Karlsruhe Reports in Informatics 2013,8
Edited by Karlsruhe Institute of Technology,
Faculty of Informatics
ISSN 2190-4782

Group Activity Recognition using Belief Propagation for P2P Mobile Devices

(Technical Report, September 2nd 2013)

Dawud Gordon, Markus Scholz, Michael Beigl

2013

KIT – University of the State of Baden-Wuerttemberg and National Research Center of the Helmholtz Association

Fakultät für Informatik

Group Activity Recognition using Belief Propagation for P2P Mobile Devices

(Technical Report, September 2nd 2013)

Dawud Gordon, Markus Scholz, Michael Beigl
Faculty of Computer Science
Karlsruhe Institute of Technology
Karlsruhe, Germany
{Firstname.Lastname}@kit.edu

***Abstract*—Human are social beings and spend most of their time in groups. Group behavior is emergent, generated by members' personal characteristics and their interactions making it difficult to recognize in peer-to-peer (P2P) systems where the emergent behavior itself cannot be directly observed. We introduce 2 novel algorithms for distributed probabilistic inference (DPI) of group activities using loopy belief propagation (LBP). We evaluate their performance using an experiment in team sports activities and show that these activities are emergent in nature through natural processes. Centralized recognition performs very well, upwards of an F-score of 0.95 for large window sizes. The distributed methods iteratively converge to solutions which are comparable to centralized methods, even surpassing them in some situations. DPI-LBP also greatly reduces energy consumption of the node, where a centralized unit or infrastructure is not required, although memory and processor consumption increases.**

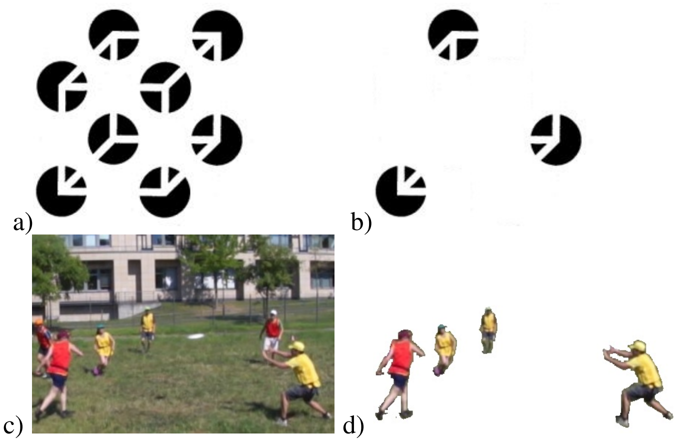

Fig. 1. Following Gestalt theory, an image of a cube emerges from distinct objects (**a**). An incomplete set obscures the emergent properties (**b**). The same is valid for group activities where from a complete image a sport can be identified from context (**c**). A partial view of player's physical behavior without context makes identification difficult (**d**).

I. Introduction

Human beings are social creatures, and as such we spend most of our time in groups [1]. It has been shown that groups are better than individuals at accomplishing tasks, which is often why they are formed in the first place [2]. Understanding group behavior and context is then crucial for intelligent environments. The process of understanding what a group is doing, or the physical attributes of the group behavior, is called group activity recognition (GAR) [3].

The behavior of the group is emergent behavior, emerging from the personal characteristics of the individual members and the group dynamic [4], [2]. Human perception of group behavior can be explained by Gestalt Theory, where only when observing the complete whole can its properties be described (see Fig. 1) [4], [2]. Recognition of that behavior is irrefutably bound to human perception, as it is the human who labels a group activity based on his/her perception. Kurt Lewin, a pioneer of modern social psychology, uses the term "emergence" to signify that the properties of the behavior of the group are fundamentally different than the properties of the behavior of the individuals, or of the "sum" of those behaviors [4], a definition which we follow. This is a generalization of many definitions of the term emergence [5], where all agree that emergence is a difference between (human) observations of micro and macro properties.

Mobile devices such as Smart Phones present an attractive platform both for human activity recognition (HAR) and the recognition of emergent group activities. Sensor information from these devices is used by a recognition algorithm to learn the ability to make the same observations as a human would. This paper shows that a global observer - a centralized detection algorithm - having the complete picture can perform detection of emergent group actives. It then analyzes if a local observer - a decentralized algorithm running on individual devices - having limited peer-to-peer communication with other peers can also deliver such observations and studies how well such local detection performs in comparison with a centralized approach. It also studies the communication range required to detect the emergent behavior with respect to the spatial size of the group, and if sparse communication can still reach acceptable detection rates compared to a global observer. We also study how much energy can be saved using the decentralized approach and how much energy needs to be invested for local processing instead.

We present novel methods for distributed GAR using distributed probabilistic inference (DPI) combined with loopy belief propagation (LBP) [6]. For each group activity, the behavior is broken down into individual clusters using unsupervised methods. Each node then estimates its belief over its local clusters for all group activities given current sensor observations, and then communicates this information to its neighbors.

All nodes then iteratively update and re-communicate their beliefs based on their local sensory evidence, the belief estimates received, and a model of individual-to-individual group dynamics. The network then iterates and converges towards a response prediction. We present two methods for LBP, one linear regression over soft posterior probabilities over user behaviors (SLBP) and one using expectations based on hard classifications (HLBP).

The novel algorithms are evaluated using an experiment in team sports. 10 subjects play 6 different sports and are monitored using Android phones as wearable sensors. The experiment naturally creates emergent group behavior where the algorithms are then evaluated in terms of their effectiveness at recognizing that behavior. The evaluation is in terms of performance with respect to the number of iterations required for convergence. The effects of P2P communication range are also evaluated by simulating local links using the devices' GPS locations. The results are then compared with centralized inference of group behavior, where the complete set of sensor data provides the complete picture of the emergent behavior.

The results show that centralized inference of emergent group behavior when presented with the complete set of group sensor data is relatively straight-forward, approaching an F-score of 0.81 for a window length of 2 seconds and 0.96 for 10 seconds. However inference using solely the data of each subject individually is poor at around 0.55 for the same window. The novel DPI-SLBP approach begins at iteration 0 at the same value as with individual subject inference, but then rapidly improves with each iteration, surpassing the centralized naive Bayes approach after three iterations and converging to an F-score of 0.84 after about 10 iterations. This method incurs an increase in the amount of local memory consumed and processing required, but reduces the amount of energy required overall for classification factor of almost 7.

The simplified DPI-HLBP algorithm performs similarly but converges to a lower value of 0.81, just under the centralized approach. However compared to DPI-SLBP, memory consumption energy required for classification drop by a factor of more than 6, which is 40 times less than that required by the centralized approach.

II. Related Work

Recognizing emergent behavior has historically been a topic in HAR for quite some time, although it has not been named as such directly. Human perception of other human activities is also governed by Gestalt Theory in that we observe a single individual instead of a collection of limbs, therefore inferring behavior of a single individual from the distributed behavior of their body is emergent [7]. The problem is however simpler, as limbs don't change roles, and there interactions with each other are mechanical in nature.

Multi-user activity recognition (MAR) is the process of recognizing the activities of multiple individuals in parallel [3]. Wearable sensing approaches leverage centralized inference structures to infer multiple activities in parallel. Subjects may be interacting with each other or may even be in the same group, but the problem presented is of a different nature, recognizing distinct activities for different subjects [8], as opposed to emergent group activities. However often these approaches gray the boundaries between MAR and GAR, where some of the activities recognized are group labels of emergent activities, where behavior of multiple subjects is necessary to infer certain activities, and others are single-user activities [9].

Approaches have also been presented to distributing the recognition process for contexts and activities across the network of sensing nodes [10]. Distributed methods leverage knowledge about the conditions which govern distributed sensing to fuse information into recognition, e.g. someone climbing a fence will create similar disturbances at multiple measurement locations [10]. However these approaches are not focused on emergent behavior. One approach which was inspirational for our research here is a method of distributed probabilistic inference for sensor calibration [11]. The approach uses the assumption that the distance between the measurement locations of nodes will provide temperature measurements which are correlated with each other, over which a potential function can be built. The approach is however fundamentally different from emergent GAR, as it does not address the human factor, where this factor is the main cause of complexity in HAR as a field in general. It is more akin to MAR, where each node must estimate its own bias under the assumption that measurements are correlated. For that application, loopy belief propagation does not converge, requiring a complex networking architecture for clique structuring and belief propagation [12].

Other sensing modalities have also been used for recognizing group activities. Video systems present an advantage as they are able to view local individual behavior and the resulting emergent group behavior simultaneously [13], and are also able to scale to larger groups. They are also able to measure certain properties of individual roles, for example a player's position in an American Football team [14]. However such approaches are accompanied with infrastructure requirements for communicating and processing the constant flow of video data, and therefore can only be applied in instrumented environments. Many human interactions are verbal, and monitoring these conversations using microphones also provides insight into the group activity [15]. An understanding of the audio situation can even allow extraction of certain types of role information present in the group behavior [16]. However for activity recognition, microphones are an orthogonal sensing approach as they do not sense the physical parameters of the behavior directly, and extracting this information from audio data is a different branch of research with its own set of challenges.

Monitoring location has also been shown to give insight into emergent properties of larger groups or crowds [17]. Here emergent spacial properties can be computed as a function of the location of multiple individuals and the properties of the space in which they are located. Adding motion sensors also allows properties such as affiliation of users to each other and to groups, building subgroups within a larger group or crowd [18]. Emergent behavior has also been addressed in the separate but related field of swarm intelligence, usually addressing this behavior in animals and insects [19]. Here the problems addressed usually have one of three different goals, either looking to simulate the emergent group behavior based on models of individuals (generation) [20], discover the rules governing individuals based on the emergent behavior produced (discovery) [21], or evaluate the correctness of assumptions about the relationship between local agents and emergent group behavior (evaluation) [20]. Our approach here differs from this field because we wish to **predict the emergent group behavior** based on observations of agents (humans) who are admittedly far too complex to model using expert knowledge. We therefore approach the problem from a machine learning standpoint in order to discover and model pertinent characteristics of agents in an automatic fashion, using only the sensing devices.

III. Concepts and Approach

In this section we present the concepts and theories which motivate the design decisions made. We begin with the fundamental principles which govern group behavior from the field of group dynamics and social psychology. Inspired by these abstract models and theories, we construct concrete models and methods for modeling and classifying group behavior in a probabilistic fashion. The goal of this section is to create models for centralized and distributed recognition of emergent group behavior, methods for evaluating them independently, and a metric for judging the degree of emergence of a recognition problem given specific models.

A. From Field Theory to Probability Theory

Kurt Lewin's "Field Theory" [4] states that the individual behavior $B^{\text{ind.}}$ of members of a group is a function of their individual attributes and characteristics c and the social environment of the group E. He quantified this as "interactionism" in Eq. (1).

$$B^{\text{ind.}} = f(c, E) \tag{1}$$

He stated that the resulting group behavior is "a dynamic whole [that] has properties which are different from the properties of [its] parts or from the sum of [its] parts" [4]. "According to Lewin, whenever a group comes into existence, it becomes a unified system with emergent properties that cannot be fully understood by piecemeal examination" [2]. However, the behavior of an individual is not only governed by their individual attributes, but also their role in the group dynamic [22]. These roles, as with group behavior, are generated as emergent norms when the group is formed, and members adapt their behavior to fit the norms for different roles [2]. As a result we can update Lewin's equation to account for emergent roles $\rho \in R$: $B^{\text{ind.}} = f(c, \rho, E)$.

From a probabilistic standpoint, we can model the probability p of all group behaviors $p(B)$ as the joint probability of all individuals. We know that this is the joint distribution of C, R and E which symbolizes the social dynamic:

$$p(B) = p(B^{\text{ind.}}_{s_1}, B^{\text{ind.}}_{s_2}, \ldots, B^{\text{ind.}}_{s_n}) = p(C, R, E) \tag{2}$$

where C is the set of characteristics of all group members $c \in C$. When Lewin used this term, he was referring to all possible relevant characteristics of the individual, psychological, sociological, physiological, metaphysical, etc., meaning the state space of C approaches infinite. However, for activity recognition we focus on the physical characteristics of contexts, activities and behaviors. These physical properties can be observed and differentiated using sensors (the premise for HAR), therefore we make the assumption that we can replace the infinite state-space of C with our observations of the physical properties of C, referred to as X. Here we use the notation x^{τ}_s to indicate a single observation, or observations over a window, for subject s at time τ. X_s refers to all observations for subject s, X^b refers to the evidence of all subjects for a single group activity, and X is the complete set of observations for all subjects and activities. We now have the following equation for the joint probabilities of group activities: $p(B) = p(B^{\text{ind.}}_{s_1}, B^{\text{ind.}}_{s_2}, \ldots, B^{\text{ind.}}_{s_n}) = p(X, R, E)$. We can break down the right hand side to approach the problem of differentiating $b \in B$ given observations and models as the following:

$$p(B|X, R, E) = \prod_{s_i \in G} p(B|X_{s_i}, \rho_{s_i}, E) \tag{3}$$

However, we still have the role of each user in the equation. Identification and annotation of roles and individual-to-role affiliation requires behavioral experts, meaning this approach lacks versatility and requires a great deal of preparation. Also, the double annotation of group activity, and role greatly increases the effort required for training. To circumnavigate this issue, we make a key assumption. The evidence X is conditionally dependent on both the individuals characteristics, and the role of the individual in the group [2]. We can therefore use this conditional dependence to gain the pertinent information about the observations and roles. This is done by combining the evidence in its conditionally dependent form using a transformation into a different space:

$$K = f(X \otimes R) = \forall_{b \in B} \forall_{s \in G} \text{clust}(X_s|b) \tag{4}$$

We cluster the evidence into clusters $\kappa \in K$, where κ^b_s is a cluster from subject s generated by group behavior b and their role ρ^b_s in that behavior. To be clear, we are not making the assumption that these clusters equate semantically to the role of the individual in the activity. Our assumption is that the clusters contain a factorization of the conditional dependencies between the evidence, the roles and the group behavior, or $p(K|B) = p(X, R|B)$. For example, assuming experts in the sport soccer inform us that one of the roles is goal-tender, no single cluster would equate to this role for a specific subject. The assumption is that the role goal-tender for a specific subject will however generate one or several clusters in which the different modalities in which this user behaves in this role are quantified. It is also possible that a similar behavior from the same or different subject in the same or different group activity could generate a cluster of the same dimensions.

B. Modeling and Classifying Group Activities

The clustering approach used is a probabilistic clustering using Expectation Maximization. For each group activity and subject, X is separated into X^b_s and then clustered, yielding clusters K^b_s. The probability density function (PDF, or P) of the clusters for a subject and group activity is given by a Gaussian mixture model (GMM) [2]:

$$P(X^b_s|K_s) = \sum_{\kappa^b_s \in K^b_s} \pi_{\kappa^b_s} \mathcal{N}(X_s|\mu_{\kappa^b_s}, \Sigma_{\kappa^b_s}) \tag{5}$$

Each node s has clusters K_s where each cluster κ_s is generated by a certain group behavior b, giving a subset of clusters for each group activity $\kappa^b_s \in K^b_s$. These clusters now build the evidence function for inference of group activities. The posterior probability distribution $p(K|X)$ can be obtained using Bayesian inference, where each posterior is normalized using the following equation:

$$p(\kappa_s^b|x_s^\tau) = \underbrace{\text{Post}(\kappa_s^b|x_s^\tau)}_{\text{GMM posterior}} \underbrace{\frac{\text{like}(K_s^b|x_s^\tau)}{\sum_{b' \in B} \text{like}(K_s^{b'}|x_s^\tau)}}_{\text{GMM likelihoods normalization}} \tag{6}$$

Here posteriors are generated over the Gaussian mixtures for each class K_s^b given an observation x_s^τ, after which the posterior distribution is normalized by the likelihood of all activity cluster models for that subject. Both the likelihood of a GMM and the posterior of a cluster given an observation are obtained by applying Bayesian inference and the Law of Total Probability [6]. It is important to note that due to the normalization in Eq. (6), the resulting probability distribution over all clusters for all activities for each subject (K_s) sums to 1. As will be explained later on, this step is necessary in order for nodes to be able to learn relative probability distributions of neighboring nodes based on histories of these distributions generated by observations. Classification of the current group activity at any point in time for a single subject is achieved by Eq. (7).

$$p(B = b|K_{s_i}, x_{s_i}^\tau) = \sum_{\kappa_{s_i}^b \in K_{s_i}^b} p(\kappa_{s_i}^b|x_{s_i}^\tau) \tag{7}$$

The classification approach of evaluating local posteriors using local evidence (Eq. (6)) can be used to evaluate the ability of a single node to infer the group activity based on local observations alone, which we call the **independent local inference (ILI)** method.

Returning to the original problem of inferring group behavior, we have now combined the user's role with the evidence in clusters. We now have the following equation: $p(B) = p(K, E)$. The term E is problematic, since it is difficult to quantify. We know, however, that group behavior can be observed by applying Gestalt Theory, meaning that observation of the whole allows it to appear in its emergent form, rather than as a sum of unrelated aspects. The indication is that for complete set of K, the effects of E are already present with respect to the interpretation of the group behavior B. The same concept can be seen in Fig.1, where presented presented with the complete image, a cube appears through the Gestalt principles in 1a. This is the emergent whole with properties different than the sum of the individual circles, which are actually not circles, but appear as such on the right in in 1b. Therefore observing all distributed observations in a single location should allow a complete view of the emergent group behavior. To examine this hypothesis, we used two methods of central inference.

The first is Bayesian inference using the complete probability distributions of K. For this purpose, $\bar{\xi}$ is constructed such that:

$$\bar{\xi}^\tau := \underset{K_s \in K}{\forall} \text{append}(p(K_s|x_s^\tau)) \tag{8}$$

For each time-step τ, $\bar{\xi}^\tau$ is then a vector of the complete normalized posters across K. Using this set as observations, a naive Bayesian classifier is constructed to model $P(\bar{\xi}|B)$ and then to infer $p(B|\bar{\xi}^\tau)$ for each time-step τ. This method is referred to as **centralized cluster-based inference (CCI)**. A more concrete description of the training process will be explained in Sec. IV. The second centralized method a more traditional naive Bayesian inference method using the observations directly. Here $P(X|B)$ is modeled as a GMM using the Expectation Maximization (EM) algorithm [2], and $p(B|X)$ can then be inferred, referred to hereon as **centralized naive Bayes (CnB)**.

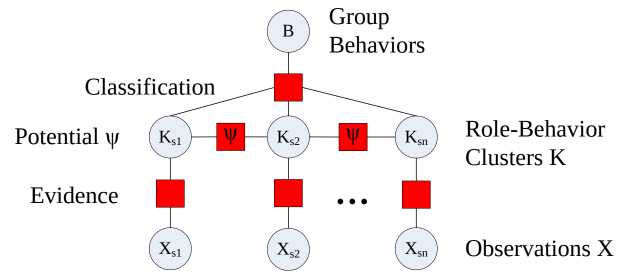

Fig. 2. Factor Graph for DPI-LBP with Evidence $p(K, X)$, Potential $\psi(K_{s_i}, K_{s_j})$ (for All i, j, Some are Omitted), and Classification $p(B, K)$

Our goal is to recognize emergent group behavior using distributed mobile phones, where E can no longer be ignored. We propose to approach this problem using DPI with LBP. The missing information sampled by other nodes which is necessary in order to infer the emergent behavior is propagated through the network in the form of beliefs from other nodes. The equation for exact inference is shown in Eq. (9) where each node calculates its own belief based on its evidence, as well as its belief of other nodes states based on its own local evidence. Evidence is propagated through the network in the form of posteriors known as beliefs

$$p(K|X) = \prod_{s_i \in G} \underbrace{p(K_{s_i}|X_{s_i})}_{\text{local evidence}} \prod_{s_j \neq i \in G} \underbrace{p(K_{s_i}|X_{s_j})}_{\text{distributed belief}} \tag{9}$$

This method has the advantage of being exact, meaning the accuracy achieve is equal to that of a centralized system [6]. However, the state space of all random variables must be modeled redundantly at every node at process at each iteration step. More attractive are methods of approximate inference where each node propagates beliefs for other nodes based on its internal beliefs and a model for relations between its random variables and those of other nodes [6].

For standard DPI problems, clique graphs can be built to factor priors using some form of expert knowledge or assumptions about conditional independence between nodes [11]. These clique graphs are structured as directed a-cyclical graphs (DAG) and then traversed for belief propagation, guaranteeing that loops do not occur. Constructing a recognition system in this manner is guaranteed to converge to the optimal solution with respect to a centralized system with a full sensory image of the emergent behavior. For group activity recognition this is not the case, as all variables within are influenced by the group dynamic E, making the entire group a single clique graph.

One approach which may or may not work in such situations is loopy belief propagation (LBP) where cyclical belief propagation paths are allowed. However several problems may occur depending on the inference problem. Several types of inference problems do not converge to single solution, and it is unclear which types of problems do and do not converge [6]. Also, the convergence rate, meaning how many iterations of belief propagation are required for convergence, are unknown. Luckily for emergent GAR, the system does converge in relatively few iterations with a resulting high accuracy, as we will see in Sec. V. The equation for loopy and non-loopy belief propagation is given in Eq. (10).

$$p(K|X) = \prod_{s_i \in G} \underbrace{p(K_{s_i}|X_{s_i})}_{\text{local evidence}} \prod_{s_{j \neq i} \in G} \underbrace{\psi_{i,j}(K_{s_i}, K_{s_j})}_{\text{potential function}} \quad (10)$$

The potential function ψ can be any positive function which defines the relationship between the variables at subject s_i and s_j [2]. For this function we used linear regression [6] to model the relationship between the variables of each pair of subjects, or K_{s_i} and K_{s_j}. As stated before, the evidence function is trained using EM for unsupervised clustering of each subjects data for each group activity. Each potential function is trained using linear regression from the variables K_{s_j} of other subjects to each cluster κ_{s_i} separately. The resulting linear mapping takes the form:

$$\psi_{i,j} = \underset{\kappa_{s_i} \in K_{s_i}}{\forall} \; \underset{s_{j \neq i} \in G}{\forall} \alpha + [\beta_1, \beta_2, \ldots, \beta_n] \times [p(K_{s_j})] \quad (11)$$

Where $[p(K_{s_j})]$ is a column vector of all cluster posteriors $\kappa_{s_j} \in K_{s_j}$. This method we call **DPI with soft LBP (DPI-SLBP)** due to the "soft" posterior probability distributions which are mapped.

Each iteration consists of a local inference step followed by several update and classification steps. In the inference step, each node s_i generates a posterior distribution over its clusters using its local evidence function from Eq. (10), creating an initial estimate of the group activity based only on local estimates. In the first update step, this information is propagated to all neighboring nodes s_j, i.e. all nodes within range of one-hop communication. These nodes then convert this estimation of the posterior probability distribution over K_{s_i} to a belief over K_{s_j} using the mappings generated from Eq. 11. These beliefs are then combined with the current beliefs of node s_j over K_{s_j} and the resulting classification of the group behavior is reevaluated using Eq. (7) in the classification step. The update and classification steps then repeated until the network is satisfied that convergence has been reached, where we will empirically evaluate how man update steps are required in Sec. V.

We also present a simplified version of the aforementioned DPI with LBP approach. That method requires each node to broadcast its posterior $p(K_s|X_s)$ to all neighboring nodes. Probabilistic classification works on the assumption that the most likely model given specific evidence is the correct model for a given instance. Based on this assumption, the most valuable information $p(K_s|X_s)$ is the most likely cluster in the most likely activity, namely $\text{argmax}_{k^b_{s_j}} \, p(k^b_{s_j})$. We present a simplified method where beliefs are calculated using only this information, instead of the full cluster posteriors $p(K_s|X_s)$. This simplified method takes the same form as Eq. (10) with a modified potential function presented in Eq. (12).

$$\psi^{\text{simp.}}_{ij} = p(K_{s_i} \,|\, \underset{k^b_{s_j}}{\text{argmax}}\, p(k^b_{s_j})) \quad (12)$$

Training for the simplified potential model is done by calculating the expectation $\mathbb{E}$ instead of the method using regression previously introduced. Training $\psi^{\text{simp.}}_{ij}$ for node s_i to s_j for $\text{argmax}_{k^b_{s_j}} \sum_{k^b_{s_j} \in K^b_{s_j}} p(k^b_{s_j}) = \kappa^b_{s_j}$ is done by creating a posterior probability for $p(K_{s_j})$ given the posteriors of instances of training data data where the most like behavior for node s_i is $\kappa^b_{s_j}$. Intuitively, we model a belief for the behavior of node j at times when node i is behaving in a specific manner. For example, if node i is behaving as a goal keeper in a soccer game, the belief that node j is playing soccer as a midfielder would (assumedly) be higher than than the belief that node j is serving a volleyball. The equation for computation of $\psi^{\text{simp.}}_{ij}$ is the following. First we define $\overline{\kappa}_{s_i}$ to be the most probable role-behavior cluster for s_i:

$$\overline{\kappa}_{s_i} = \underset{k^b_{s_i} \in K^b_{s_i}}{\text{argmax}}\, p(k^b_{s_i}|x^t_{s_i}) \quad (13)$$

Then, for each cluster κ_{s_i} the expectation is calculated given K_{s_i} and $x^\tau_{s_i}$:

$$\psi^{\text{simp.}}_{ij}(\kappa_{s_i}) = \underset{\tau | \overline{\kappa}_{s_j} = \kappa^b_{s_j}}{\forall} \mathbb{E}(\kappa_{s_j}|K_{s_i}, x^\tau_{s_i}) \quad (14)$$

where the probability of $K_{s_i}, x^\tau_{s_i}$ is given by Eq. (6). We refer to this method as **DPI with hard LBP (DPI-HLBP)** due to the hard role-behavior classification in the potential function. Lewin's definition of emergence in group behavior as the whole having properties different than the parts or the "sum" of those parts [4], and emergence is a function of observational difference between the micro and the macro [5]. We define a metric for evaluating this disparity. For a physical activity recognition system, trained to recognize a set of group activities identified by human observations, we define the **"degree of emergence"** ϵ as the proportional information gain, quantified using the F-score, of activity recognition with the complete picture, to the mean of activity recognition of all nodes using their local observations.

$$\epsilon(B|X) = \frac{\text{F-score}(p(B|X)) - \frac{\sum_{s \in G} \text{F-score}(p(B_s|X_s))}{|G|}}{\text{F-score}(p(B|X))} \quad (15)$$

This measure is dependent on and specific to the models used, the subjective observations (labels), and only for the behavior recognition problem, and does not necessarily be generalized over these parameters, other definitions of emergence, or other recognition problems.

IV. Experiment and Procedure

To evaluate the approach detailed in the preceding section we constructed an experiment with emergent group activities. The activities performed were team sports, where the emergent behavior is the sport being played itself, based on the observations of the physical behavior of the individuals.

The devices used LG Nexus 4 Android devices with a custom application. The software sampled the accelerometer,

Fig. 3. Device Placement (Left) for the Team Sports Group Activity Experiment (Middle) with Relative Subject Locations on the Field (Right)

magnetometer and gyroscope, each a 3 axis vector value, with the maximum sampling rate. The accelerometer measures on-body acceleration, the magnetometer delivers orientation and heading information relative to the local ambient magnetic field, and the gyroscope samples rotation information. Effectively a sample rate of about 50 Hz was delivered for the accelerometer and gyroscope, while the orientation sensor only delivered an approximate sample rate of 20 Hz. In addition to the behavioral sensing, the devices sampled their absolute location using the GPS sensor. The location information was not used for group activity recognition, but was used for simulation of performance of a the P2P recognition system.

The devices where attached at the right side of the hip, as the hip has been shown to be the most beneficial single location for activity sensing [7]. This was done using an elastic sports belt for the device, where the phones where inserted into the belt with the face outwards and the top of the phone forward as shown in Fig. 3. 6 different team sports where performed by all subjects: **volleyball, badminton, football (soccer), ultimate Frisbee, touch rugby, and flunky-ball**. Each sport was performed for 10 minutes, with a break between each type of sport. The experiment was conducted outdoors in an open field with a natural turf of dimensions 15m by 20m, and a video recording was made from an elevated standpoint of the experiment.The day was sunny with high temperatures around $29°C$, making breaks between activities necessary.The subjects were made up of 7 males and 3 females. On a scale of 0 (no experience in any of the sport) to 10 (very experienced in all of the sports) over all sport types, the average experience was 4.5 with a variance of 3.5.

The data recorded was synchronized and input into an offline sensor replay mechanism in a MATLAB simulation environment, where the algorithms are implemented. 50% of the data is used to train the algorithms, and the other 50% for evaluating algorithmic performance. All sensor measurements where then hold-resampled to 50 Hz to provide equidistant measurements for feature calculation. GPS location annotations where also resampled and smoothed to account for asynchronous updates. This sensor data was cut into windows of lengths from 1 to 10 seconds, where the window is advanced by 0.5 seconds each iteration over which features where calculated. The features used were the mean and variance of the total acceleration signal, the mean and variance of the azimuth orientation with respect to the subject's body, and the mean and variance of the rotation around the X and Z axes (see Fig. 3 for orientation). These features calculated for subject s then represent the observations X_s of the subject, where τ is the last timestamp of a sensor data window. For each window length, all models are retrained and reevaluated using the features generated over the windows.

Based on these locations we simulated performance under different communication capabilities. We then simulated performance for a communication range ϕ of 5m, 10m, 15m, and 20m sequentially, compared to the diagonal of the field of 25m which is also a good approximation of the radius of the group. We used the relative Euclidean distance between two subjects dist(s_i, s_j) based on their GPS coordinates, and judged them to be able to communicate if dist$(s_i, s_j) \leq \phi$. The timestamp used to evaluate dist(s_i, s_j) is the final timestamp of the window τ, as this is the point where the network is able to evaluate the distributed evidence functions and communicate beliefs. we simulated performance with full inter-connectivity of all nodes in the network, meaning the range local P2P communication was greater than the maximum distance between any two subjects during the course of the experiment, or $\phi = \infty$. No multi-hop communication is implemented, simulated or required for the methods presented here. The results are generated using only loopy belief propagation and the models previously trained for this window length.

During the course of the simulation we evaluate the F-score of the described algorithms by constructing confusion matrices over the output of the algorithm for each node. How the output is determined based on a node's belief is described in Eq. (7). This is monitored for each node, at each iteration of the belief propagation algorithm. We also monitored the processor time required for each operation and iteration, as well as the memory required to store and process information.

V. Evaluation

The goal of the algorithms presented in Sec. III is to allow distributed mobile devices sensing the physical activities of individuals to be able to recognize the activity of the group. Since group behavior is emergent, the correct response is not dependent on any single node, but the combine implications of all distributed measurements. The presented DPI with LBP methods may be a solution, but there are open questions in the literature about their performance under GAR circumstances. For one, it is unclear if the algorithms will converge to response. If they do, it is unclear what the accuracy of that response will be, or how long it will take to converge.

A. Centralized Recognition of Emergent Group Behavior

To analyze performance of centralized inference of emergent group behavior using the complete picture of sampled

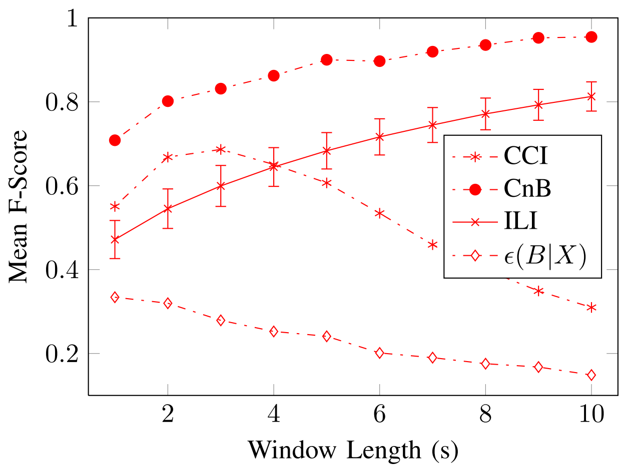

Fig. 4. Performance for Centralized Algorithms, Distributed Independent Classification and the Resulting Degree of Emergence

sensor data, we looked at 3 different approaches which where explained in Sec. III, namely CnB, CCI, ILI and the degree of emergence ϵ of this specific problem.

The results of the centralized analysis are presented in Fig. 4 where performance is shown in the form of the F-measure over varying lengths of the feature analysis window. The CnB algorithm performs the best, with F-measures of 0.71 for a window of 1 second, increasing up to a recognition rate of 0.96 for a window of 10 seconds. The implications are that for the given scenario and set of conditions, the emergent group behavior can be recognized using relatively straight forward methods, *if* observations of all members of the group are present. Admittedly, there are many other issues in GAR which are not present in this experiment, such as variance of group members and the number of group members over time, device location, etc. [23], however these problems are outside the scope of this work.

The CCI approach yields an F-score of 0.52 for an observation window length of 1 second, with an optimum of 0.70 for a window length of 3 seconds, after which it subsides towards random classifications with an F-score of 0.31 at 10 seconds. This would appear to indicate that posteriors over role-behavior clusters do not contain the pertinent information required to infer group behavior. However, as we will see later, the is not the case. The implication is therefore only that naive Bayesian inference is not the correct method for inference using these posteriors. This is due to the fact that Bayesian inference using GMMs separates the data probabilistically using EM for clustering, but the posteriors themselves do not separate well into such clusters.

The evaluation of the accuracy of the ILI method provides insight into the nature of the experiment. For a window size of 1 second, the mean F-measure of all nodes across all experiments was 0.48, with a variance of 0.05. For 10 seconds, the mean increases to 0.82 and variance drops slightly to 0.03. The longer the time-line of data used to classify the group activity, the better the group activity can be recognized, both for the centralized as distributed evidence functions. Also the quantified emergence of the group activity shrinks with the size of the window from 0.34 for 1 second to 0.16 for 10 seconds.

Sports activities in general are very dynamic in nature, where players change roles rapidly. For a longer observation time, a single player may change roles enough, allowing a classifier to observe the majority of role-behaviors from a single subject in that time, and therefore improve classification of the emergent behavior. This effect cannot be generalized to other forms of group activities such as social gatherings or meetings and is specific to the experiment conducted here. For the remaining evaluation of the novel distributed algorithms, a window size of 2 seconds has been selected, as it represents a a good level of emergence, and none of the algorithms in Fig. 4 have saturated or reached their peak results, allowing us to compare relative values.

B. *DPI with LBP*

The results of DPI with LBP for a window size of 2 seconds and a communication range of $\phi = \infty$ are displayed in Fig. 5. The shape of the curve presented demonstrates clearly that the distributed algorithm does indeed converge to a solution. This solution is reached after 15 iterations at an F-measure of 0.86. At iteration 0, the lower bound is given by the evaluation of the local evidence functions of each node separately, and corresponds to the value for a window size of 2s in Fig. 4. This value even exceeds the centralized approach at 0.81 after 3 iterations where 95% of convergence, a value of 0.84 is already reached after 6 iterations. It must be noted here that the indication is not distributed inference performs better, but that the potential performance using posteriors over K is higher than the performance of a nB classifier over X. The standard deviation across nodes is 0.045 for iteration 0, but drops to 0.027 already after one iteration and then converges to a value of 0.021.

The results of DPI with LBP with the same parameters ($\phi = \infty$, ws = 2) but with the simplified potential function $\psi^{\text{simp.}}$ is shown in Fig. 6. Iteration 0 also begins at the same lower bound as in Fig. 5. A similar convergence is also clearly visible, but convergence occurs at 0.80, as compared to a value of 0.86 for the full potential method. The standard deviation also drops dramatically after one iteration from 0.045 to 0.037, and then iteratively converges to 0.031. This value is however greater than the standard deviation of 0.021 for the regression-based potential function. Here again, 95% of convergence is reached fairly quickly after 5 iterations.

The effects of the simplified potential function are clear. Convergence occurs slightly faster (1 iteration less for 95%), but converges to an optimum 7% less than when using a full regression-based potential function, and the standard deviation across nodes also increases by 68%. As we will see later, the reduced F-measure and increased standard deviation come with reductions in resource consumption, where the performance trade-off can be advantageous for certain applications.

C. *Effects of P2P Communication*

The two novel distributed methods where also simulated for various communication ranges. The range ϕ was simulated for 5m, 10m, 15m, 20m, and ∞, or full connectivity. The mean F-measure results for regression-based potential function are displayed in Fig. 7. There the man value for $\phi = \infty$ corresponds to the same mean in Fig. 6. Mean values for 20m and 15m

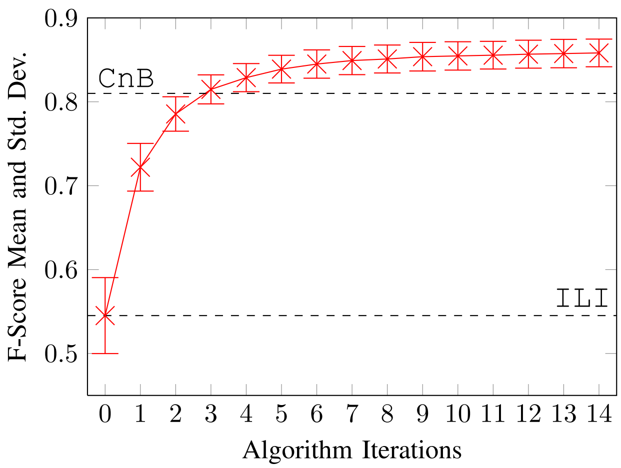

Fig. 5. DPI-SLBP for a Window of 2 Seconds and Full Inter-Connectivity ($\phi = \infty$)

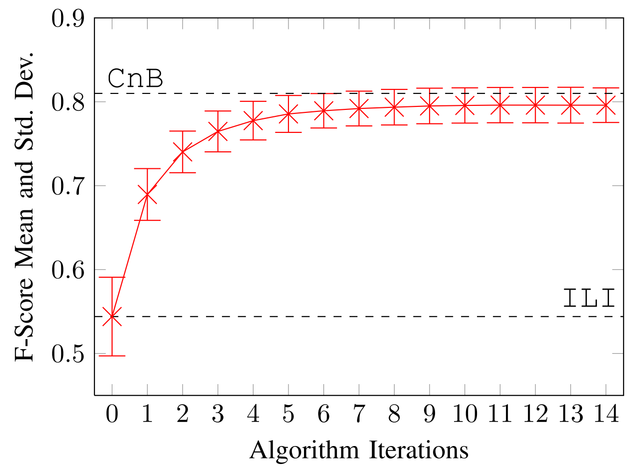

Fig. 6. DPI-HLBP for a Window of 2 Seconds and Full Inter-Connectivity ($\phi = \infty$)

perform similarly to full connectivity, converging at an almost identical rate to a value of 0.85 and 0.84 respectively, compared to 0.86 for full connectivity. Reducing communication to 10m however incurred larger losses, converging to a value of 0.80, although with an identical rate of convergence as well. At 5m, convergence only achieved and F-measure of 0.68, although the rate of convergence remained constant.

Similar behavior was also observed for performance using the simplified potential function for the same simulated communication distances in Fig. 8. Communication ranges of 20m and 15m iteratively incur a loss of less than one F-measure point, although 95% of convergence requires one further iteration, namely 6 iterations. At 10m, convergence occurred at an F-measure of 0.76 with 95% reached after 8 iterations. Reducing communication further to 5m also required 8 iterations and converged to an F-measure of 0.65.

A survey of convergence values for both algorithms after 5 iterations can be seen in Tab. I, where the coverage is simply the ratio of the of ϕ to the diameter of the group, assumed to be the diagonal of the field 25m. From full connectivity to 15m range there is little effect on the convergence times, although the using the simplified potential function incurred a greater reduction of 4.9 percentage points (pp) as apposed to 2.4 pp for regression-based potentials. This effect is due to the speed of belief propagation for the two algorithms. For $\psi^{\text{simpl.}}$ the propagation takes more "effort" as a node must receive enough belief contrary to its current state before its internal belief about its must probable cluster changes. For the regression-based approach, this occurs more quickly as beliefs are integrated and propagated in a continuous manner. For these communication ranges, the large majority of nodes are in the same network with occasional disconnection of individuals as they leave the group, e.g. to collect the ball. Hence, only the small changes in recognition rates over these ranges as belief propagates over intermediary nodes throughout the network.

For a communication distance of 10m, both algorithms propagate information at the same speed as before, but the network breaks apart into disjoint sub-networks as groups of nodes and individuals are out of range of each other. This is also the cause of the reduced recognition rates in Figs. 7 and 8 for a range of 10m, where necessary information cannot propagate to all nodes due to the lack of a link between nodes in different sub-networks. For a range of 5m the problem is exacerbated as the network breaks up into many different subgroups, and nodes only have one or two other nodes in their neighborhoods, many disjunct neighborhoods appear. The results can be seen clearly in the low convergence rates in Figs. 7 and 8. However, convergence occurs quickly, as beliefs are only propagated to small subgroups of G.

TABLE I. Convergence in % After 5 Iterations

Range	Coverage (%)	Convergence SLBP (%)	Convergence HLBP (%)
∞	100	91.2	94.6
20m	80	89.6	91.5
15m	60	88.8	89.7
10m	40	86.5	86.5
5m	20	91.1	91.7

D. Resource Consumption Analysis

The resource consumption is only for recognition, where training would incur higher costs and is more efficient when conducted in a centralized manner. The values presented here are simplified approximations, calculated from the bitrate and power consumption of different communication technologies [24][1]. The results of the embedded resource consumption analysis for the different approaches are presented inn Tab. II. For the CCI and CnB algorithms, we simulated communication of local information to a centralized instance using 3G networks. For the DPI algorithms, 10 iterations are assumed which is well over the amount required for 95% convergence presented in Tab. I. Here DPI-SLBP reduced power consumption due to communication by 84% compared to CnB, and DPI-HLBP presents a reduction of 97.5%. In terms of time required for a classification, DPI-SLBP increases response time by a factor of 2.5, although server-side calculations for CnB are not taken into account [3]. DPI-HLBP however reduces the reaction time of the system by 51% with respect to CnB, which is around 5.5 times less then the reaction time of DPI-SLBP. It is important

[1] http://www.csr.com/sites/default/files/white-papers/comparisons_between_low_power_wireless_technologies.pdf

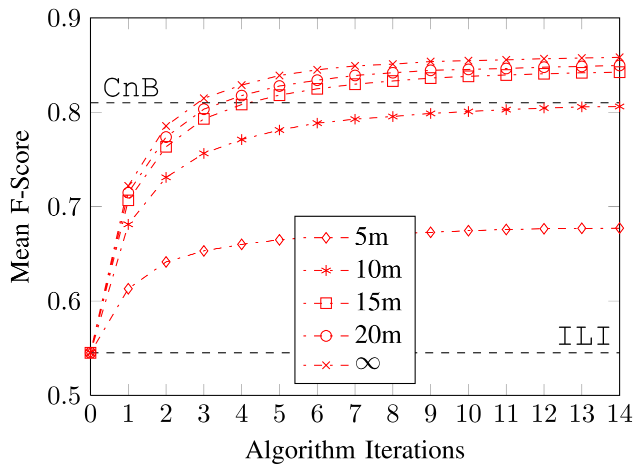

Fig. 7. Convergence Curves for DPI-SLBP for Varying Ranges ϕ

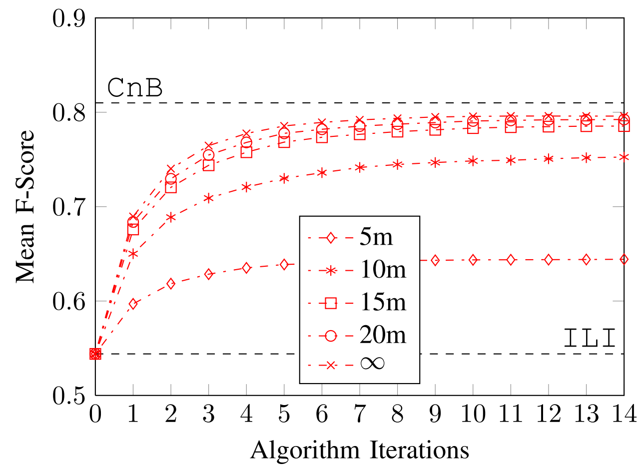

Fig. 8. Convergence Curves for DPI-HLBP for Varying Ranges ϕ

to note than the necessity to communicate with a server or centralized instance is removed for DPI-LBP algorithms.

The memory required to perform CnB is only the amount of memory required to store 1 window of sensory data. For DPI-SLBP, around 30 times more storage is required or almost 100 kB. DPI-HLBP only requires around 5 times more memory than CnB, representing a reduction of over 83% compared to DPI-SLBP due to the reduced size of the expectation look-up table compared to linear regression mappings.

VI. Discussion

The large reductions in resource consumption and low convergence time make DPI-HLBP an attractive approach. However for many applications there are some drawbacks. The effect of reducing simulated communication range was more pronounced than with DPI-SLBP. For both algorithms, conversion time increases as the group grows proportional to the communication range (see Tab. I), but it grows slower for DPI-SLBP then for DPI-HLBP. For applications where the surface area of the group is large proportional to the communication range of the group, e.g. groups or crowds in public areas, propagation rates for DPI-SLBP could be greatly affected. For such applications the indications are that DPI-SLBT is the best approach to take, although performance and scalability to large groups was not evaluated here. However through the use of LBP, each node is only dependent on neighboring nodes, meaning the approach is very scalable, where the limiting factor is the time required for information to propagate over the group. For small groups such as the one analyzed here, this time is negligible. However if the required response time of the system drops below the processing time required, the number of iterations possible becomes limited and may not suffice for convergence.

For both algorithms however, it is important that the communication range be proportional to the surface area of the group such that the vast majority of group members are connected to at least one other member by one link, and to all members by at least one multi-hop path so that belief may propagate. In the case of sport activities, this requirement is fulfilled by a range of around 12.5m-15m, or 50% of the surface area of the group.

For the presented experiment, it is conceivable that distributed majority voting techniques could achieve high GAR rates. However, in general for GAR, this will not be applicable. For problems with a higher degree of divergence ϵ, majority voting will inevitably degrade into noise by the definition of majority voting and ϵ. For this reason an analysis and comparison of such methods has been omitted here.

In the field of group activity recognition there are other aspects which are not addressed here [23]. Group members can come and go over time, leading to changing group sizes and changes in individual and group behavior characteristics. These aspects are outside the scope of this work and must still be researched, for GAR in general and for GAR using DPI-LBP. Integration of explicit roles into the approaches presented here, along with generalized models for each role and automatic role detection is a path for future research which we will follow, and which could potentially further address these open issues in the field of GAR.

VII. Conclusion

Group activities are emergent from the individual characteristics of group members, their roles in the group, and the group dynamic [2]. The group behavior therefore has properties which are different from the properties of the behavior of the individuals, as well as the “sum” of those individual properties [4]. Recognition of these activities is the process of inferring the properties of the whole, based on the properties of the individual behaviors.

We have shown that the emergent behavior of the group can be inferred using centralized inference methods where the distributed observations of all members are present with F-scores upwards of 95% possible. We use clustering to address the problem of inference without explicitly requiring role. We presented two methods of inferring emergent behavior in a distributed fashion, based local estimations (distributed probabilistic inference DPI) and exchange of belief estimates (loopy belief propagation LPB). The first (DPI-SLBP) propagates beliefs based on linear potentials over posteriors from subject to subject. The second (DPI-HLBP) propagates beliefs as expectation based on the most likely behavior of an individual.

TABLE II. RESOURCE CONSUMPTION ANALYSIS FOR ALL ALGORITHMS FOR 1 CLASSIFICATION AFTER 10 ITERATIONS

Approach	Memory Used (kB)	Comm. Per Classification (B)	Comm. Time (ms)	Comm. Energy (mJ)	Proc. time (ms)	Proc. energy (mJ)	Total Time (ms)	Total Energy (mJ)
CnB (3G)	3.6	3600	71.53	32.47	2.13	0	73.65	32.47
CCI (3G)	14.48	51.5	1.02	0.46	7.16	0	8.18	0.47
ILI	14.48	0	0	0	2.56	0	2.56	0
DPI-SLBP (BT)	99.35	5150	119.5	4.57	76.31	0.38	195.81	4.96
DPI-HLBP (BT)	16.6	800	18.56	0.71	17.7	0.09	36.26	0.8

DPI-SLBP and DPI-HLBP converged to relatively high rates of recognition, with F-scores of 0.84 and 0.80 respectively compared to a centralized inference of 0.81 for the same parameters. Reducing the the communication range to 50% of the diameter of the group only marginally affected the value which the distributed algorithms converged to, as long as the range did not create disjunct networks out of the single group. However it did affect convergence time, where the effect on DPI-HLBP was greater, increasing the number of iterations needed. For larger groups such as crowds where local communication range is small in proportion to the surface area of the group, DPI-SLBP is then preferable. However, DPI-HLBP greatly reduces local resource consumption compared to DPI-SLBP, making it attractive for small group applications. In total, the distributed approaches allow inference of emergent group behavior using only local observations and classification from the mobile devices themselves, without the need for a centralized instance or infrastructure. They also reduce local energy consumption of the nodes themselves, although for both algorithms the memory required locally increases, although still remaining under 100 kB. Response time also increases slightly for DPI-SLBP, although DPI-HLBP reduces response time against a cloud or server based centralized system.

REFERENCES

[1] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz, "The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics," *PLoS ONE*, vol. 5, no. 4, p. 7, 2010.

[2] D. Forsyth, *Group Dynamics*, 4th ed., ser. International student edition. Thomson/Wadsworth, 2006.

[3] D. Gordon, J.-H. Hanne, M. Berchtold, A. Shirehjini, and M. Beigl, "Towards collaborative group activity recognition using mobile devices," *Mobile Networks and Applications*, pp. 1–15, 2012.

[4] K. Lewin, *Field theory in social science: selected theoretical papers*, ser. Social science paperbacks, D. H. Cartwright, Ed. New York: Harper, 1951.

[5] J. Deguet, L. Magnin, and Y. Demazeau, "Elements about the emergence issue: A survey of emergence definitions," *ComPlexUs*, vol. 3, pp. 24–31, 2006.

[6] K. P. Murphy, *Machine learning: a probabilistic perspective*. Cambridge, MA: MIT Press, 2012.

[7] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive*, 2004, pp. 1–17.

[8] T. Gu, Z. Wu, L. Wang, X. Tao, and J. Lu, "Mining emerging patterns for recognizing activities of multiple users in pervasive computing," in *The Sixth Annual International Conference on Mobile and Ubiquitous Systems*, july 2009, pp. 1 –10.

[9] L. Wang, T. Gu, X. Tao, H. Chen, and J. Lu, "Recognizing multi-user activities using wearable sensors in a smart home," *Pervasive Mob. Comput.*, vol. 7, no. 3, pp. 287–298, Jun. 2011.

[10] G. Wittenburg, N. Dziengel, C. Wartenburger, and J. Schiller, "A system for distributed event detection in wireless sensor networks," in *IPSN '10: Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*. New York, NY: ACM, 2010.

[11] M. A. Paskin and C. E. Guestrin, "Robust probabilistic inference in distributed systems," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, ser. UAI '04. Arlington, Virginia: AUAI Press, 2004, pp. 436–445.

[12] M. Paskin, C. Guestrin, and J. McFadden, "A robust architecture for distributed inference in sensor networks," in *Proceedings of the 4th international symposium on Information processing in sensor networks*, ser. IPSN '05. Piscataway, NJ, USA: IEEE Press, 2005.

[13] M.-C. Chang, N. Krahnstoever, S. Lim, and T. Yu, "Group level activity recognition in crowded environments across multiple cameras," in *Advanced Video and Signal Based Surveillance, IEEE Conference on*. Los Alamitos, CA, USA: IEEE Computer Society, 2010.

[14] R. Li, R. Chellappa, and S. Zhou, "Learning multi-modal densities on discriminative temporal interaction manifold for group activity recognition," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, june 2009, pp. 2450 –2457.

[15] J. Y.-J. Hsu, C.-C. Lian, and W.-R. Jih, "Probabilistic models for concurrent chatting activity recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 4:1–4:20, January 2011.

[16] W. Dong, B. Lepri, A. Cappelletti, A. S. Pentland, F. Pianesi, and M. Zancanaro, "Using the influence model to recognize functional roles in meetings," in *Proceedings of the 9th international conference on Multimodal interfaces*, ser. ICMI '07. New York, NY, USA: ACM, 2007, pp. 271–278.

[17] M. Wirz, T. Franke, E. Mitleton-Kelly, D. Roggen, P. Lukowicz, and G. Tröster, "Coenosense: A framework for real-time detection and visualization of collective behaviors in human crowds by tracking mobile devices," in *Proceedings of European Conference on Complex Systems*. Springer, 2012.

[18] D. Roggen, M. Wirz, G. Tröster, and D. Helbing, "Recognition of Crowd Behavior from Mobile Sensors with Pattern Analysis and Graph Clustering Methods," *Networks*, vol. 00, no. 0, pp. 1–24, 2011.

[19] D. Grnbaum, S. Viscido, and J. Parrish, "Extracting interactive control algorithms from group dynamics of schooling fish," in *Cooperative Control*, ser. Lecture Notes in Control and Information Sciences, V. Kumar, N. Leonard, and A. Morse, Eds. Springer Berlin / Heidelberg, 2005, vol. 309, pp. 447–450.

[20] A. Huth and C. Wissel, "The simulation of fish schools in comparison with experimental data," *Ecological Modelling*, vol. 75-76, no. 0, pp. 135 – 146, 1994, state-of-the-Art in Ecological Modelling proceedings of ISEM's 8th International Conference.

[21] R. Lukeman, Y.-X. Li, and L. Edelstein-Keshet, "Inferring individual rules from collective behavior," *Proceedings of the National Academy of Sciences*, vol. 107, no. 28, pp. 12 576–12 580, 2010.

[22] K. D. Benne and P. Sheats, "Functional roles of group members," *Journal of Social Issues*, vol. 4, no. 2, pp. 41–49, 1948.

[23] D. Gordon, M. Scholz, Y. Ding, and M. Beigl, "Global peer-to-peer classification in mobile ad-hoc networks: a requirements analysis," in *Proceedings of the 7th conference on Modeling and using context (CONTEXT)*. Springer, 2011, pp. 108–114.

[24] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, "Energy consumption in mobile phones: a measurement study and implications for network applications," in *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, ser. IMC '09. New York, NY, USA: ACM, 2009, pp. 280–293.