreckenfahrzeuge (4) wohnen in suburbanen/kleinstädtischen Gebieten. Darum ist es möglicherweise nicht notwendig, in andere Raumtypen zu pendeln, um die täglichen Bedürfnisse zu befriedigen. Im Gegensatz dazu beginnen und enden lediglich 61 % (urban–urban: 3,6 %; suburban/kleinstädtisch–suburban/kleinstädtisch: 42,7 %; ländlich–ländlich: 14,7 %) der Fahrten beim Mittelstreckenpendelfahrzeug im gleichen

Raumtyp. Wie bereits in Tabelle 5.6 gezeigt, „wohnen“ im Vergleich zu den anderen FNPs die meisten Mittelstreckenpendelfahrzeuge in ländlichen Gebieten (34,0%). Durch die geringe Anzahl von POIs werden viele Mittelstreckenpendelfahrzeuge dazu genutzt, in urbanere Gebiete zu pendeln. Dieser Zusammenhang spiegelt sich auch in den Quelle-Ziel-Raumtypmatrizen wieder. 9,7% der Fahrten beginnen in ländlichen Gebieten und enden in suburbanen/kleinstädtischen Gebieten. Bei keinem anderen FNP ist dieser Wert so hoch (1: 8,4%; 3: 5,3%; 4: 5,7%; 5: 7,4%).

Tabelle 5.7: Quelle-Ziel-Raumtypmatrizen: Anteil der Fahrten (mittlere Fahrdistanz) nach Fahrzeugnutzungsprofilen

FNP	Raumtyp (Quelle)	Raumtyp (Ziel)		
		Urban	Suburban/ kleinstädtisch	Ländlich
1	Urban	7,6 % (5 km)	6,4 % (20 km)	1,7% (48 km)
	Suburban/kleinstädtisch	6,2 % (19 km)	43,5 % (7 km)	8,5 % (30 km)
	Ländlich	1,8 % (47 km)	8,4 % (32 km)	16,1 % (6 km)
2	Urban	3,6 % (4 km)	8,2 % (18 km)	1,6 % (26 km)
	Suburban/kleinstädtisch	8,1 % (19 km)	42,7 % (12 km)	9,8 % (17 km)
	Ländlich	1,6 % (26 km)	9,7 % (18 km)	14,7 % (8 km)
3	Urban	5,8 % (4 km)	6,8 % (10 km)	0,6 % (30 km)
	Suburban/kleinstädtisch	6,7 % (10 km)	54,4 % (5 km)	5,3 % (13 km)
	Ländlich	0,6 % (34 km)	5,3 % (14 km)	14,6 % (4 km)
4	Urban	9,6 % (4 km)	6,8 % (10 km)	1,1 % (64 km)
	Suburban/kleinstädtisch	6,5 % (9 km)	46,8 % (3 km)	5,9 % (26 km)
	Ländlich	1,2 % (62 km)	5,7 % (24 km)	16,4 % (3 km)
5	Urban	7,9 % (5 km)	8,9 % (19 km)	1,1 % (96 km)
	Suburban/kleinstädtisch	8,5 % (19 km)	48,5 % (6 km)	7,7 % (23 km)
	Ländlich	1,1 % (66 km)	7,4 % (24 km)	8,8 % (3 km)

In Unterkapitel 2.3 wurde ein Ansatz zur Segmentierung von Fahrzeugen auf der Basis objektiver Kriterien (z.B. Fahrzeuglänge) vorgestellt. Für jedes betrachtete Fahrzeug ist eine Zuweisung zu einem Segment und Konzept möglich. Tabelle 5.8 zeigt die Verteilung der einzelnen Segmente bzw. Konzepte nach dem Raumtyp des approximierten Wohnortes. Vorab sollte darauf hingewiesen werden, dass die Stichproben durch die Unterteilung nach Konzept bzw. Segment sehr gering sind. Zudem ist, wie bereits gesagt, die Auswahl der Fahrzeuge willkürlich erfolgt, weshalb auch die Verteilung der Segmente und Konzepte

nicht zwingend repräsentativ für die Gesamtflotte sein muss. Darüber hinaus sind hier nur Fahrzeuge eines Premiumautomobilherstellers enthalten, was dazu führt, dass kleinere und kostengünstigere Fahrzeuge (UKL, KKL) unterrepräsentiert sind. Deshalb sollten die Ergebnisse und ihre Interpretation mit einer gewissen Vorsicht behandelt werden. Vielmehr fungieren diese Analysen als Beispiele, um die vielfältigen Verknüpfungsmöglichkeiten zu demonstrieren. Durch die Skalierbarkeit kann diese Logik zukünftig auf viele Fahrzeuge über mehrere Länder ausgeweitet werden. Bei den Segmenten ist auffällig, dass vor allem kleine Fahrzeuge (UKL1 (z.B. VW Polo): 22,9%; UKL2 (z.B. VW Golf): 17,0%; KKL (z.B. VW Passat): 11,0%; KKL2 (z.B. BMW 3er): 12,5%) ihren approximierten Wohnort in urbanen Gebieten haben. Dies ist durchaus nachvollziehbar, da gerade in urbanen Gebieten der Platz begrenzt ist (z. B. durch enge Straßen oder kleine Parklücken). Aus diesem Grund werden möglicherweise kleine Fahrzeuge aufgrund der Fahrzeuggröße in urbanen Gebieten bevorzugt eingesetzt. Im Kontrast dazu ist der Anteil größerer Fahrzeuge in urbanen Gebieten fast nicht vorhanden (MKL (z.B. BMW 5er): 5,3%; GKL (z.B. BMW X7): 6,9%). Vielmehr befindet sich der approximierte Wohnort bei größeren Fahrzeugen (GKL: 65,5%; MKL: 66,5%) in suburbanen/kleinstädtischen Gebieten. In diesen ist zudem auch die Kaufkraft (Tabelle 5.4) mit € 22 512 am höchsten. Die Unterteilung der Fahrzeuge nach Konzepten zeigt auf, dass Cabrios (19,0%), Coupé (18,8%) sowie Hatch (22,4%) eher im urbanen Gebiet zum Einsatz kommen. Speziell Cabrios sind eher weniger für den Transport sperriger Gegenstände geeignet. Diese Konzepte werden eher für seltene Ereignisse, wie beispielsweise Wochenendausflüge, vorgehalten. Auch hier könnte es sein, dass in urbanen Gebieten Fahrzeuge für bestimmte Zwecke (z. B. Freizeit) gekauft und eingesetzt werden. Die restlichen Aktivitäten werden möglicherweise mit anderen Verkehrsmitteln erledigt. Die Nutzung von Estates findet weniger in urbanen Gebieten (9,5%) statt, sondern eher im ländlichen Raum (32,7%). Sie eignen sich durch die Verbindung zwischen Fahrgast- und Kofferraum besonders gut zum Transport von sperrigen Gegenständen. Möglicherweise ist gerade auf dem Land der Transport großer Gegenstände relevanter als in urbanen Gebieten. SAVs sind durch das Geländewagenkonzept eher im ländlichen Bereich zu erwarten, allerdings ist die Verteilung mit der durchschnittlichen vergleichbar. SAT- und Sedan-Konzepte sind überdurchschnittlich in suburbanen/kleinstädtischen Gebieten vertreten. Möglicherweise sind diese Konzepte eine gute Zwischenlösung, um die Probleme großer Autos im urbanen Umfeld und gleichzeitig die Abhängigkeit von Fahrzeugen im ländlichen Bereich zu berücksichtigen.

Tabelle 5.8: Verteilung der Fahrzeuge nach Raumtyp des approximierten Wohnortes und Segment/Konzept

Segment	Wohnort			Stichprobe
	Urban	Suburban/ kleinstädtisch	Ländlich	
GKL	6,9 %	65,5 %	27,6 %	29
MKL	5,3 %	66,5 %	28,2 %	170
KKL2	12,5 %	62,3 %	25,0 %	8
KKL	11,0 %	62,3 %	26,7 %	236
UKL2	17,0 %	54,4 %	28,6 %	147
UKL1	22,9 %	60,0 %	17,1 %	35
Durchschnitt	11,4 %	61,6 %	27,0 %	625
Konzept				
Cabrio	19,0 %	57,1 %	23,8 %	21
Coupé	18,8 %	56,3 %	25,0 %	16
Estate	9,5 %	57,7 %	32,7 %	168
Hatch	22,4 %	51,3 %	26,3 %	76
MPV	–	66,7 %	33,3 %	12
SAT	6,0 %	66,0 %	28,0 %	50
SAV	11,7 %	64,4 %	23,9 %	180
Sedan	6,9 %	69,6 %	23,5 %	102
Durchschnitt	11,4 %	61,6 %	27,0 %	625

Durch die Verknüpfung mit dem UI kann die Fahrzeugnutzung differenziert nach Raumtypen betrachtet werden. Darüber hinaus können durch die Darstellung von Quelle-Ziel-Matrizen auch Pendelverflechtungen zwischen den Raumtypen in Abhängigkeit von verschiedenen Faktoren (z.B. approximierter Wohnort, FNP) untersucht werden. Die Verknüpfung des Fahrzeugtyps zur Bestimmung des Segments bzw. Konzepts eines Fahrzeugs liefert eine weitere Perspektive für eine differenzierte Betrachtung der Fahrzeugnutzung. Diese Herangehensweise schafft eine flexible Grundlage zur Untersuchung von Fahrzeugnutzung auf der Basis von Sensordaten und bietet Raum für zukünftige Untersuchungen:

- Der zugrundeliegende Datensatz ist mit 625 Fahrzeugen und einem durchschnittlichen Beobachtungszeitraum von ca. einem Monat noch relativ klein. Um robustere Ergebnisse zu generieren, sollte, sobald die Daten verfügbar und aufbereitet sind, die Stichprobe

dementsprechend vergrößert werden. Dadurch sind Regelmäßigkeiten in der Fahrzeugnutzung, die Approximation des Wohnortes sowie die wichtigen Standorte stabiler identifizierbar.

- Um den UI maximal zu nutzen, sollten, sobald die Fahrzeugnutzungsdaten für weitere Länder vorhanden sind, auch internationale Untersuchungen vorgenommen werden. Damit wäre es erstmalig möglich, Unterschiede in der Fahrzeugnutzung zwischen Ländern und Raumtypen zu identifizieren.
- Durch Sensordaten ist der Beobachtungszeitraum beliebig skalierbar. Zukünftige Untersuchungen sollten das Fahrzeugalter als weitere Perspektive in den Analysen berücksichtigen. Dadurch könnten die verschiedenen Lebensphasen eines Fahrzeugs besser beobachtet werden.

6 Fazit und Ausblick

Das letzte Kapitel greift die eingangs definierten Forschungsfragen wieder auf und beantwortet sie mit Hilfe der in dieser Arbeit geschaffenen Erkenntnisse. Darauf aufbauend wird ein übergreifendes Fazit gezogen und zukünftige Anknüpfungspunkte für inhaltliche und technologische Weiterentwicklungen zur Identifikation von FNPs diskutiert. Eine Aufzählung zusätzlicher Analyse- und Verknüpfungsmöglichkeiten zur Untersuchung der Fahrzeugnutzung rundet das Kapitel ab.

In welchen Fachabteilungen eines Automobilherstellers und auf welche Art und Weise gibt es einen Bedarf zur Untersuchung von Fahrzeugnutzung?

Um die Anforderungen zur Untersuchung von Fahrzeugnutzung zu identifizieren, wurden 22 Interviews geführt. Dabei konnte vor allem der Bedarf aus Industrieperspektive aufgezeigt werden. Speziell die Untersuchung von Fahrzeugnutzung (im Mobilitätskontext) in der Beobachtungsphase im Feld und die Verknüpfung der gewonnenen Erkenntnisse mit der Vorentwicklungsphase stellen ein großes Bedarfsfeld dar. Sowohl in der Vorentwicklungsphase (Unterkapitel 2.3) als auch in der Literatur (Abschnitt 4.2.1) sind Segmentierungen ein weitverbreitetes Instrument, welches zum Verständnis eines Marktes einen Mehrwert liefern kann. Deshalb wurde der Fokus dieser Arbeit auf die Entwicklung eines nutzungsorientierten Segmentierungsansatzes und die Identifikation von FNPs gesetzt. Durch die Verknüpfung mit bestehenden Segmentierungen (Kunden- und Fahrzeugsegmentierung) entsteht eine umfassende Betrachtungsweise hinsichtlich des Automobilmarkts und dessen Facetten. Darüber hinaus liefern FNPs für die Wissenschaft eine ganzheitliche Abbildung der Fahrzeugnutzung, mit deren Hilfe bisher wenig erforschte Bereiche aus der Mobilitätsforschung (z. B. Fernverkehr in Abhängigkeit vom Wohnort, einheitliche Definition von Fahrtzwecken über verschiedene Länder sowie Lebensphasen eines Fahrzeugs) zukünftig näher untersucht werden können.

Welche Dimensionen und Einflussfaktoren der Fahrzeugnutzung sind für den zugrundeliegenden Anwendungsbereich relevant?

Die generelle Fahrzeugnutzung unterteilt sich in die „Fahrzeugnutzung (im Mobilitätskontext)“, die „Funktionsnutzung“ und die „Fahrdynamik“. Die Funktionsnutzung kann dabei in Infotainment, Fahrerassistenzsystem (FAS) sowie Funktionen im Interieur und Exterieur aufgeteilt werden. Bei der Fahrdynamik ist eine Unterscheidung zwischen Quer- und Längsbeschleunigung möglich. Die Dimensionen der Fahrzeugnutzung (im Mobilitätskontext)

und die Einflussfaktoren auf die Fahrzeugnutzung, welche auch im Fokus dieser Arbeit standen, konnten über eine Literaturrecherche (Unterkapitel 2.1) bestimmt werden. Die qualitative Literaturrecherche zeigte auf, dass bei Betrachtung der Fahrzeugnutzung (im Mobilitätskontext) die Dimensionen Zeit, Distanz und Fahrtzweck im Mittelpunkt vieler Analysen stehen. Die Einflussfaktoren auf die Fahrzeugnutzung wurden in Kontext, Fahrzeug und Kunde aufgeteilt (Abbildung 2.1). Speziell die Betrachtung von Urbanität, welche ein Teilgebiet des Kontextes darstellt, wurde, durch die Entwicklung eines international vergleichbaren Urbanitätsindex (UI) auf PLZ-Ebene, in dieser Arbeit intensiv diskutiert. Der Index bildet die Grundlage für länderübergreifende Analysen zur Fahrzeugnutzung in Abhängigkeit von Urbanität.

Wie und mit welchen Daten lassen sich Nutzungsmuster identifizieren?

Zur Analyse von Fahrzeugnutzung (im Mobilitätskontext) und der Identifikation von Nutzungsmustern gibt es zwei Datenquellen: Umfrage- und Sensordaten. Beide Quellen bringen Vor- und Nachteile mit sich, welche detailliert an verschiedenen Stellen dieser Arbeit diskutiert wurden. Speziell Sensordaten bilden die objektive Nutzung eines Fahrzeugs im internationalen Kontext ab. Umfragedaten hingegen können auch Faktoren wie Einstellungen und soziodemografische Informationen in Relation zur Fahrzeugnutzung setzen. Allerdings sind Umfragedaten hinsichtlich des Beobachtungszeitraums, des nationalen Fokus und der geringen Stichprobe im Nachteil. In dieser Arbeit wurden die Vorteile beider Datenquellen verknüpft, um neue Erkenntnisse zu gewinnen (von Premium- bzw. Elektrofahrzeugen mit gewöhnlichen Fahrzeugen, Validierung der Methodik zur Approximation von Fahrtzwecken). Vor allem aufgrund der länderübergreifenden Betrachtung von Fahrzeugnutzung und der einheitlichen Erhebungsmethode wurde sich in dieser Arbeit ab Kapitel 4 auf die Verwendung von Sensordaten fokussiert.

Welche Eigenschaften haben FNPs?

Eine allumfassende Definition von FNPs mit speziellen Eigenschaften gibt es nicht und wäre aufgrund der unterschiedlichen Einsatzbereiche auch nicht zielführend. Um die Flexibilität in Abhängigkeit vom Anwendungsfall zu garantieren, muss der entsprechende Ansatz (A-Priori-/A-Posteriori-Segmentierung, Einbindung in bestehende FNPs) zur Identifikation von FNPs gewählt werden. Das Vorgehen zur Approximation des Fahrtzwecks (Unterkapitel 4.1) und die Variablen zur Beschreibung der drei Dimensionen (Unterkapitel 4.2) liefern dabei die Grundlage zur Analyse der Fahrzeugnutzung (im Mobilitätskontext) unter Anwendung von Sensordaten. Diese Herangehensweise bietet zudem eine Reihe zukünftiger Forschungsmöglichkeiten, welche im Ausblick noch weiter diskutiert werden.

Wie unterscheiden sich FNPs nach differenzierenden Faktoren?

Es gibt viele Einflussfaktoren, welche sich auf die Intensität und Variabilität der Fahrzeugnutzung auswirken. Unter anderem spielt der Kontext, in dem das Fahrzeug hauptsächlich genutzt wird, eine wichtige Rolle. Hierbei sind alternative Verkehrsmittel zum Pkw, aber auch die POIs in der unmittelbaren Umgebung zu berücksichtigen. Durch die Verknüpfung des approximierten Wohnortes mit dem UI konnte die Fahrzeugnutzung differenziert nach Raumtyp betrachtet werden. Hinsichtlich des Kontextes könnten in zukünftigen Diskussionen soziodemografische Informationen sowie meteorologische Aspekte näher untersucht werden. Darüber hinaus könnte das Fahrzeug an sich (Konzept, Segment) als differenzierender Faktor noch detaillierter analysiert werden. Diese Untersuchung würde Erkenntnisse liefern, wie stark die zukünftige Fahrzeugnutzung bereits beim Kauf eines Fahrzeugs berücksichtigt wird. Hierbei stellt sich aber die Frage, ob und wie unterschiedliche Datenquellen verknüpft werden können.

Die dargestellte Metrik liefert eine Momentaufnahme, deren Bausteine auf den Anwendungsbereich zugeschnitten sind. Je nachdem welches Ziel bei der Untersuchung verfolgt wird, kann sich daraus auch die inhaltliche und methodische Herangehensweise verändern. Aus diesem Grund bildet die Definition der Fragestellung das Fundament für das beschriebene Vorgehen. Mit der präsentierten Metrik ist es möglich, die Variablen zu verändern oder neue zu kreieren. Dies ist, speziell in der gegenwärtig sehr dynamischen Umwelt, auch notwendig. Natürlich wurden viele Aspekte in dieser Arbeit vernachlässigt oder nur am Rande erwähnt. Die Fahrzeugnutzung ist ein sehr komplexer Vorgang, welcher durch unterschiedliche Faktoren definiert und beeinflusst werden kann. Auch die Stärke des Einflusses kann sich zwischen den Individuen unterscheiden. In Konsequenz ist die Skizzierung eines vollständigen Bildes zur Beschreibung der Fahrzeugnutzung und deren Hintergründe nahezu unmöglich. Besonders die Verknüpfung von Kundeninformationen und Fahrzeugnutzung stellt eine große Herausforderung dar. Vor allem der in dieser Arbeit öfter diskutierte Perspektivenwechsel (Perspektive von Individuen und des Fahrzeugs) macht die Verknüpfung schwierig. Durch die stetige Weiterentwicklung von Alternativen zum Pkw und die individuelle Einstellung zu aktuellen Themen (z. B. Klimaschutz) ist diese Kombination für die Wissenschaft und Industrie durchaus relevant. Mögliche Datenquellen zur Beschreibung der Fahrzeugnutzung und Identifikation von FNPs sind vielseitig. Oftmals ist eine Eins-zu-eins-Verknüpfung nicht möglich. In Konsequenz können nicht alle benötigten Informationen für eine Untersuchung aus einer Datenquelle herangezogen werden. Um dieses Problem abzuschwächen, wurden zwei Ansätze genutzt: Zum einen wurden bestehende FNPs als Gerüst verwendet, um zwei Datenquellen vergleichbar zu machen, zum anderen fand eine Anreicherung von Sensordaten mit weiteren Informationen (Fahrtzweck) statt. Allerdings sollten bei einer

Fusionierung und Anreicherung von Daten immer die Einschränkungen, wie unterschiedliche Datenverteilungen und Annahmen, transparent dargestellt werden. Bei Einflussfaktoren sollte immer darauf geachtet werden, dass diese auch für unterschiedliche Datenquellen messbar und anwendbar sind. Beispielsweise kann der UI auch für andere Datenquellen verwendet werden, vorausgesetzt, die Informationen der PLZs sind im Datensatz enthalten. Dies bietet zudem die Möglichkeit, unterschiedliche Datenquellen auf der Basis der geografischen Einheit zu verknüpfen. Bei der Interpretation des Einflusses auf die Fahrzeugnutzung sollten die nicht berücksichtigten Faktoren in die Argumentation mit einfließen. Wichtig ist in diesem Zusammenhang immer, dass die Metriken dynamischer Natur sind und Veränderungen in der Abbildung von Fahrzeugnutzung erkannt und angepasst werden können. Deshalb werden im Folgenden mögliche inhaltliche und methodische Weiterentwicklungen für die verschiedenen beobachteten Bereiche geliefert.

Ansatz zur Identifikation von Fahrzeugnutzungsprofilen

In dieser Arbeit wurden die Anzahl der FNPs und deren Eigenschaften entweder über bereits existierende FNPs oder über die Verteilung der Daten selbst bestimmt. Dadurch konnten zum einen vergleichende Analysen und zum anderen eine Basissegmentierung mit dem A-posteriori-Ansatz generiert werden. Die Basissegmentierung unterteilt die grundlegende Fahrzeugnutzung in Profile. In der Wissenschaft wie auch in der Praxis werden aber für bestimmte Zwecke spezielle Nutzungsprofile gesucht, welche auch extreme Profile darstellen können. Diese spiegeln sich möglicherweise nicht in einer Basissegmentierung wider. In weiterführenden Arbeiten sollten die generierten Variablen über Zeit, Distanz und Fahrtzweck a priori verwendet werden, um anwendungsorientierte Segmentierungskriterien festzulegen. Darauf aufbauend wäre es beispielsweise möglich, den Anteil der Fahrzeuge mit den Segmentierungskriterien (z. B. Kriterien für die Eignung von ODM) an der Gesamtheit zu messen oder Korrelationen zu den Einflussfaktoren herzustellen. Durch die Verknüpfung von A-posteriori- und A-priori-Segmentierungen könnten sowohl allgemeine als auch anwendungsorientierte FNPs identifiziert werden.

Weiterentwicklung der Bausteine genereller Fahrzeugnutzung und deren Einflussfaktoren

Der Fokus dieser Arbeit lag auf der Betrachtung von Fahrzeugnutzung im Mobilitätskontext. Dabei wurden Elemente aus der Funktionsnutzung (Infotainment, Funktionen im Interieur und Exterieur sowie FAS) und der Fahrdynamik (Quer- und Längsbeschleunigung) nicht berücksichtigt. Für eine ganzheitliche Darstellung der generellen Fahrzeugnutzung sollten auch diese Bausteine betrachtet werden. Beispielsweise könnten hier, analog zur Vorgehensweise für

die Identifikation von FNPs, ebenfalls Funktionsnutzungs- und Fahrdynamikprofile erstellt und in Relation zu den anderen Profilen und den beschriebenen Einflussfaktoren gesetzt werden.

Nutzung technologischer Fortschritte in der Datengenerierung und -verarbeitung zur detaillierteren Beschreibung der Fahrzeugnutzung

Die Technologien zur Übertragung von Sensordaten werden stetig weiterentwickelt. Vor allem die Übertragungsraten und die Datengröße sind durch cloudbasierte Systeme beliebig veränderbar. Durch diese nahezu unbegrenzten Möglichkeiten ergeben sich weitere Ansätze, die Fahrzeugnutzung noch detaillierter über Sensordaten zu beschreiben. Darüber hinaus wird es zukünftig möglich sein, durch die zunehmende Digitalisierung des Fahrzeugs neue Signale flexibel und schnell abzurufen. Beispielsweise könnte durch ein Signal zur elektrischen Reichweite oder zum Tankfüllstand der Lade- bzw. Tankvorgang eindeutig identifiziert werden. Mit den hier verwendeten Daten ist das nur über die Verknüpfung von POI-Informationen möglich. Des Weiteren wäre über die Sitzplatzbelegung, welche in den Daten zur Bestimmung von Zeit-Distanz-Fahrtzweck-FNPs nicht verfügbar ist, der Fahrtzweck „Holen/Bringen“ ebenfalls eindeutig bestimmbar. Eine Unterscheidung von privat und geschäftlich genutzten Fahrzeugen ist aufgrund technischer Einschränkungen gegenwärtig nicht möglich. Eine zukünftige Unterscheidung dieser Nutzerklassen würde viele neue Erkenntnisse zur Nutzung in Abhängigkeit vom Besitzverhältnis beisteuern. Sensordaten betrachten die objektive Nutzung eines Fahrzeugs. Subjektive Faktoren können dabei nicht eingesehen werden. Durch Befragungen im Fahrzeug (beispielsweise über den Bildschirm) könnten die Einstellung und Werteorientierung abgefragt werden. Mit diesem Ansatz wäre es möglich, sowohl objektive Fahrzeugnutzungsdaten als auch subjektive Informationen des Fahrzeugnutzers zu generieren und eindeutig miteinander zu verknüpfen. Darüber hinaus würde die entwickelte Logik lokal auf einen begrenzten Datensatz angewendet. Um die effiziente und nachhaltige Nutzung der Logik für Millionen von Fahrzeugen und Milliarden von Fahrten zu garantieren, sollte eine Anpassung des zugrundeliegenden Skripts zur Anwendung in der Produktivumgebung im Big-Data-Bereich stattfinden. Hierbei müssten die Skriptsprache (von Python zu PySpark) für verschiedene Verarbeitungsschritte (z. B. Ersetzen von Schleifen mit Fensterfunktionen) sowie die verwendete Entwicklungsumgebung (von lokaler zu cloudbasierter (z. B. Amazon Web Services) Umgebung) dementsprechend angepasst werden.

Weitere Analyse- und Verknüpfungsansätze

In Unterkapitel 5.3 wurden verschiedene Analyse- und Verknüpfungsmöglichkeiten aufgezeigt. Darüber hinaus gibt es noch eine Reihe weiterer Möglichkeiten, die Fahrzeugnutzung in Abhängigkeit von verschiedenen Faktoren zu betrachten:

Analyseansätze

- Von Behren et al. (2018) bestimmen die objektive und subjektive Autoabhängigkeit von Individuen. Wenn die Abhängigkeiten der Individuen in einem Fahrzeug gebündelt werden und gleichzeitig die Fahrzeugnutzung betrachtet wird, könnten die objektive und subjektive Autoabhängigkeit entlang der FNPs untersucht werden.
- Bei der Analyse von Fahrzeugnutzung wurde der Wohnort anhand der meisten Übernachtungen an einem Ort im Beobachtungszeitraum identifiziert. Hierbei wurde nur ein Standort als Wohnort angenommen. In der Realität können Fahrzeugnutzer aber auch mehrere Wohnorte haben (Wochenendpendler), was wiederum Einfluss auf die Fahrzeugnutzung und die Stellung des Fahrzeugs hat. Deshalb ist eine Weiterentwicklung der Methodik durch die Approximation von mehreren Wohnorten denkbar.
- Der Fernverkehr wird bei diesem Ansatz aus einem anderen Blickwinkel betrachtet, insofern als die Entfernung vom Wohnort das Kriterium für den Fernverkehr ist. Weitere Untersuchungen sollten den Zusammenhang zwischen Fahrzeugnutzung und der Entfernung zum Wohnort im Detail betrachten. Speziell bei der Elektromobilität und der Annahme des Ladens am Wohn- und Arbeitsort könnte der Aktionsradius des Fahrzeugs ein wichtiges Akzeptanzkriterium sein.
- Die Verknüpfung der beschriebenen Segmentierungsansätze (Fahrzeug-, Kunden- und Fahrzeugnutzungssegmentierung) könnte die Untersuchung der Fahrzeugnutzung und deren Einflussfaktoren komplettieren. Um die Kundensegmentierung (soziale Milieus) und die Fahrzeugnutzung miteinander zu verknüpfen, müssen weitere Annahmen getroffen werden. Vor allem sollten hierbei die unterschiedlichen Betrachtungsweisen (Individuen- und Fahrzeugebene) bei der Verknüpfung berücksichtigt werden.
- Die Identifikation von Halterwechseln bei Fahrzeugen ist in der Praxis nicht trivial. Zukünftige Analysen könnten sich mit der Frage auseinandersetzen, inwieweit eine Identifikation von Halterwechseln auf der Basis der Fahrzeugnutzung möglich ist.

Verknüpfungsansätze

- Über Sensordaten können Fahrzeuge im Längsschnitt betrachtet werden, was eine Untersuchung der Fahrzeugnutzung entlang des Fahrzeugalters ermöglicht.
- Durch die Approximation des Wohnortes eines Fahrzeugs können soziodemografische Informationen auf PLZ-Ebene verknüpft werden. Viele Datenlieferanten (z. B. Gesellschaft für Konsumforschung (GfK), Nexiga etc.) haben sich auf Geomarketing spezialisiert und bieten Informationen auf verschiedenen geografischen Ebenen an.

- Die Anzahl der Fahrzeugnutzer, die Haushaltsgröße sowie die damit verbundene Stellung des Fahrzeugs im Haushalt wurden in dieser Arbeit nicht berücksichtigt. Dadurch könnten die Bündelung der Fahrzeugnutzung und deren Fahrtzwecke auf die einzelnen Fahrzeugnutzer aufgeteilt und der Einsatzzweck eines Fahrzeugs noch besser verstanden werden.
- Meteorologische Kennzahlen wurden bei der Bestimmung der FNP's nicht untersucht. Übertragungstechnologien könnten zukünftig auch Kontextinformationen (z. B. Außentemperatur) in ihrem Skript speichern, was eine Untersuchung der Zusammenhänge zwischen der Intensität der Fahrzeugnutzung und bestimmten Wettergegebenheiten ermöglichen würde. Diese Analysen würden weitere Erkenntnisse hinsichtlich der Einstellung zum und Abhängigkeit vom Auto liefern.

Mit dieser Arbeit wurde erstmalig eine grundlegende Metrik zur ganzheitlichen Beschreibung der Fahrzeugnutzung und deren Einflussfaktoren entwickelt. Darüber hinaus wurden die zugrundeliegenden Sensordaten nicht nur in ihrer vorhandenen Form untersucht, sondern auch um den Fahrtzweck angereichert. Dies schaffte ein erweitertes Bild zur Beschreibung der Fahrzeugnutzung (im Mobilitätskontext) über alle drei Dimensionen (Zeit, Fahrtzweck, Distanz). Um die Einflussfaktoren auf die Fahrzeugnutzung zu quantifizieren, wurde ein international vergleichbarer Urbanitätsindex entwickelt, welcher zum ersten Mal die Untersuchung der Zusammenhänge zwischen Fahrzeugnutzung und Urbanität auf internationaler Ebene ermöglicht. All die angewandten Logiken und Methoden leisten auf der einen Seite einen wissenschaftlichen Beitrag und garantieren auf der anderen Seite die praxisnahe Anwendung im industriellen Kontext. Darüber hinaus schafft diese Arbeit die Grundlage zur Beantwortung vieler Fragestellungen, welche eine kontinuierliche Betrachtung der Fahrzeugnutzung über einen längeren Beobachtungszeitraum voraussetzen.

Literaturverzeichnis

- Abe, S. (2005). *Support vector machines for pattern classification* (Bd. 2). Springer.
- Adamou, A., Clerides, S. & Zachariadis, T. (2014). Welfare implications of car feebates: A simulation analysis. *The Economic Journal*, 124 (578), F420–F443.
- Administration, F. H. (2017). 2017 National Household Travel Survey, U.S. Department of Transportation. Zugriff auf <https://nhts.ornl.gov>
- Aldous, D. J. (1985). Exchangeability and related topics. In *École d'été de probabilités de saint-flour xiii—1983* (S. 1–198). Springer.
- Alexander, L., Jiang, S., Murga, M. & González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 58, 240–250.
- Alonso-González, M. J., Hoogendoorn-Lanser, S., van Oort, N., Cats, O. & Hoogendoorn, S. (2020). Drivers and barriers in adopting mobility as a service (maas) – a latent class cluster analysis of attitudes. *Transportation Research Part A: Policy and Practice*, 132, 378–401.
- Arellano, B. & Roca, J. (2018). The urbanization impact in China: A prospective model (1992–2025). In *Remote Sensing and Modeling of Ecosystems for Sustainability XV* (Bd. 10767, S. 107670A).
- Arsanjani, J. J., Zipf, A., Mooney, P. & Helbich, M. (2015). OpenStreetMap in GIScience. *Lecture notes in geoinformation and cartography*.
- Auer, M., Bogenberger, K., Rehborn, J., Koller, M. & Palmer, J. (2017). Mobilitätskennwerte für den motorisierten Individualverkehr aus Flottendaten. *Straßenverkehrstechnik*, 61 (1), 87–93.
- Aultman-Hall, L., Sears, J., Dowds, J. & Hines, P. (2012). Travel demand and charging capacity for electric vehicles in rural states: Vermont case study. *Transportation Research Record*, 2287 (1), 27–36.
- Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfuser, G. & Haupt, T. (2002). Observing the rhythms of daily life: A six-week travel diary. *Transportation*, 29 (2), 95–124.
- Bäumer, M., Hautzinger, H., Kuhnimhof, T. & Pfeiffer, M. (2018). The German Vehicle Mileage Survey 2014: Striking the balance between methodological innovation and continuity. *Transportation Research Procedia*, 32, 329–338.

- Bernard, C. R. (1966). The scree test for the number of factors. *Multivariate Behavioural Research*.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Blei, D. M., Jordan, M. I. et al. (2006). Variational inference for dirichlet process mixtures. *Bayesian analysis*, 1 (1), 121–143.
- BMVI. (2020). *Verkehr in Zahlen 2019/2020*. DVV Media Group GmbH Hamburg.
- BMW. (2020a). BMW Group: Offizielle Zulassungszahlen nach Länder. *Internes Dokument*.
- BMW. (2020b). BMW Group Segmentierung: Handbuch. *Internes Dokument*.
- Bohte, W. & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies*, 17 (3), 285–297.
- Bonges III, H. A. & Lusk, A. C. (2016). Addressing electric vehicle sales and range anxiety through parking layout, policy and regulation. *Transportation Research Part A: Policy and Practice*, 83, 63–73.
- Brand, C., Anable, J. & Tran, M. (2013). Accelerating the transformation to a low carbon passenger transport system: The role of car purchase taxes, feebates, road taxes and scrappage incentives in the uk. *Transportation Research Part A: Policy and Practice*, 49, 132–148.
- Brezzi, M., Dijkstra, L. & Ruiz, V. (2011). *OECD extended regional typology: The economic performance of remote rural regions*. OECD.
- Bricka, S. G., Sen, S., Paleti, R. & Bhat, C. R. (2012). An analysis of the factors influencing differences in survey-reported and GPS-recorded trips. *Transportation Research Part C: Emerging Technologies*, 21 (1), 67–88.
- Buehler, R. (2011). Determinants of transport mode choice: A comparison of Germany and the USA. *Journal of Transport Geography*, 19 (4), 644–657.
- Buehler, R., Pucher, J., Gerike, R. & Götschi, T. (2017). Reducing car dependence in the heart of Europe: Lessons from Germany, Austria, and Switzerland. *Transport Reviews*, 37 (1), 4–28.
- Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr., J. & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, 26, 301–313.
- Cervero, R. & Kockelman, K. (1997). Travel demand and the 3ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2 (3), 199–219.
- Chen, C., Ma, J., Susilo, Y., Liu, Y. & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies*, 68, 285–299.

- Cheng, L., De Vos, J., Shi, K., Yang, M., Chen, X. & Witlox, F. (2019). Do residential location effects on travel behavior differ between the elderly and younger adults? *Transportation Research Part D: Transport and Environment*, 73, 367–380.
- Chlond, B., Weiss, C., Heilig, M. & Vortisch, P. (2014). Hybrid modeling approach of car uses in Germany on basis of empirical data with different granularities. *Transportation Research Record*, 2412 (1), 67–74.
- Delhomme, P. & Gheorghiu, A. (2016). Comparing French carpoolers and non-carpoolers: Which factors contribute the most to carpooling? *Transportation Research Part D: Transport and Environment*, 42, 1–15.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39 (1), 1–22.
- Deng, Z. & Ji, M. (2010). Deriving rules for trip purpose identification from GPS travel survey data and land use data: A machine learning approach. In *Traffic and Transportation Studies 2010* (S. 768–777).
- Dieleman, F. M., Dijst, M. & Burghouwt, G. (2002). Urban form and travel behaviour: Micro-level household attributes and residential context. *Urban studies*, 39 (3), 507–527.
- Dietterich, T. G. & Bakiri, G. (1994). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2, 263–286.
- Dijkstra, L. & Poelman, H. (2014). A harmonised definition of cities and rural areas: The new degree of urbanisation. *European Commission Directorate-General for Regional and Urban Policy*, WP01/2014, 24.
- Ding, S., Zhu, Z. & Zhang, X. (2017). An overview on semi-supervised support vector machine. *Neural Computing and Applications*, 28 (5), 969–978.
- Dunteman, G. H. (1989). *Principal components analysis* (Nr. 69). Sage.
- Ecke, L., Chlond, B., Magdolen, M., Eisenmann, C., Hilgert, T. & Vortisch, P. (2019). *Deutsches Mobilitätspanel (MOP) – Wissenschaftliche Begleitung und Auswertungen, Bericht 2017/2018: Alltagsmobilität und Fahrleistung*.
- Ehteshamrad, S., Saffarzadeh, M. & Mamdoohi, A. R. (2017). Role of children in parents' car use behavior. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, 27 (1), 179–188.
- Eisenmann, C. (2019). *Mikroskopische Abbildung von Pkw-Nutzungsprofilen im Längsschnitt* (Bd. 74). KIT Scientific Publishing.
- Eisenmann, C. & Buehler, R. (2018). Are cars used differently in Germany than in California? Findings from annual car-use profiles. *Journal of Transport Geography*, 69, 171–180.
- Elango, V. V., Guensler, R. & Ogle, J. (2007). Day-to-day travel variability in the commute Atlanta, Georgia, study. *Transportation Research Record*, 2014 (1), 39–49.

- Esch, F.-R. & Manger, I. (2019). The brand comes first: Marktsegmentierung und Zielgruppendifinition auf Basis der Marke. In *Handbuch Markenführung* (S. 235–254). Springer.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X. et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Bd. 96, S. 226–231).
- Euroensors. (2018). SIGMA Milieus 2018. *Internes Dokument*.
- Feng, J., Dijst, M., Prillwitz, J. & Wissink, B. (2013). Travel time and distance in international perspective: A comparison between Nanjing (China) and the Randstad (The Netherlands). *Urban Studies*, 50 (14), 2993–3010.
- Figenbaum, E. & Nordbakke, S. (2019). *Battery electric vehicle user experiences in norway's maturing market* (Bericht).
- Fix, E. (1951). *Discriminatory analysis: Nonparametric discrimination, consistency properties*. USAF school of Aviation Medicine.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232.
- Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.
- Gianotti, A. G. S., Getson, J. M., Hutyrá, L. R. & Kittredge, D. B. (2016). Defining urban, suburban, and rural: A method to link perceptual definitions with geospatial measures of urbanization in central and eastern massachusetts. *Urban ecosystems*, 19 (2), 823–833.
- Giuliano, G. & Narayan, D. (2003). Another look at travel patterns and urban form: The US and Great Britain. *Urban studies*, 40 (11), 2295–2312.
- Gong, L., Kanamori, R. & Yamamoto, T. (2018). Data selection in machine learning for identifying trip purposes and travel modes from longitudinal GPS data collection lasting for seasons. *Travel Behaviour and Society*, 11, 131–140.
- Gong, L., Liu, X., Wu, L. & Liu, Y. (2016). Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43 (2), 103–114.
- Gong, L., Morikawa, T., Yamamoto, T. & Sato, H. (2014). Deriving personal trip data from GPS data: A literature review on the existing methodologies. *Procedia-Social and Behavioral Sciences*, 138 (Supplement C), 557–565.
- Gong, L., Sato, H., Yamamoto, T., Miwa, T. & Morikawa, T. (2015). Identification of activity stop locations in GPS trajectories by density-based clustering method combined with support vector machines. *Journal of Modern Transportation*, 23 (3), 202–213.
- Gong, L., Yamamoto, T. & Morikawa, T. (2016). Comparison of activity type identification from mobile phone GPS data using various machine learning methods. *Asian Transport Studies*, 4 (1), 114–128.

- Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep learning*. MIT press.
- Gorges, C., Öztürk, K. & Liebich, R. (2019). Impact detection using a machine learning approach and experimental road roughness classification. *Mechanical Systems and Signal Processing*, 117, 738–756.
- Görür, D. & Rasmussen, C. E. (2010). Dirichlet process gaussian mixture models: Choice of the base distribution. *Journal of Computer Science and Technology*, 25 (4), 653–664.
- Guensler, R., Ko, J., Kim, D., Khoeini, S., Sheikh, A. & Xu, Y. (2019). Factors affecting Atlanta commuters' high occupancy toll lane and carpool choices. *International Journal of Sustainable Transportation*, 1–12.
- Hardman, S., Jenn, A., Tal, G., Axsen, J., Beard, G., Daina, N., ... Kinnear, N. (2018). A review of consumer preferences of and interactions with electric vehicle charging infrastructure. *Transportation Research Part D: Transport and Environment*, 62, 508–523.
- Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Haustein, S. & Nielsen, T. A. S. (2016). European mobility cultures: A survey-based cluster analysis across 28 European countries. *Journal of Transport Geography*, 54, 173–180.
- Haustein, S. & Siren, A. (2015). Older people's mobility: Segments, factors, trends. *Transport Reviews*, 35 (4), 466–487.
- Heller, K. (2008). *Efficient bayesian methods for clustering* (Unveröffentlichte Dissertation). University of London.
- Hidayah, L. & Wulandari, C. (2018). DBSCAN algorithm and decision tree to automate trip purpose detection. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 3 (4), 305–310.
- Hildebrand, E. D. (2003). Dimensions in elderly travel behaviour: A simplified activity-based model using lifestyle clusters. *Transportation*, 30 (3), 285–306.
- IEA. (2019). *World energy outlook 2019*.
- Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J. & Varshavsky, A. (2011). Identifying important places in people's lives from cellular network data. In *International Conference on Pervasive Computing* (S. 133–151).
- Jakobsson, N., Gnann, T., Plötz, P., Sprei, F. & Karlsson, S. (2016). Are multi-car households better suited for battery electric vehicles? – Driving patterns and economics in Sweden and Germany. *Transportation Research Part C: Emerging Technologies*, 65, 1–15.
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An introduction to statistical learning* (Bd. 112). Springer.
- Katenkamp, O. & Schröder, A. (2003). *Praxishandbuch: Empirische Sozialforschung: Arbeit, Gesundheit, Frauenfragen, Altersforschung, Sozialpolitik, Armut*. LIT Verlag.

- Kelly, P., Krenn, P., Titze, S., Stopher, P. & Foster, C. (2013). Quantifying the difference between self-reported and global positioning systems-measured journey durations: a systematic review. *Transport Reviews*, 33 (4), 443–459.
- Khan, M. & Kockelman, K. M. (2012). Predicting the market potential of plug-in electric vehicles using multiday GPS data. *Energy Policy*, 46, 225–233.
- King, C., Griggs, W., Wirth, F. & Shorten, R. (2013). Using a car sharing model to alleviate electric vehicle range anxiety. In *Proceedings of the sixteenth yale workshop on adaptive and learning systems* (S. 130–135).
- Klambauer, G., Unterthiner, T., Mayr, A. & Hochreiter, S. (2017). Self-normalizing neural networks. In *Advances in Neural Information Processing Systems* (S. 971–980).
- Koztra, B., Büttner, G., Hazeu, G. & Arnold, S. (2017). Updated CLC illustrated nomenclature guidelines. *European Environment Agency: Wien, Austria*, 1–124.
- Kühne, K., Mitra, S. K. & Saphores, J.-D. M. (2018). Without a ride in car country – A comparison of carless households in Germany and California. *Transportation Research Part A: Policy and Practice*, 109, 24–40.
- Kullingsjö, L.-H. & Karlsson, S. (2012). The Swedish car movement data project. In *Proceedings to EVEC Brussels, Belgium, November 19-22, 2012*.
- Lamnek, S. (2005). *Qualitative Sozialforschung: Lehrbuch*. Beltz.
- Lanzendorf, M. (2002). Mobility styles and travel behavior: Application of a lifestyle approach to leisure travel. *Transportation Research Record*, 1807 (1), 163–173.
- Levinson, D. M. & Kumar, A. (1994). The rational locator: Why travel times have remained stable. *Journal of the American Planning Association*, 60 (3), 319–332.
- Liu, J. P., Beyca, O. F., Rao, P. K., Kong, Z. J. & Bukkapatnam, S. T. (2016). Dirichlet process gaussian mixture models for real-time monitoring and their application to chemical mechanical planarization. *IEEE Transactions on Automation Science and Engineering*, 14 (1), 208–221.
- Löchl, M. (2005). Stability of travel behaviour: Thurgau 2003. *Travel Survey Metadata Series*, 16.
- Maaten, L. v. d. & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9 (Nov), 2579–2605.
- Mackett, R. L. (2002). Increasing car dependency of children: Should we be worried? In *Proceedings of the institution of civil engineers-municipal engineer* (Bd. 151, S. 29–38).
- MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Bd. 1, S. 281–297).

- Magdolen, M., Ecke, L., Hilgert, T., Chlond, B. & Vortisch, P. (2020). Identification of non-routine tours in everyday travel behavior [in press]. In *99th Transportation Research Board Annual Meeting, Washington D.C., January 12-16, 2020*.
- Magdolen, M., von Behren, S., Chlond, B., Hunecke, M. & Vortisch, P. (2019). Combining attitudes and travel behavior – A Comparison of Urban Mobility Types Identified in Shanghai, Berlin and San Francisco. In *98th Transportation Research Board Annual Meeting, Washington, D.C., January 13–17, 2019*.
- Marz, N. & Warren, J. (2015). *Big data: Principles and best practices of scalable real-time data systems*. New York; Manning Publications Co.
- Molin, E., Mokhtarian, P. & Kroesen, M. (2016). Multimodal travel groups and attitudes: A latent class cluster analysis of Dutch travelers. *Transportation Research Part A: Policy and Practice*, 83, 14–29.
- Morton, C., Anable, J. & Nelson, J. D. (2017). Consumer structure in the emerging market for electric vehicles: Identifying market segments using cluster analysis. *International Journal of Sustainable Transportation*, 11 (6), 443–459.
- Mosler, K. C., Dyckerhoff, R. & Scheicher, C. (2009). *Mathematische Methoden für Ökonomen*. Springer.
- Motte-Baumvol, B., Massot, M.-H. & Byrd, A. M. (2010). Escaping car dependence in the outer suburbs of Paris. *Urban studies*, 47 (3), 604–619.
- Murphy, K. P. (2013). *Machine learning: A probabilistic perspective*. MIT Press.
- Newman, D., Wells, P., Donovan, C., Nieuwenhuis, P. & Davies, H. (2014). Urban, suburban or rural: Where is the best place for electric vehicles? *International Journal of Automotive Technology and Management*, 14 (3-4), 306–323.
- Nicholas, M. A., Tal, G. & Turrentine, T. S. (2017). Advanced plug-in electric vehicle travel and charging behavior interim report. *Inst Transp Stud*.
- Niklas, U., Magdolen, M., von Behren, S. & Vortisch, P. (2021). Identifying trip purposes on trip level for vehicle sensor data [in press]. In *100th Transportation Research Board Annual Meeting, Washington D.C., January, 2021*.
- Niklas, U., von Behren, S., Chlond, B. & Vortisch, P. (2020). Electric factor – A comparison of car usage profiles of electric and conventional vehicles by a probabilistic approach. *World Electric Vehicle Journal*, 11 (2), 36.
- Niklas, U., von Behren, S., Eisenmann, C., Chlond, B. & Vortisch, P. (2020). Premium factor – Analyzing usage of premium cars compared to conventional cars. *Research in Transportation Business & Management*, 100456.
- Niklas, U., von Behren, S., Soylu, T., Kopp, J., Chlond, B. & Vortisch, P. (2020). Spatial Factor – Using a Random Forest Classification Model to Measure an Internationally Comparable Urbanity Index. *Urban Science*, 4 (3), 36.

- Nobis, C. & Kuhnimhof, T. (2018). *Mobilität in Deutschland - MiD: Ergebnisbericht*.
- Öğdül, H. G. (2010). Urban and rural definitions in regional context: A case study on Turkey. *European Planning Studies*, 18 (9), 1519–1541.
- Olafsson, A. S., Nielsen, T. S. & Carstensen, T. A. (2016). Cycling in multimodal transport behaviours: Exploring modality styles in the Danish population. *Journal of Transport Geography*, 52, 123–130.
- Pagany, R., Ramirez Camargo, L. & Dorner, W. (2019). A review of spatial localization methodologies for the electric vehicle charging infrastructure. *International Journal of Sustainable Transportation*, 13 (6), 433–449.
- Paisley, J. (2010). *A simple proof of the stick-breaking construction of the dirichlet process* (Bericht). Technical report, Princeton University, Department of Computer Science.
- Pan, H., Shen, Q. & Zhang, M. (2009). Influence of urban form on travel behaviour in four neighbourhoods of Shanghai. *Urban studies*, 46 (2), 275–294.
- Panahandeh, G. & Åkerblom, N. (2015). Clustering driving destinations using a modified dbSCAN algorithm with locally-defined map-based thresholds. In *European Congress on Computational Methods in Applied Sciences and Engineering* (S. 97–103).
- Pasaoglu, G., Fiorello, D., Martino, A., Scarcella, G., Alemanno, A., Zubaryeva, A. & Thiel, C. (2012). Driving and parking patterns of European car drivers - A mobility survey. *Luxembourg: European Commission Joint Research Centre*.
- Pasaoglu, G., Fiorello, D., Martino, A., Zani, L., Zubaryeva, A. & Thiel, C. (2014). Travel patterns and the potential use of electric cars – Results from a direct survey in six European countries. *Technological Forecasting and Social Change*, 87, 51–59.
- Pearre, N. S., Kempton, W., Guensler, R. L. & Elango, V. V. (2011). Electric vehicles: How much range is required for a day's driving? *Transportation Research Part C: Emerging Technologies*, 19 (6), 1171–1184.
- Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R. & Ratti, C. (2010). Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *International Workshop on Human Behavior Understanding* (S. 14–25).
- Ramos, É. M. S., Bergstad, C. J. & Nässén, J. (2020). Understanding daily car use: Driving habits, motives, attitudes, and norms across trip purposes. *Transportation Research Part F: Traffic Psychology and Behaviour*, 68, 306–315.
- Raschka, S. & Mirjalili, V. (2017). *Python machine learning*. Packt Publishing Ltd.
- Rasmussen, C. E. (2000). The infinite gaussian mixture model. In *Advances in Neural Information Processing Systems* (S. 554–560).
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.

- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323 (6088), 533–536.
- Schafer, A. (2000). Regularities in travel demand: An international perspective. *Journal of Transportation and Statistics*, 3, 1–31.
- Scherer, M. & Dziekan, K. (2012). Bus or rail: An approach to explain the psychological rail factor. *Journal of Public Transportation*, 15 (1), 5.
- Schönfelder, S. (2006). *Urban rhythms: Modelling the rhythms of individual travel behaviour* (Unveröffentlichte Dissertation). ETH Zurich.
- Schuessler, N. & Axhausen, K. W. (2009). Processing raw data from global positioning systems without additional information. *Transportation Research Record*, 2105 (1), 28–36.
- Seign, R. & Bogenberger, K. (2013). Prescriptions for the successful diffusion of carsharing with electric vehicles. In *Cofat*.
- Shen, L. & Stopher, P. R. (2014). Review of GPS travel survey and GPS data-processing methods. *Transport Reviews*, 34 (3), 316–334.
- Siedentop, S., Roos, S. & Fina, S. (2013). Ist die „Autoabhängigkeit“ von Bewohnern städtischer und ländlicher Siedlungsgebiete messbar?: Entwicklung und Anwendung eines Indikatorenkonzepts in der Region Stuttgart. *Raumforschung und Raumordnung Spatial Research and Planning*, 71 (4), 329–341.
- Smith, A. (2010). *The wealth of nations: An inquiry into the nature and causes of the wealth of nations*. Harriman House Limited.
- Sodenkamp, M., Wenig, J., Thiesse, F. & Staake, T. (2019). Who can drive electric? Segmentation of car drivers based on longitudinal GPS travel data. *Energy policy*, 130, 111–129.
- Soylu, T. (2017). *Räumliche Dynamik der Agglomeration Istanbul*. KIT Scientific Publishing.
- Stopher, P., FitzGerald, C. & Xu, M. (2007). Assessing the accuracy of the Sydney Household Travel Survey with GPS. *Transportation*, 34 (6), 723–741.
- Susilo, Y. O. & Axhausen, K. W. (2014). Repetitions in individual daily activity–travel–location patterns: A study using the Herfindahl–Hirschman Index. *Transportation*, 41 (5), 995–1011.
- Tamor, M. A. & Milačić, M. (2015). Electric vehicles in multi-vehicle households. *Transportation Research Part C: Emerging Technologies*, 56, 52–60.
- Tully, C. J. & Baier, D. (2007). *Mobiler Alltag: Mobilität zwischen Option und Zwang-Vom Zusammenspiel biographischer Motive und sozialer Vorgaben*. Springer-Verlag.
- UN. (2018). World urbanization prospects: The 2018 revision, key facts. *Technical report*.
- VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. O'Reilly Media, Inc.

- van Haaren, R. (2011). Assessment of electric cars' range requirements and usage patterns based on driving behavior recorded in the National Household Travel Survey of 2009. *Earth and Environmental Engineering Department, Columbia University, Fu Foundation School of Engineering and Applied Science, New York*, 51, 53.
- Verreault, H. & Morency, C. (2011). Transcending the typical weekday with large-scale single-day survey samples. *Transportation Research Record*, 2230 (1), 38–47.
- von Behren, S., Bönisch, L., Niklas, U. & Chlond, B. (2020). Linking individuals' affective and instrumental motives to their car use: An application of an integrated choice and latent variable model. In *99th Transportation Research Board Annual Meeting, Washington D.C., January 12-16, 2020*.
- von Behren, S., Minster, C., Esch, J., Hunecke, M., Vortisch, P. & Chlond, B. (2018). Assessing car dependence: Development of a comprehensive survey approach based on the concept of a travel skeleton. *Transportation Research Procedia*, 32, 607–616.
- Weiss, C., Chlond, B., Heilig, M. & Vortisch, P. (2014). Capturing the usage of the German car fleet for a one year period to evaluate the suitability of battery electric vehicles – A model based approach. *Transportation Research Procedia*, 1 (1), 133–141.
- Weiß, C., Chlond, B., von Behren, S., Hilgert, T. & Vortisch, P. (2016). Deutsches Mobilitätspanel (MOP) – Wissenschaftliche Begleitung und Auswertungen Bericht 2015/2016: Alltagsmobilität und Fahrleistung. Hg. v. *KIT-Karlsruher Institut für Technologie*. Online verfügbar unter http://mobilitaetspanel.ifv.kit.edu/downloads/Bericht_MOP_15_16.pdf, zuletzt geprüft am, 3, 2017.
- Wiersma, J., Bertolini, L. & Straatemeier, T. (2017). Adapting spatial conditions to reduce car dependency in mid-sized 'post growth' European city regions: The case of South Limburg, Netherlands. *Transport Policy*, 55, 62–69.
- Wittwer, R. (2014). Zwangsmobilität und Verkehrsmittellorientierung junger Erwachsener: Eine Typologisierung. *Schriftenreihe des Institutes fuer Verkehrsplanung und Strassenverkehr der TU Dresden* (16).
- Wolf, J., Guensler, R. & Bachman, W. (2001). Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. *Transportation Research Record*, 1768 (1), 125–134.
- Wolf, J., Guensler, R., Washington, S. & Frank, L. (2000). The use of electronic travel diaries and vehicle instrumentation packages in the year 2000 atlanta regional household travel survey: Test results, package configurations, and deployment plans. In *9th International Association of Travel Behaviour Research Conference*.
- Wolf, J., Oliveira, M. & Thompson, M. (2003). Impact of underreporting on mileage and travel time estimates: Results from global positioning system-enhanced household travel survey. *Transportation Research Record*, 1854 (1), 189–198.

- Wright, M. N. & Ziegler, A. (2015). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *arXiv preprint arXiv:1508.04409*.
- Xiao, G., Juan, Z. & Zhang, C. (2016). Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization. *Transportation Research Part C: Emerging Technologies*, 71, 447–463.
- Xu, S., Ye, Y. & Xu, L. (2017). Complex power: An analytical approach to measuring the degree of urbanity of urban building complexes. *International Journal of High-Rise Buildings*, 6 (2), 165–175.
- Yan, W. Q. (2019). *Introduction to intelligent surveillance: Surveillance data capture, transmission, and analytics*. Springer.
- Zhao, Q. & Hastie, T. (2019). Causal interpretations of black-box models. *Journal of Business & Economic Statistics*, 1–10.
- Zumkeller, D. (2009). The dynamics of change - 15 years German Mobility Panel. In *88th Transportation Research Board Annual Meeting, Washington D.C., January 11-15, 2009*.
- Zumkeller, D., Chlond, B., Last, J. & Manz, W. (2006). Long-distance travel in a longitudinal perspective: The INVERMO approach in Germany. In *85th Transportation Research Board Annual Meeting, Washington D.C., January 22-26, 2006*.

Anhang

A Hauptkomponentenanalyse

Die Hauptkomponentenanalyse (PCA) ist ein lineares Verfahren zur Reduzierung der Variablenmenge. Anhand von Korrelationen zwischen den Variablen werden die Richtungen maximaler Varianz in hochdimensionalen Daten identifiziert und auf einen dimensionreduzierten Unterraum transformiert (Raschka & Mirjalili, 2017). Formal werden x Variablen mit d Dimensionen ($\mathbf{x} = [x_1, x_2, x_3, \dots, x_d]$, $\mathbf{x} \in \mathbb{R}^d$) zu z Variablen mit k Dimensionen ($\mathbf{z} = [z_1, z_2, z_3, \dots, z_k]$, $\mathbf{z} \in \mathbb{R}^k$) transformiert, wobei $k < d$ gilt.

Um eine Matrix M mit d Dimensionen (Spalten) zu reduzieren, müssen in einem ersten Schritt die Varianzen und Kovarianzen zwischen den Variablen berechnet werden:

$$\begin{aligned}\sigma_j &= \frac{1}{n} \sum_{i=1}^n (x_j^{(i)} - \mu_j)^2 \\ \sigma_{jk} &= \frac{1}{n} \sum_{i=1}^n (x_j^{(i)} - \mu_j)(x_k^{(i)} - \mu_k)\end{aligned}\tag{.1}$$

Die Varianz σ_j entspricht der durchschnittlichen Abweichung vom Mittelwert $\mu_j = \frac{1}{n} \sum_{i=1}^n x_j^{(i)}$. Die Kovarianz σ_{jk} kombiniert die Varianzen zweier Variablen. Basierend auf den berechneten Varianzen und Kovarianzen kann im nächsten Schritt die $d \times d$ Varianz-Kovarianz-Matrix Σ bestimmt werden:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \dots & \sigma_{2d} \\ \vdots & & \ddots & & \\ \sigma_{d1} & \sigma_{d2} & \sigma_{d3} & \dots & \sigma_{dd} \end{bmatrix}\tag{.2}$$

Mit Hilfe der Varianz-Kovarianz-Matrix sind die Eigenwerte λ_d und Eigenvektoren für jede Dimension d bestimmbar (Mosler, Dyckerhoff & Scheicher, 2009). Die Eigenwerte werden in absteigender Reihenfolge sortiert und k Eigenvektoren mit den größten Eigenwerten verwendet, um eine Matrix W mit k Dimensionen (Spalten) zu erstellen. Der Erklärungsgehalt eines Eigenwertes λ_d entspricht dem Anteil an der Summe aller Eigenwerte:

$$\frac{\lambda_d}{\sum_{d=1}^D \lambda_d}\tag{.3}$$

Die Anzahl der Hauptkomponenten hängt von der erklärten Variation im Datensatz ab. Beispielsweise schlägt das Kaiser-Kriterium als Faustregel einen Eigenwert größer 1 vor, d.h. alle Hauptkomponenten mit einem Eigenwert größer eins werden für weitere Berechnungen verwendet. Anhand der kumulierten Variation kann der Erklärungsanteil der ausgewählten Hauptkomponenten bestimmt werden.

B k-Means-Verfahren

Beim k-Means-Verfahren handelt es sich um ein iteratives distanzbasiertes Clusteringverfahren, welches Datenpunkte über Distanzmaße (z.B. euklidische Distanz) Gruppen zuordnet (James et al., 2013; MacQueen et al., 1967). Jede Instanz (z.B. \mathbf{x}) wird als Vektor mit verschiedenen Attributen ($\mathbf{x} = (x_1, x_2, x_3, \dots, x_n)$) repräsentiert.

Die euklidische Distanz zweier Vektoren \mathbf{x}_a und \mathbf{x}_b mit gleichen Attributen ist definiert durch

$$d(\mathbf{x}_a, \mathbf{x}_b) = \sqrt{\sum_{i=1}^n (x_{ai} - x_{bi})^2} = |\mathbf{x}_a - \mathbf{x}_b| \quad (.4)$$

Nachdem die Anzahl der K Cluster festgelegt wurde, werden K Clusterzentren \bar{x}_k so bestimmt, dass die Streuungsquadratsumme

$$SQ(K) = \sum_{i=1}^n \sum_{g \in C_k} d^2(x_i, \bar{x}_k) \quad (.5)$$

minimal ist. Die Streuungsquadratsumme berechnet sich über die Summe der quadrierten euklidischen Distanzen zwischen den Datenpunkten x_i und K Clusterzentren \bar{x}_k . Die Minimierung der Streuungsquadratsumme erfolgt iterativ. Im ersten Schritt werden zufällige Punkte als Clusterzentren definiert. Mit diesen Clusterzentren wird jeder Datenpunkt dem Cluster mit der geringsten euklidischen Distanz zugeordnet. Die Clusterzentren werden angepasst, indem der Mittelpunkt \bar{x}_k für jedes Cluster über die zugeteilten Datenpunkte neu berechnet wird. Dieses Prozedere wird solange wiederholt, bis die Streuungsquadratsumme konvergiert. Um lokale Minima zu vermeiden, findet die Initialisierung der Clusterzentren mehrmals statt. Wenn $SQ(K)$, trotz wiederholter Initialisierung der Clusterzentren, zu ähnlichen Streuungsquadratsummen konvergieren, ist das Risiko lokaler Minima gering.

C Metriken zur Bewertung von Clusterlösungen

Für die Bewertung von Clusterlösungen gibt es verschiedene Verfahren und Herangehensweisen. Wittwer (2014) nutzt u.a. das sogenannte Elbow-Kriterium zur grafischen Bestimmung der optimalen Anzahl an Cluster. Hierbei wird die Streuungsquadratsumme $SQ(K)$ in Abhängigkeit der Clusteranzahl geplottet (Bernard, 1966). Dabei entsteht eine streng monoton fallende Funktion, da $SQ(K)$ automatisch mit ansteigender Anzahl an Clustern abnimmt. Im Extremfall ist $SQ(K) = 0$ wenn jeder Datenpunkt einem Cluster entspricht. Die Erhöhung der Clusteranzahl kann keinen signifikanten Beitrag zur Minimierung von $SQ(K)$ leisten, sobald $SQ(K)$ nicht mehr so stark abnimmt. Visuell ist das an einem Knick zu erkennen. Der Knick und die damit verbundene Anzahl an Cluster ist ein Indiz für die, gegeben den zugrundeliegenden Daten, optimale Clusterlösung.

Ein weiteres Gütemaß zur Bewertung von Clusterlösungen ist der Silhouettenkoeffizient (Rousseeuw, 1987). Der Silhouettenkoeffizient $s(g)$ wird für alle Clusterlösungen von $1-K$ Cluster berechnet. Mit Hilfe der verschiedenen Koeffizienten kann entschieden werden, welche Anzahl an K Cluster die beste Wahl ist und ob die beste Lösung mit K Clustern die Daten gut voneinander trennen kann. Der Silhouettenkoeffizient eines Objektes wird mit

$$s(g) = \frac{b(g) - a(g)}{\max(a(g), b(g))} \quad (.6)$$

definiert. Dabei stellt $a(g)$ die durchschnittliche Distanz zu allen anderen Objekten innerhalb des selben Clusters und $b(g)$ die durchschnittliche Distanz zu Objekten zum nächstgelegenen Cluster dar.

Der Silhouettenkoeffizient $S(C_k)$ eines Clusters C_k ist der Durchschnitt aller Silhouettenkoeffizienten der Objekte eines Clusters und wird mit

$$S(C_k) = \frac{1}{|C_k|} \sum_{g \in C_k} s(g) \quad (.7)$$

berechnet. Der Silhouettenkoeffizient $S(K)$ einer Clusterlösung mit K Clustern ist das Maximum der Silhouettenkoeffizienten aller Cluster C_k und kann wie folgt beschrieben werden

$$S(K) = \max_{C_k} (S(C_k)) \quad (.8)$$

Silhouettenkoeffizienten einer Clusterlösung können Werte im Intervall zwischen -1 und 1 annehmen. Je größer der Wert, desto besser ist die Clusterlösung. Ist der Silhouettenkoeffizient größer als 0,5 bzw. 0,75, entspricht dies einer brauchbaren bzw. starken Clusterlösung.

D Satz von Bayes

Mit Hilfe des Satzes von Bayes lassen sich bedingte Wahrscheinlichkeit effizient über andere, leicht berechenbare Wahrscheinlichkeiten bestimmen. Für die Herleitung des Satzes von Bayes und zur Vereinfachung wird angenommen, dass z und x je zwei Ausprägungen ($z \in \{N, L\}$; $x \in \{A, B\}$) besitzen. Die Gesamtwahrscheinlichkeiten der vier Ausprägungen setzen sich wie folgt zusammen:

$$\begin{aligned}
 p(x = A) &= \frac{n(x = A, z = N) + n(x = A, z = L)}{n} \\
 p(x = B) &= 1 - p(x = A) \\
 p(z = N) &= \frac{n(x = A, z = N) + n(x = B, z = N)}{n} \\
 p(z = L) &= 1 - p(z = N)
 \end{aligned} \tag{.9}$$

$p(x = A)$ setzt sich aus der Anzahl der Objekte in $x = A$, unabhängig von der Ausprägung z , dividiert durch die Anzahl aller Objekte zusammen. Da in diesem Beispiel zwei Ausprägungen betrachtet werden, entspricht $p(x = B)$ der Gegenwahrscheinlichkeit ($1 - p$) von $p(x = A)$. Analog zum beschriebenen Vorgehen lassen sich die Gesamtwahrscheinlichkeiten von $p(z = N)$ und $p(z = L)$ berechnen.

Die gemeinsamen Wahrscheinlichkeiten sind abhängig von der Ausprägung von x und z :

$$\begin{aligned}
 p(x = A, z = N) &= \frac{n(x = A, z = N)}{n} \\
 p(x = B, z = N) &= \frac{n(x = B, z = N)}{n} \\
 p(x = A, z = L) &= \frac{n(x = A, z = L)}{n}
 \end{aligned} \tag{.10}$$

Die gemeinsame Wahrscheinlichkeit stellt immer zwei Zustände (z.B. $x = A$ und $z = N$) zum gleichen Zeitpunkt dar. Beispielsweise entspricht $p(x = A, z = N)$ der Anzahl der Objekte z mit der Ausprägung N , welche sich gleichzeitig in $x = A$ befinden, dividiert durch die Anzahl aller Objekte.

Bedingte Wahrscheinlichkeiten (z.B. $p(z = L|x = A)$) beschreiben einen Zustand (z.B. $z = L$), unter eines bereits eingetretenen Zustands (z.B. $x = A$):

$$p(z = L|x = A) = \frac{n(x = A, z = L)}{n(x = A, z = N) + n(x = A, z = L)} \tag{.11}$$

Im Zähler befindet sich die Anzahl der Objekte $n(x = A, z = L)$, welche die beiden Zustände ($x = A$ und $z = L$) erfüllen. Im Gegensatz zu Gesamtwahrscheinlichkeiten und gemeinsamen Wahrscheinlichkeiten, werden bei bedingten Wahrscheinlichkeiten nicht mehr alle Objekte n zur Berechnung der Wahrscheinlichkeit berücksichtigt, sondern lediglich die Objekte, welche einen bestimmten Zustand ($x = A$) erfüllen ($n(x = A, z = N) + n(x = A, z = L)$).

Die gemeinsame Wahrscheinlichkeit von $p(x = A, z = L)$ kann nun, um den Term ($n(x = A, z = N) + n(x = A, z = L)$) aus der bedingten Wahrscheinlichkeit von Gleichung .11 erweitert und wie folgt ausgedrückt werden:

$$p(x = A, z = L) = \frac{n(x = A, z = L)}{n} \times \frac{n(x = A, z = N) + n(x = A, z = L)}{n(x = A, z = N) + n(x = A, z = L)} \quad (.12)$$

Die Gleichung kann wie folgt umgestellt werden:

$$p(x = A, z = L) = \frac{n(x = A, z = L)}{n(x = A, z = N) + n(x = A, z = L)} \times \frac{n(x = A, z = N) + n(x = A, z = L)}{n} \quad (.13)$$

$$p(x = A, z = L) = p(z = L|x = A) \times p(x = A)$$

Im Ergebnis setzt sich nun die gemeinsame Wahrscheinlichkeit $p(x = A, z = L)$ aus der bedingten Wahrscheinlichkeit $p(z = L|x = A)$ und der Gesamtwahrscheinlichkeit $p(x = A)$ zusammen.

Durch Umformung nach der bedingten Wahrscheinlichkeit entsteht folgende Gleichung:

$$p(z = L|x = A) = \frac{p(x = A, z = L)}{p(x = A)} \quad (.14)$$

Da gemeinsame Wahrscheinlichkeiten symmetrisch sind, kann $p(x = A, z = L)$ durch $p(z = L, x = A)$ ersetzt werden. $p(z = L, x = A)$ entspricht dann der bedingten Wahrscheinlichkeit $p(x = A|z = L)$ multipliziert mit der Gesamtwahrscheinlichkeit $p(z = L)$. Dies entspricht wiederum dem Satz von Bayes:

$$p(z = L|x = A) = \frac{p(x = A|z = L)p(z = L)}{p(x = A)} \quad (.15)$$

Verallgemeinert dargestellt, stimmt die Gleichung .15 mit der Gleichung 4.14 überein:

$$p(z_k = 1|x_i) = \frac{p(x_i|z_k = 1)p(z_k = 1)}{p(x_i)} \quad (.16)$$