Success Factors for Recommender Systems From a Customers' Perspective

Timo Schreiner, Alexandra Rese and Daniel Baier

Abstract Recommender systems have become an integral part of today's ecommerce landscape and are no longer only deployed on websites but also increasingly serve as a basis for the delivery of personalized product recommendations in various communication channels. Within this paper, we present a brief overview of popular and commonly used recommender algorithms as well as current cutting-edge algorithmic advances. We examine consumers' preferences regarding product recommendations in advertisements across different media channels within the apparel industry by applying choice-based conjoint analysis. The findings of studies for young male (n = 170) and female (n = 162) consumers show that the recommender algorithm is not necessarily of upmost importance. In contrast, the advertising channel is of highest relevance with banner advertising being the least preferred channel. Moreover, differences between male and female respondents are outlined. Finally, implications for retailers and advertisers are discussed and a brief outlook on future developments is presented.

Timo Schreiner · Alexandra Rese · Daniel Baier
University of Bayreuth, Chair of Marketing & Innovation
UniversitätsstraSSe 30, 95447 Bayreuth, Germany
✉ timo.schreiner@uni-bayreuth.de
✉ alexandra.rese@uni-bayreuth.de
✉ daniel.baier@uni-bayreuth.de

ARCHIVES OF DATA SCIENCE, SERIES A
(ONLINE FIRST) DOI: 10.5445/KSP/1000098012/02
KIT SCIENTIFIC PUBLISHING ISSN 2363-9881
Vol. 6, No. 2, 2020

1 Introduction

Nowadays, consumers are constantly exposed to various advertisements throughout their everyday lives, both offline and online. The omnipresent exposure to advertisements forces companies and advertisers, especially in an online context, to make their ads as relevant and appealing as possible to increase the advertising effectiveness in terms of conversion (e.g. click-through rates). Therefore, personalization methods that allow for tailoring advertising messages to individual preferences, e.g. based on customers' recent online browsing behavior, are increasingly used by online advertisers and retailers (Bleier and Eisenbeiss, 2015; Estrada-Jiménez et al., 2017).

Recommender systems are a distinct and widespread method of personalization (Kaptein and Parvinen, 2015). They offer benefits for both firms and customers: On the one hand, recommender systems can help to increase product sales by enabling cross- and upselling opportunities, and thus be of great value to firms (Aggarwal, 2016). On the other hand, recommender systems can enhance consumers' decision-making quality in ecommerce and they reduce information overload as well as search costs (Xiao and Benbasat, 2007). Besides their usage on websites and within web shops, they are also used when presenting product recommendations in email campaigns (Linden et al., 2003). Such personalized product recommendations have recently even been successfully deployed in offline print mailings such as package inserts when delivering online orders (Borchers, 2016). In order to maximize the effectiveness of product recommendations in advertisements, companies have to consider several design aspects of recommender systems, such as which algorithm to use or how many recommendations to present at a time (Jugovac and Jannach, 2017; Knijnenburg et al., 2012; Xiao and Benbasat, 2007).

Motivated by the increasing usage of product recommendations within various communication channels, our research goal was to identify the ideal design of personalized product recommendations in advertisements from a customers' perspective. Therefore, we first present a classification scheme of popular, commonly used, and recent recommender algorithms as well as a brief summary of success factors for the design of recommender systems that have been researched so far (Section 2). Then, in Section 3, the research method and design as well as the investigated success factors of recommender systems (attributes and attribute levels) for the choice-based conjoint experiments are outlined, followed by the presentation of the main results. Section 4 closes with a brief discussion of results, potential implications for retailers and advertisers as well as promising future developments in the field of recommender systems.

2 Recommender Systems: Approaches and Success Factors

The term "recommender systems" has its origin in the early 1990s and has been mainly coined by Resnick and Varian (1997). According to a refined definition by Burke (2002, p.331) the term refers to

> " (. . .) *any system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options."*

For achieving the overarching goal of increasing product sales, recommendations need to be relevant to the respective users (Aggarwal, 2016). Next, the recommendation of novel or serendipitous items – recommendations that are unexpected by the consumer – can also be beneficial. Moreover, recommendation sets should include diverse items instead of only similar products for increasing the probability that the consumer will like at least one object from the set: It might, for instance, be unfavorable to present movies of only one specific genre or only t-shirts with similar color and shape within a recommendation set. If users do not like the specific movie genre or rather wish for recommendations of complementary outfits, such highly similar recommendation sets will be rather unsuccessful.

The generation of product recommendations is based on the underlying data sources (Burke, 2002): Background data refers to already existing data such as preferences of other users for certain items or features of specific items. Input data refers to information that needs to be elaborated explicitly or implicitly by the user to the system (e.g. ratings of a specific user for certain items vs. purchase history). Different algorithms can be used for the generation of product recommendations by combining background and input data. On the basis of the data sources used, several recommendation techniques can be distinguished. The two most common and widely used approaches are collaborative-filtering (CF) and content-based filtering (CBF). Furthermore, hybrid recommender systems combining several particular methods are being increasingly used in order to counterbalance disadvantages of single methods by benefits of others (Burke, 2002).

In CF, recommendations for a specific user are based on previous ratings by other users (Adomavicius and Tuzhilin, 2005). Such ratings can either refer to explicitly stated user feedback collected via e.g. numerical rating scales (1–5 star rating), or to implicitly collected user feedback e.g. via unconsciously analyzing the consumers' online shopping behavioral data (Schafer et al., 2001).

By contrast, in CBF, recommendations for a specific user are based on his previous, already known preferences (ratings) for certain features of objects (Adomavicius and Tuzhilin, 2005). In the case of a movie recommender relevant features might for instance be actors, directors or the genre of the movie.

In general, CF and CBF can be classified into heuristics-based approaches where utility predictions are calculated by heuristic methods, and model-based ones which develop – "learn" – a model predicting preferences based on the user database (Adomavicius and Tuzhilin, 2005; Breese et al., 1998).

In *heuristics-based CF*, predictions are directly based on the entire data set of user-item ratings (Breese et al., 1998). Accordingly, there are two ways how predictions of ratings can be retrieved (Aggarwal, 2016):

a) *Item-to-item CF:* Recommendations are based on similar items. Similarity scores of items might, for instance, be positively impacted when products are often purchased together (Linden et al., 2003). This approach is nowadays widely used across various domains mainly inspired by Amazon's successful item-to-item CF approach (Linden et al., 2003).

b) *User-to-user CF:* As opposed to this, recommendations are based on similar users, i.e. users with similar profiles who are providing similar ratings for multiple items (Adomavicius and Tuzhilin, 2005).

Popular algorithms used within the CF approaches include nearest-neighbor classifiers (e.g. cosine, correlation), clustering-based methods as well as graph models (Adomavicius and Tuzhilin, 2005).

Heuristics-based CBF mainly relies on information retrieval methods such as the term frequency-inverse document frequency (TF-IDF) weight which is used to determine the importance of keywords/features within documents/items (Adomavicius and Tuzhilin, 2005). User profiles are then generated by "*an-*

alyzing the content of the items previously seen and rated by the user" (Adomavicius and Tuzhilin, 2005, p.736). Subsequently, for instance, similarity measures can be used for predicting similar items (e.g. cosine similarity measures). Despite their widespread use, these heuristics-based approaches suffer from several issues (Table 1).

Table 1: Major drawbacks of heuristics-based approaches (based on Adomavicius and Tuzhilin (2005); Bobadilla et al. (2013)).

Collaborative filtering	**Content-based filtering**
Cold start issue for new users: The recommender system cannot provide accurate recommendations to new users until the user has rated a sufficient number of items.	
Cold start issue for new items: The CF recommender system is not capable of providing recommendations for new items within the environment until the new item has been rated by a sufficient number of users.	*Limited content analysis:* A CBF recommender system is limited by features that have been explicitly associated with items (either manually or automatically).
Data sparsity / limited coverage: Especially for neighborhood-based algorithms (*k Nearest Neighbors* (kNN) algorithm), the recommendation quality clearly suffers in case of sparse rating data as only few neighbors (for items or users) can be used to predict ratings.	*Overspecialization:* CBF recommender systems tend to recommend items that are highly similar to previously rated items.
Scalability issues: With increasing amounts of data, especially neighborhood-based approaches become too slow.	

Model-based approaches have been developed in order to address major disadvantages of heuristics-based recommender systems. Such

> "(...) *model-based techniques calculate utility (rating) predictions based not on some ad hoc heuristic rules, but, rather, based on a model learned from the underlying data using statistical and machine learning techniques"* (Adomavicius and Tuzhilin, 2005, p.740).

While basically all model-based approaches can be classified as machine learning-based methods, deep learning-based approaches are a more specific sub-field currently receiving a great deal of attention and being widely researched within the recommender systems literature (Figure 1). Deep learning methods can be defined as

> *" representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level"* (LeCun et al., 2015, p. 436).

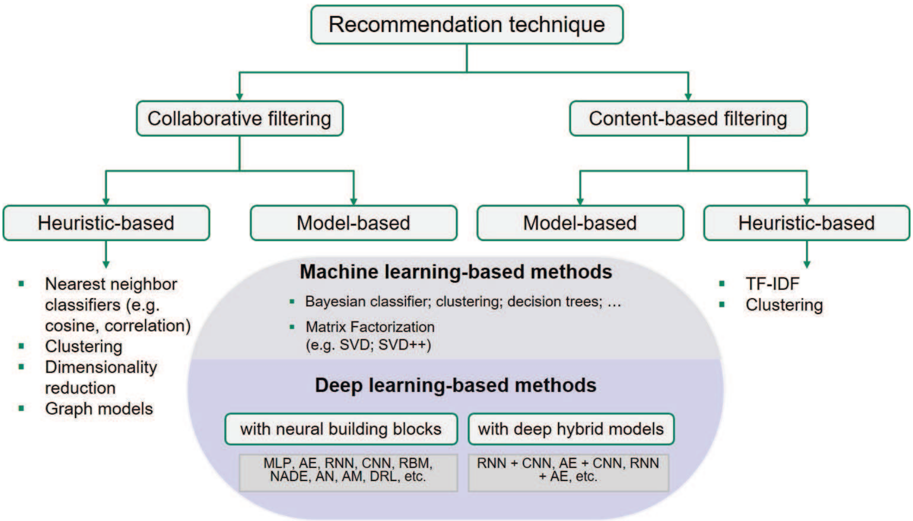

Figure 1: Classification of frequently used recommender algorithms (based on Adomavicius and Tuzhilin (2005); Zhang et al. (2019)). Abbreviations: Multilayer perceptron (MLP); autoencoder (AE); recurrent neural network (RNN); convolutional neural network (CNN); restricted Boltzmann machine (RBM); neural autoregressive distribution estimation (NADE); adversarial networks (AN); attentional models (AM); deep reinforcement learning (DRL).

For recommender systems, deep learning-based approaches can be divided into two categories (Zhang et al., 2019):

1. "*Recommender systems with neural building blocks*":
   Here, Zhang et al. (2019) mainly differentiate between several specific methods ranging from basic feed-forward neural networks (MLP) to more recent developments tailored to specific recommendation is-

sues such as recurrent neural networks (RNN) which are capable of modeling temporal dynamics.

2. "*Recommendation with deep hybrid models*":
   Deep learning methods that combine several specific techniques at a time.

Various researchers have already successfully applied different types of deep learning algorithms to recommender systems in various domains (for an overview see Zhang et al., 2019). For instance, Cheng et al. (2016) created the so-called "*wide & deep*" learning model by a combination of deep neural networks (multilayer perceptrons) with wide, linear models (single layer perceptrons). By doing so, their model is capable of capturing both memorization and generalization, and thus enhancing both the accuracy as well as the diversity of the recommendations (Cheng et al., 2016). The authors evaluated their algorithm within a live environment for the context of app recommendations in Google Play and clearly demonstrated its superiority: Compared to a wide-only algorithm, app acquisitions increased by 3.9 %, while – compared to a deep-only approach – an increase of 1.0 % was observed, too.

Another example of an application of deep learning within recommender systems is the session-based recommender system GRU4Rec which addresses the special issue of generating recommendations when no long-term user data is available (Hidasi and Karatzoglou, 2018; Hidasi et al., 2016). This issue is of high practical relevance, e.g. for smaller online retailers which are not tracking user ID's, when generating recommendations for first time visitors to a website or for domains in which recommendations should particularly refer to short-term user preferences within one session (e.g. news or music recommendations). As commonly used methods such as neighborhood models and matrix factorization methods

> "*are only taking into account the last click of the user, in effect ignoring the information of the past clicks*" (Hidasi et al., 2016, p. 2),

Hidasi et al. (2016) developed a session-based recommender system based on a RNN with Gated Recurrent Units (GRU). The input to the system is the item of the current event in the session and the output is the item of the next event in the session. In an offline experiment for two data sets of videos and click stream data of an ecommerce retailer the authors determined a clear accuracy gain (~ 20–30 %) of the GRU-based approach compared to the best performing baseline algorithm (item-kNN). In addition, a revised version of the GRU4Rec recommender system (using another loss function) clearly outperformed the initial algorithm in a live environment for the recommendation of online videos in terms of watch time (+ 5 %), video plays (+ 5 %) and clicks (+ 4 %; Hidasi and Karatzoglou, 2018).

Those examples clearly illustrate that the utilization of deep learning algorithms might be very beneficial for the success of recommender systems. Nevertheless, besides algorithmic advances in the field of recommender systems and the evaluation in terms of the algorithm's predictive accuracy, current research increasingly emphasizes user-centric evaluation methods (Pu et al., 2011). In user-centric evaluations, the direct interaction of users with a system is measured (Cremonesi et al., 2013; Herlocker et al., 2004). Thus, such evaluations are either based on user survey data or on the analysis of user behavior within a live environment. In user surveys, algorithms are increasingly assessed in terms of various aspects besides accuracy, often including the similarity, novelty, serendipity or diversity of recommendations (e.g. Ekstrand et al., 2014; Said et al., 2013). Moreover, recently more holistic approaches for the evaluation of recommender systems, e.g. considering the entire user experience with such systems, have been presented identifying further success factors beyond algorithms (Jugovac and Jannach, 2017; Knijnenburg et al., 2012; Pu et al., 2011; Schafer et al., 2001; Xiao and Benbasat, 2007).

An extensive literature review (for the detailed overview, see Schreiner et al., 2019) shows that, for instance, also the number of recommendations presented at a time or the provision of an explanation on why certain items are being recommended can have a major impact on the consumers' perception of and willingness to interact with recommender systems:

- As the recommender algorithm defines which products are being recommended, it is a key success factor for a recommender system. Current literature focuses greatly on state-of-the-art algorithms such as deep learning methods. Yet, in practice still rather basic, heuristic-based approaches such as CF are commonly used across websites and ecommerce companies (Smith and Linden, 2017).

- Short captions accompanying the product recommendations (e.g. "customers who bought this item also bought") are often used to explain how recommendations have been generated, thus increasing transparency of and trust in the system (Herlocker et al., 2000).

- The amount of products presented within one recommendation set might also clearly impact the success of a recommender system. However, there is no consensus yet in the current literature whether a large or a small number of recommendations might be more beneficial (Schreiner et al., 2019).

Therefore, for the study at hand, it was of high interest to examine these success factors from a customers' perspective.

3 Empirical Study: Research Methods and Results

Based on previous research on success factors for designing recommender systems and also taking into account current literature dealing with the effectiveness of advertisements in different media channels (e.g. Baek and Morimoto, 2012; Yu and Cude, 2009), we deployed choice based conjoint analysis (CBC) to determine the ideal design of product recommendations in advertisements from a customers' perspective. CBC was considered to be the most suitable approach for our research as it allows for a collection of customers' preferences in a very realistic way (Cohen, 1997).

In CBC, respondents have to select their most preferred option from a set of alternatives including the possibility to select a "none" option - indicating their aversion to all other presented stimuli (Cohen, 1997; Louviere and Woodworth, 1983). This choice decision is repeated several times and the respondents' overall evaluations of objects are subsequently decomposed into part worth utilities for specific attributes as well as attribute levels (Green and Srinivasan, 1978). Besides highly relevant success factors for designing product recommendations, namely the underlying recommender algorithm (levels: CF algorithm vs. recommendation of bestselling products), the number of recommendations presented at a time (levels: 4 vs. 8 vs. 12) as well as the explanation accompanying the recommended products (levels: specific item-style explanation vs. unspecific explanation), different media channels (levels: package inserts vs. email advertising vs. banner advertising) as well as specific providers/retailers (levels: Amazon vs. a local mail-order company:

Baur vs. a fictitious company: Vestes Deis) have been included as attributes for our CBC experiment.

A product in the apparel industry, i.e. the bestselling pullover at Amazon on November 21st, 2017 for males and females respectively, was chosen as field of application for our study. Choosing a product from the apparel industry seemed especially suitable for our research context as recommender systems are commonly deployed by leading apparel online retailers such as Zalando or Amazon within their online shops as well as within their communication with customers through email newsletters or banner advertisements. Two CBC experiments have been created – one for males and one for females – and have been analyzed in comparison to identify relevant differences between both genders. Such a gender-specific investigation seemed to be very promising as previous literature points to clear differences between men and women in terms of their fashion shopping behavior and motivations (Blázquez, 2014). For generating product recommendations, Amazon's recommendations for the corresponding pullover as well as other best selling pullovers were taken (Schreiner et al., 2019). All attributes and attribute levels used for the CBC have been presented solely visual by creating 108 (3^3 x 2^2) different stimuli per experiment. For instance, the advertising channel was visualized by integrating the product recommendations in the image of a package insert, an email interface of a renowned German email provider or in the banner advertisement of a German news portal.

A reduced design was created using Sawtooth Software by deploying the *balanced overlap* method which enables a moderate degree of level overlap and provides reliable estimates of main effects.

After instructing the respondents to imagine having purchased a specific, displayed pullover previously online, they had to complete 16 choice tasks in which they had to decide whether they would consider a product recommendation or not. Four so-called holdout tasks served for evaluation of validity.

The data collection took place at one faculty of a mid-sized German university on four days in November and December 2017 via an online-aided survey. The target group of the survey were students as part of the group of so-called *Digital Natives* – young adults born after 1980 that have grown up with the internet and digital technologies. After data cleaning of two respondents who either completed the survey faster than half of the average survey duration or were older than 37 (hence, not part of the target group of Digital Natives), a total of

332 students remained for analysis. 170 respondents were male (51.2 %) and 162 were female (48.8 %) representing the population of students in the faculty well in terms of gender. An overwhelming majority of study participants was aged 23 years or younger (76.2 %), thus mainly born 1994 or later. For data analysis the *Analysis Manager* of Sawtooth's Lighthouse Studio as well as IBM SPSS Statistics version 21 were used and led to the following results:

With regard to internal validity the root likelihood (RLH) values were greater than 0.5 and satisfactory for both samples. The same holds for the mean first choice hit rates (FCHR) and predictive validity near 80 % (Table 2).

Table 2: Goodness of fit and predicitve validity of the utility estimation (source: Schreiner et al. (2019)).

	Male (n = 170)	**Female (n = 162)**
RLH		
Aggregate	0.727	0.707
Individual	0.736	0.724
FCHR		
Holdout task 1	74.12 %	74.07 %
Holdout task 2	79.41 %	70.37 %
Holdout task 3	77.06 %	83.95 %
Holdout task 4	85.88 %	82.10 %
Mean	79.12 %	77.62 %

The results of the CBC/HB estimation illustrated in Table 3 clearly show that the advertising channel is by far the most important attribute for males and females when deciding whether to use or follow product recommendations in advertisements.

While banner advertising is least preferred by both subgroups, males prefer ads in package inserts and females email advertising the most. The second most important attribute for both samples is the number of recommendations presented at a time. For males, the smallest set of four product recommendations is of greatest utility whereas females prefer the largest set of twelve recommendations. In terms of the underlying recommender algorithm, there are also significant differences between both groups. The algorithm is almost as important as the number of recommendations to females. However, overall, females only slightly prefer recommendations generated by the CF algorithm (with a great variance/standard deviation in utility scores on an individual level). By contrast, the utility of the product recommendations is far less influenced by the recommender algorithm for males. Beyond that, recommendations of bestselling products even outperfom recommendations generated by the item-to-item CF algorithm for the male sample.

Table 3: CBC/HB results: part-worth utilities and attribute importances for both samples in comparison (source: Schreiner et al. (2019)).

Attributes, levels	Importance (%) / part-worth utility / (Standard deviation)	
	Male (n = 170)	**Female** (n = 162)
Advertising channel**	47.06 % (18.5383)	42.52 % (16.4170)
Package inserts	41.6981 (86.9194)	27.2990 (92.1058)
Email advertising*	18.3888 (91.8203)	35.3836 (62.6270)
Banner advertising	−60.0869 (119.8352)	−62.6827 (100.2497)
Algorithm***	11.47 % (8.0500)	18.52 % (13.3744)
CF***	−20.0924 (28.7406)	5.8049 (56.9336)
Bestselling product***	20.0924 (28.7406)	−5.8049 (56.9336)
Explanation	7.41 % (5.4502)	6.69 % (4.7139)
Item style***	−1.9414 (22.9594)	−9.3738 (18.2219)
Unspecific***	1.9414 (22.9594)	9.3738 (18.2219)
Number of recommendations	20.49 % (12.8676)	18.89 % (9.0242)
4***	41.3089 (51.9067)	−8.7204 (47.2899)
8	−22.7051 (29.5661)	−21.3111 (26.8464)
12***	−18.6038 (43.8844)	30.0315 (41.8294)
Provider	13.57 % (7.2459)	13.38 % (7.0440)
Amazon	23.4981 (33.0486)	25.5799 (29.1617)
Baur	−13.2115 (27.4731)	−17.1631 (19.9480)
Vestes Deis	−10.2866 (24.9089)	−8.4169 (29.0588)
“None” option	169.5986 (220.7524)	100.1100 (121.5072)

***, **, * indicate two-sided significant differences of importances or part-worth utilities between both groups at $p < 0.01$, $p < 0.05$ and $p < 0.1$, respectively.

4 Discussion, Implications and Outlook on Future Developments

These results lead to a few recommendations for action for advertisers and retailers:

1. First of all, personalized product recommendations for men should contain as few relevant items as possible (up to a maximum of four), whereas the recommendation set for women should entail significantly more products (at least twelve). One reason for the different preferences regarding the number of recommendations might be that females might have a higher level of involvement with apparel as males. However, it is important to note here that results might differ for other products and domains indicating promising possibilities for future research.
2. Secondly, retailers should increasingly focus on designing personalized product recommendations in email advertising and package inserts instead of only relying on banner ads. The relatively high part-worth utilities for advertising in package inserts illustrate that traditional (print) advertising media must also not be disregarded for younger, digitally-savvy audiences.
3. Thirdly, while currently a major focus is on tailoring product recommendations to individual or segment-specific product needs by applying a personalized recommender algorithm, our research demonstrates that the underlying algorithm is not necessarily of utmost importance. Accordingly, retailers and advertisers have to assure that product recommendations will also be personalized to individual preferences with regard to other design aspects such as the number of recommendations, the advertising channel or the degree of personalization of the text accompanying the product recommendations (e.g. personalized vs. unpersonalized greetings in email newsletters).

   ⇒ This implication is in line with Jeff Bezos' vision of personalized online shops from more than 20 years ago. In an interview with the *Washington Post* the founder and CEO of Amazon pronounced:

> "*If we have* 4.5 *million customers, we shouldn't have one store.* (...) *We should have* 4.5 *million stores*" (Jeff Bezos in Walker, 1998).

Despite this early idea of personalized marketing, such an extreme form of online personalization "with a target segment of size one" is still far away from being reality (Arora et al., 2008, p.306). An early prototypical implementation of such a personalized user interface within a university context is discussed by Geyer-Schulz et al. (2001).

More recently, a state-of-the art industry example from Netflix shows that companies nowadays are already taking into account also other aspects when delivering personalized recommendations (Gomez-Uribe and Hunt, 2015): Netflix uses a combination of different recommendation algorithms on its website to deliver relevant, novel as well as diverse movie recommendations. Recommendations are presented in different rows – each deploying a different recommender algorithm. Furthermore, pages are constructed using another personalized algorithm

> "*taking into account the relevance of each row to the member as well as the diversity of the page*" (Gomez-Uribe and Hunt, 2015, p. 4–5).

Consequently, each Netflix member sees an individually designed homepage in terms of page layout when logging into his or her Netflix account. By doing so, according to Gomez-Uribe and Hunt (2015) 80 % of hours streamed at Netflix are triggered by its own recommender systems.

4. Last but not least, the widespread used item-to-item CF algorithm might not necessarily be a beneficial approach by default for all domains and use cases. For the specific sector of apparel, our research suggests that a similarity-based CF approach does not necessarily lead to ideal product recommendations from a customers' perspective as females only slightly prefer the CF algorithm and male respondents even prefer the recommendation of bestselling products over the recommendations generated by the CF algorithm. Answers to an open question further support this finding: Approximately one-fifth of all answers referred to

the desire to receive recommendations for complementary products or suggestions for entire outfits *from head to toe*. Accordingly, future research should evaluate recommendations for apparel products in advertisements generated by other algorithms that e.g. also take aspects like diversity, novelty and serendipity of recommendations more into account.

⇒ *Complementary recommender systems* which aim at recommending items that are not similar to the previously bought / viewed item but are often times bought together with it are a currently trending area in recommender systems addressing this specific issue (Hwangbo et al., 2018; Yu et al., 2019). The major challenge in such complementary recommender systems is to identify relevant complementary products as the pure co-purchase of items might not be sufficiently defining supplements (Hwangbo et al., 2018). For instance, some co-purchase relationships might only work unidirectional: A power bank is often times bought as a supplement to mobile phones. Yet, mobile phones are not bought as a supplement to a power bank.

⇒ Another currently highly relevant development in the recommender systems literature having the potential to overcome this issue are multi- and cross-domain recommender systems which enable the recommendation of products from one domain or product category, e.g. music/shoes, also within another domain, e.g. movies/pullovers (Cantador et al., 2015; Cremonesi et al., 2011; Khan et al., 2017). Alternatively, recommendations can be generated on a joint basis of two domains (e.g. music and movies).

Transferring knowledge acquired in one domain to another might especially be beneficial for large ecommerce retailers and platforms like Amazon or eBay. Such systems might reduce cold-start issues for new users and items and help to create cross-selling opportunities for products from different domains. Crucial here is the identification of two or more highly related domains with reference to user preferences (Cantador et al., 2015; Cremonesi et al., 2011).

With an ever increasing amount of available data sources and customer information it is also becoming increasingly important to define the right composition of data and variables that should be taken into account by the recommender algorithm in order to reach the important goal of optimizing the individual recommendation quality. Context-aware recommender systems are a current development that seem to be very promising with reference to this goal. Such systems are capable of considering contextual information when generating recommendations such as user profiles, time, location, purpose of purchase, social situation, emotions, mood, etc. (Haruna et al., 2017). By adding contextual information to the traditionally used data sources (users and items) for predicting user's preferences, such approaches have the potential to increase the degree of personalization and individual fit of the recommended items tremendously. Major challenges in designing such context-aware recommender systems might be to identify the most relevant contextual information per user, product category and domain: While for the illustrated use case of apparel products gender was identified to be a relevant contextual information, this might, for instance, much less or not at all apply to other domains such as recommendations for music or articles in scientific journals.

Finally, another future challenge or rather opportunity will be to leverage new types of data sources such as speech data in conversational commerce from voice assistants (Baier et al., 2018) or user-generated content within reviews, blogs or comments in social networks (Chen et al., 2015).

As illustrated by our research and current developments in the field of recommender systems, there are still a lot of open issues that need to be addressed for enhancing the efficiency and quality of recommender systems from a customers' perspective. We hope that our investigation can help to advance research on the design of product recommendations in advertisements as well as deliver some food for thought on future research directions.

References

Adomavicius G, Tuzhilin A (2005) Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering 17(6):734–749. DOI: 10.1109/TKDE.2005.99.

Aggarwal CC (2016) Recommender Systems: The Textbook. Springer, Cham (Switzerland). ISBN: 978-3-319296-57-9, DOI: 10.1007/978-3-319-29659-3.

Arora N, Dreze X, Ghose A, Hess J, Iyengar R, Jing B, Joshi Y, Kumar V, Lurie N, Neslin S, Sajeesh S, Su M, Syam N, Thomas J, Zhang Z (2008) Putting One-to-one Marketing to Work: Personalization, Customization, and Choice. Marketing Letters 19(3-4):305–321. DOI: 10.1007/s11002-008-9056-z.

Baek TH, Morimoto M (2012) Stay Away From Me: Examining the Determinants of Consumer Avoidance of Personalized Advertising. Journal of Advertising 41(1):59–76. DOI: 10.2753/JOA0091-3367410105.

Baier D, Rese A, Röglinger M (2018) Conversational User Interfaces for Online Shops? A Categorization of Use Cases. In: Proceedings of the 39th International Conference on Information Systems (ICIS2018), Association for Information Systems, Atlanta (USA), pp. 1–17. ISBN: 978-0-996683-17-3.

Bleier A, Eisenbeiss M (2015) Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where. Marketing Science 34(5):669–688. DOI: 10.1287/mksc.2015.0930.

Blázquez M (2014) Fashion Shopping in Multichannel Retail: The Role of Technology in Enhancing the Customer Experience. International Journal of Electronic Commerce 18(4):97–116. DOI: 10.2753/JEC1086-4415180404.

Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Recommender Systems Survey. Knowledge-Based Systems 46:109–132. DOI: 10.1016/j.knosys.2013.03.012.

Borchers D (2016) Personalisieren in Print. ONEtoONE 9(16):48. URL: `https://prudsys.de/wp-content/uploads/2016/10/ONEtoONE_Klingel_Case_09_2016.pdf`.

Breese JS, Heckerman D, Kadie C (1998) Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In: Cooper GF, Moral S (eds.), Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI'98), Morgan Kaufmann Publishers Inc., San Francisco, pp. 43–52. ISBN: 978-1-558605-55-8.

Burke R (2002) Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction 12(4):331–370. DOI: 10.1023/A:1021240730564.

Cantador I, Fernández-Tobías I, Berkovsky S, Cremonesi P (2015) Cross-Domain Recommender Systems. In: Ricci F, Rokach L, Shapira B (eds.), Recommender Systems Handbook. Springer, Boston (USA), pp. 919–959. DOI: 10.1007/978-1-4899-7637-6_27.

Chen L, Chen G, Wang F (2015) Recommender Systems Based on User Reviews: The State of the Art. User Modeling and User-Adapted Interaction 25(2):99–154. DOI: 10.1007/s11257-015-9155-5.

Cheng HT, Koc L, Harmsen J, Shaked T, Chandra T, Aradhye H, Anderson G, Corrado G, Chai W, Ispir M, Anil R, Haque Z, Hong L, Jain V, Liu X, Shah H (2016) Wide & Deep Learning for Recommender Systems. In: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems (DLRS2016), Association for Computing Machinery (ACM), New York (USA), pp. 7–10. DOI: 10.1145/2988450.2988454.

Cohen SH (1997) Perfect Union: CBCA marries the best of conjoint and discrete choice models. Marketing Research 9(1):12–17.

Cremonesi P, Tripodi A, Turrin R (2011) Cross-Domain Recommender Systems. In: 2011 IEEE 11th International Conference on Data Mining Workshops, Institute of Electrical and Electronics Engineers (IEEE), New York (USA), pp. 496–503. DOI: 10.1109/ICDMW.2011.57.

Cremonesi P, Garzotto F, Turrin R (2013) User-Centric vs. System-Centric Evaluation of Recommender Systems. In: Kotzé P, Marsden G, Lindgaard G, Wesson J, Winckler M (eds.), Human-Computer Interaction – INTERACT 2013, Springer, Berlin, Heidelberg (Germany), Lecture Notes in Computer Science, Vol. 8119, pp. 334–351. DOI: 10.1007/978-3-642-40477-1_21.

Ekstrand MD, Harper FM, Willemsen MC, Konstan JA (2014) User Perception of Differences in Recommender Algorithms. In: Proceedings of the 8th ACM Conference on Recommender Systems (RecSys ’14), Association for Computing Machinery (ACM), New York (USA), pp. 161–168. DOI: 10.1145/2645710.2645737.

Estrada-Jiménez J, Parra-Arnau J, Rodríguez-Hoyos A, Forné J (2017) Online Advertising: Analysis of Privacy Threats and Protection Approaches. Computer Communications 100:32–51. DOI: 10.1016/j.comcom.2016.12.016.

Geyer-Schulz A, Hahsler M, Jahn M (2001) Educational and Scientific Recommender Systems: Designing the Information Channels of the Virtual University. International Journal of Engineering Education 17(2):153–163.

Gomez-Uribe CA, Hunt N (2015) The Netflix Recommender System: Algorithms, Business Value, and Innovation. ACM Transactions on Management Information Systems 6(4):1–19. DOI: 10.1145/2843948.

Green PE, Srinivasan V (1978) Conjoint Analysis in Consumer Research: Issues and Outlook. Journal of Consumer Research 5(2):103–123. DOI: 10.1086/208721.

Haruna K, Akmar Ismail M, Suhendroyono S, Damiasih D, Pierewan AC, Chiroma H, Herawan T (2017) Context-Aware Recommender System: A Review of Recent Developmental Process and Future Research Direction. Applied Sciences 7(12). DOI: 10.3390/app7121211.

Herlocker JL, Konstan JA, Riedl J (2000) Explaining Collaborative Filtering Recommendations. In: Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW ’00), Association for Computing Machinery (ACM), New York (USA), pp. 241–250. DOI: 10.1145/358916.358995.

Herlocker JL, Konstan JA, Terveen LG, Riedl JT (2004) Evaluating Collaborative Filtering Recommender Systems. ACM Transactions on Information Systems (TOIS) 22(1):5–53. DOI: 10.1145/963770.963772.

Hidasi B, Karatzoglou A (2018) Recurrent Neural Networks with Top-k Gains for Session-based Recommendations. In: Cuzzocrea A, Allan J, Paton N, Divesh Srivastava RA, Broder A, et al. (eds.), Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18), Association for Computing Machinery (ACM), New York (USA), pp. 843–852. DOI: 10.1145/3269206.3271761.

Hidasi B, Karatzoglou A, Baltrunas L, Tikk D (2016) Session-based Recommendations with Recurrent Neural Networks. In: Proceedings of the International Conference on Learning Representations (ICLR '16), pp. 1–10. URL: `https://arxiv.org/pdf/1511.06939.pdf`. ArXiv:1511.06939.

Hwangbo H, Kim YS, Cha KJ (2018) Recommendation System Development for Fashion Retail E-Commerce. Electronic Commerce Research and Applications 28:94–101. DOI: 10.1016/j.elerap.2018.01.012.

Jugovac M, Jannach D (2017) Interacting with Recommenders-Overview and Research Directions. ACM Transactions on Interactive Intelligent Systems (TiiS) 7(3):1–46. DOI: 10.1145/3001837.

Kaptein M, Parvinen P (2015) Advancing E-Commerce Personalization: Process Framework and Case Study. International Journal of Electronic Commerce 19(3):7–33. DOI: 10.1080/10864415.2015.1000216.

Khan MM, Ibrahim R, Ghani I (2017) Cross Domain Recommender Systems: A Systematic Literature Review. ACM Computing Surveys (CSUR) 50(3):36–1-36:34. DOI: 10.1145/3073565.

Knijnenburg BP, Willemsen MC, Gantner Z, Soncu H, Newell C (2012) Explaining the User Experience of Recommender Systems. User Modeling and User-Adapted Interaction 22(4-5):441–504. DOI: 10.1007/s11257-011-9118-4.

LeCun Y, Bengio Y, Hinton GE (2015) Deep Learning. Nature 521(7553):436–444. DOI: 10.1038/nature14539.

Linden G, Smith B, York J (2003) Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Internet Computing 7(1):76–80. DOI: 10.1109/MIC.2003.1167344.

Louviere JJ, Woodworth G (1983) Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data. Journal of Marketing Research 20(4):350–367. DOI: 10.1177/002224378302000403.

Pu P, Chen L, Hu R (2011) A User-centric Evaluation Framework for Recommender Systems. In: Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11), Association for Computing Machinery (ACM), New York (USA), pp. 157–164. DOI: 10.1145/2043932.2043962.

Resnick P, Varian HR (1997) Recommender Systems. Communications of the ACM 40(3):56–58. DOI: 10.1145/245108.245121.

Said A, Fields B, Jain BJ, Albayrak S (2013) User-centric Evaluation of a K-furthest Neighbor Collaborative Filtering Recommender Algorithm. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13), Association for Computing Machinery (ACM), New York (USA), pp. 1399–1408. DOI: 10.1145/2441776.2441933.

Schafer JB, Konstan JA, Riedl J (2001) E-Commerce Recommendation Applications. Data Mining and Knowledge Discovery 5(1-2):115–153. DOI: 10.1023/A:1009804230409.

Schreiner T, Rese A, Baier D (2019) Multichannel Personalization: Identifying Consumer Preferences for Product Recommendations in Advertisements Across Different Media Channels. Journal of Retailing and Consumer Services 48:87–99. DOI: 10.1016/j.jretconser.2019.02.010.

Smith B, Linden G (2017) Two Decades of Recommender Systems at Amazon.com. IEEE Internet Computing 21(3):12–18. DOI: 10.1109/MIC.2017.72.

Walker L (1998) Amazon Gets Personal With E-Commerce. The Washington Post, p. H1. URL: `https://www.washingtonpost.com/wp-srv/washtech/daily/nov98/amazon110898.htm`.

Xiao B, Benbasat I (2007) E-commerce Product Recommendation Agents: Use, Characteristics, and Impact. MIS Quarterly 31(1):137–209. DOI: 10.2307/25148784.

Yu H, Litchfield L, Kernreiter T, Jolly S, Hempstalk K (2019) Complementary Recommendations: A Brief Survey. In: 2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), Institute of Electrical and Electronics Engineers (IEEE), New York (USA), pp. 73–78. DOI: 10.1109/HPBDIS.2019.8735479.

Yu JH, Cude B (2009) "Hello, Mrs. Sarah Jones! We recommend this product!" Consumers' Perceptions About Personalized Advertising: Comparisons Across Advertisements Delivered via Three Different Types of Media. International Journal of Consumer Studies 33(4):503–514. DOI: 10.1111/j.1470-6431.2009.00784.x

Zhang S, Yao L, Sun A, Tay Y (2019) Deep Learning Based Recommender System: A Survey and New Perspectives. ACM Computing Surveys (CSUR) 52(1):1–38. DOI: 10.1145/3285029.