2nd International Conference on Advanced Information Technologies (ICAIT), Nov. 1 - 2, 2018, Yangon, Myanmar

Towards the Discovery of Genuine Social Groups from Mobility Data

Htoo Htet Aung
edward@aidatech.io
AIDA Technologies

Nay Aung Lwin
lwin.nayaung@gmail.com
EIE Co. Ltd

Phyo Phyo San
phyophyo.san@suncorp.com.au
Suncorp Group

Abstract

Extraction of social groups from human mobility datasets has been regarded as convoy mining problem since a convoy (defined as a group of people which are spatially close to each other across time) is customarily assumed to represent a social group. However, social groups cannot be modelled trivially as a convoy as empirical evidence suggests convoy mining will report many 'false positives' and miss some 'false negatives'. We propose a two-step method to discover social groups from human mobility data in real-time. To the best of our knowledge, this paper is the first attempt to discover genuine social groups from mobility data. Experiments on the real-life dataset indicate that our two-step approach can accurately and efficiently discover genuine social groups.

Keywords- Real-time Data Mining, Mobility Data

1. Introduction

Human mobility data collected by systems like [1] can be analyzed either off-line or in real-time to obtain actionable insights, useful information and knowledge. For instance, Aung and Tan [2] has extracted frequently used routes from GPS traces of trucks and human users.

An important mobility analytics on pedestrian data is to extract social connection among the people tracked in the dataset as this information can lead to better operations of venues. For example – venue operators can notify a social group (a family) that someone from their group (an underage child) has gone missing soonest possible as a real-time social-group tracking system can monitor social-groups in public places.

Another useful application of social group information is in understanding of the population under analysis. For example, operators of museums and shopping malls will have a better understanding of the demographics of their visitors (families or singles etc.), which is essential to perform targeted advertisement or promotions. It also plays an important role in crowd control as noted in [3].

A naive approach to extract social groups from human mobility data is to model a social group as a convoy — traditionally defined as a group of users moving together — and employ one of the convoy mining techniques [4, 5, 6] to obtain social groups. However, consider the three independent pedestrians ('*a*', '*b*' and '*c*') walking down a narrow corridor in Fig. 1. Since they are moving together, the naive approach will detect them as a convoy and wrongfully report them as a social group even though they are not. We will term such instances 'false positive.' On the other hand, the family ('*p*', '*q*' and '*r*') is a social group yet the traditional convoy mining algorithms will not capture them as a convoy as the child '*p*' does not always move together with her parents, '*q*' and '*r*'. The naive approach will, therefore, will not report this family — a 'false negative.' The naive approach can capture only the couples ('*x*' and '*y*') as they move together.

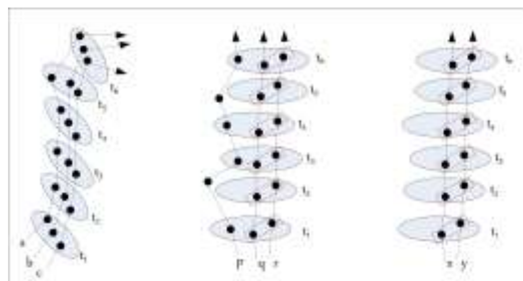


Figure. 1. An Example of Mobility Dataset

The false-positives and the false-negatives negatively impact the application in question. Take, for example, the three independent pedestrians ('*a*', '*b*' and '*c*') in Fig. 1. Alerting the group that their members are lost after each pedestrian take separate ways will be a waste. Likewise, only detecting the parents ('*q*' and '*r*') as a social group and not alerting when the child ('*p*') wonders away from the group for an extended period is not desirable.

To capture these 'false negatives' as social groups, Aung and Tan [7] introduced the notion of 'dynamic convoys' to capture the family in Fig. 1 correctly as a social group yet their approach still unable to weed out the false positives such as the group of '*a*', '*b*' and '*c*'. Zanlungo *et al.* [3] suggested that a genuine social group may be distinguishable from group of independent people moving together by using group features such as group formation and group velocity. For instance, from the fact that '*a*', '*b*' and '*c*' does not have a group formation (they don't have a member keeping in view of other members) and their high velocities, it is subtly hinted that they are not very likely to be a social group. In contrast, the couple and the parents in the family maintains a group formation ('*q*' and '*x*' maintain '*r*' and '*y*' respectively in their field of view by falling behind slightly as highlighted in

ISBN 978-99971-0-585-1 © 2018 ICAIT

rectangles). However, their findings are limited to simulating crowd dynamics and are not readily adaptable as models and algorithms for discovering social groups.

Indeed, users in a social group exhibit far more complex movement behaviors than a simple model like a convoy can characterize. Empirical evidence (See Sect. 6 for details) that naively modelling of social groups as convoys does not yield satisfactory results in accuracy. In addition, we postulate that a social group may exhibit different movement patterns in different environments, i.e. a group formation in a crowded corridor may be different from that of the same group in a park. Hence, a single rule-set or a model to describe movement patterns of social groups is impractical. Therefore, we propose to employ machine learning techniques to model the genuine social group by learning the movement behaviors of known social groups, after which the model can be used to effectively discover social groups in a similar dataset.

In this paper, we:

1) **report that a traditional convoy do not necessarily indicates a social group**. In other words, discovery of social groups using convoy mining methods will contain 'false positives' and miss 'false negatives'.

2) **proposed a two-step framework to extract genuine social group from human mobility data**. Our two-step framework consists of a mining step to get potential social groups and a classification step to determine which potential social group is a social group.

To the best of our knowledge, this is the first work that addresses the discovery of genuine social groups both accurately and efficiently. The first step in our two-step framework is based on the dynamic convoy mining algorithm [7] and can capture groups which do not always move together. The second step is a machine learning based classifier that identify social groups from the groups mined in the first step. It currently uses gradient boosting machine classifier we built using multiple weak classifiers. Experiment results show that our two-step framework is more efficient and accurate than the convoy approaches.

2. Related Works

In this section, we will discuss variants of convoy models and background on classification algorithms.

Flocks. Gudmundsson and Kreveld [4] defined a Flock pattern *f lock* (m, r, w) as a pattern formed by a set G of at least m objects staying in a *moving* circle of radius r for at least w consecutive time-stamps. They reported that the complexity to compute the all of flock patterns is NP-Hard. Vieira *et al*. [5] reported polynomial-time algorithms to find Flock instances of fixed durations.

Convoys. Flock model has a short-coming —a circular area with maximum radius r cannot fit more than a certain number of members — termed as lossy-flock problem by Jeung *et al.* [6]. Using the definition of density-connectivity (two density-connected objects have a chain of dense neighbors that connect them) from [9], they defined a Convoy as a group of at least m objects being density-connected with each other throughout w consecutive time-points, where m, w, and the two DBSCAN parameters ε and *min-pts* = m are provided by the user. Since a convoy can occupy an arbitrary shaped region in its lifetime, there is no lossy-flock problem.

Dynamic Convoys. Aung and Tan [7] proposed Dynamic Convoy to mitigate shortcomings of convoy models requiring to have all its members spatially close from convoy's formation to its disintegration. For given parameters: m, k and w, a set of moving objects D forms a dynamic convoy during a period $P = [t_{\text{start}}(D), t_{\text{end}}(D)]$ if it contains (i) at least m persistent-members, all of which are density-connected in each time-stamp t in P and (ii) zero or more dynamic members, each of which must be density- connected with the persistent-members at least k times for any w sliding window in P. The definition of Dynamic Convoy allows dynamic members to move away from the main body and, thus, reduces "false-negatives".

Background on Classification. Traditional tools and notions used in query processing become impractical and inadequate when the underlying semantics are complex to capture in a simple rule/query-model. This is more apparent in social group discovery as the same social group may behave and move in different ways in different applications' environments. Machine Learning techniques can overcome this issue by building the model from data.

Supervised machine learning is a type of machine learning algorithm that learns from the training dataset, to classifies unseen instances in the test dataset. It works in two phases – training phase, where the machine learns from the training dataset to produce a model and the prediction phase, where the machine uses the model to classify the test data. An instances of supervised machine learning technique is the decision tress based gradient boosted machine (GBM).

GBM Model. McCaffrey *et al.* [9] proposed GBM model, which is a piecewise constant model for prediction of dichotomous outcomes. Initially, it starts with a single simple regression tree while other trees are constructed and added at each new iteration, in which the new tress is determined to provide the best fit to the residuals of the model from the previous iteration and provides the greatest increase to the log likelihood for the data. Exploiting the connection between boosting and optimization, Friedman [10] introduced the gradient boosted machines. GBM can build a strong classifier by combining weak classifiers.

3. Problem Definition

Mobility Dataset — For a given period of time $T = \{t_1,t_2,...,t_\tau\}$ and a set of users $U = \{u_1,u_2,...,u_n\}$, a set $\mathbb{D}$ of records of the form $\langle u,t,x,y\rangle$, where $u \in U$, $t \in T$ and $(x,y) \in \mathbb{R}^2$, is a Mobility Dataset. In a Mobility Dataset, each record $\langle u,t,x,y\rangle$ represents user u is at location (x,y) as sampled on time-stamp t. Without loss of generality, we assume that we can access the mobility dataset in ascending order of time-stamps. This reflect real-world application scenarios with streaming data. Fig. 1 and Fig. 2 depict examples of human mobility datasets chronicling movement records of $U_{tr} = \{a,b,c,p,q,r,x,y\}$ and $U_{ts} = \{d,e,f,i,j,s,t,u\}$ respectively. The ovals indicate proximity.

Social Group Information — Given a mobility dataset $\mathbb{D}$, its social group information $\mathbb{G}$ is a set $\{G : G \subset U$ and G is a social group $\}$. For example, consider the mobility dataset in Fig. 2, its social group information contains the family and the couple, $G_{ts} = \{\{s,t,u\},\{i,j\}\}$. Since social associations do not necessarily translate into a set of identifiable movement patterns and vice versa, finding social groups usually is labor-intensive and requires both mobility datasets and other sources of data.

Thus, the discovery of social group – to find all social group information from a given mobility dataset – has become a challenge as the mobility datasets explode.

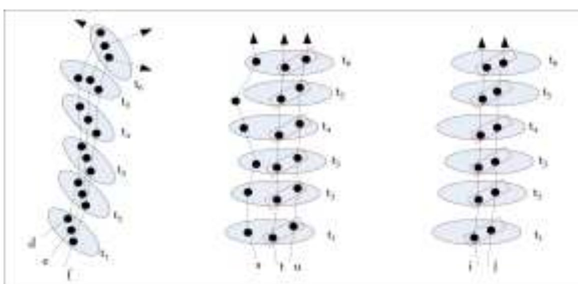


Figure. 2. Mobility Dataset for Test

4. Attempts to Find Social Groups

Convoy — For given parameters: m, k and $w = w_c$ ($m > 1$, $1 \leq k \leq w$), a set of moving objects C forms a convoy in time period P of length w_c if it:

- Contains at least m persistent-members, all of which are density-connected in all t in P and
- Contains dynamic members, each of which must be density-connected with the persistent-members at least k times in P

This definition of convoy is that of a dynamic convoy defined in [8] except that it lasts exactly a period of length w_c. Consider the family $\{p, q, r\}$ in Fig. 1. For $m = 2$, $w = 6$ and $k = 4$, it forms a dynamic convoy as q and r form the main body of the convoy and p is registered as a dynamic member. Using this definition, we can build a program to mine all convoys in a given mobility dataset.

Following the assumption that a social group will be always together, we first attempt to approximate a social group with convoys by regarding each convoy is a social group. Following our running example in Fig. 2, there are three convoys formed by $\{d, e, f\}$, $\{s, t, u\}$ and $\{i, j\}$. Therefore, this approximation method reports three social groups including the group of three independent pedestrians $\{d, e, f\}$. We term such an instance of wrongfully reporting a group with no social connection a 'false positive.' Likewise, since not all social groups move like a convoy, this attempt can also miss to report a genuine social group, which we term as a 'false negative.'

Limitations. In our experiments (see Sect. 6.2), where we compare a human-labelled social groups with convoy results, we discovered that this approximation method yields very few false negatives but many false positives.

Since we observed that the convoy model derived from fixed movement patterns cannot represent a social group, we consider modelling social groups using a machine learning technique. To do this, we compute group features and manually label each group. From the computed features and labels (the training data), machine learning algorithm learns and produces a model.

If we use the mobility dataset shown in Fig. 1 as training data, we need to compute features and mark label for all $G \subset \{a, b, c, p, q, r, x, y\}$. Table 1 shows a few groups along with three features and their labels. The three features are x_1 = whether the group is elongated towards the group direction, x_2 = whether the members keep other members in view. The features computed for the family $\{p,q,r\}$ is $x_1 = N$ and $x_2 = Y$ (in the second row).

Table 1. Examples of Features and Labels

G Group	x_1 (elongated)	x_2 (keep in view)	y
$\{a, b, c\}$	Y	N	N
$\{p, q, r\}$	N	Y	Y
$\{x, y\}$	N	Y	Y

Learning from the training data, the machine will produce a model, using which we can identify social groups from unseen/future test data. In this example, the output model will be "Social groups are groups with $x_2 = Y$ and $x_1 = N$". Applying this rule-set (model) to identify social groups in the test dataset shown in Fig. 2 results in $\{s,t,u\}$ and $\{i, j\}$ correctly outputted as social groups..

Limitations. Although machine learning improves accuracy, it is not practical because we need to calculate features for all groups in the given dataset. The number of all the possible groups are exponential to the number of users. For the test dataset (Fig. 2), we need to calculate features for 247 groups (2^8 = 256 subsets of 8 users minus 9 subsets with size less than 2). Since real-life applications have thousands of users, this method is not feasible.

5. Proposed Framework

Since approximating social groups with convoys is efficient and identifying social groups using models built by machine learning is accurate, we combine the merit of these two methods into a framework to discover social groups. Since convoy approximation usually contain fewer false negatives than false positives, we use it as a filtering step to reduce the number of groups for feature computation to a manageable size. Then, in the next step, we use a machine-learned model to identify social groups.

Box 1. outlines our proposed framework. In training phase, convoys are mined in training dataset D_{tr} (line 1).. Then, features extracted from convoys and labelled groups are used to build a classifier *clf* using GBM (line 2). In test phase, for each sliding time-window, two steps are performed to get the social groups. The first step is mining potential social groups using a convoy mining algorithm (line 4). The second step is calculating features for the potential groups (line 5) and classifying if each potential social group is a genuine social group using the classifier *clf* (line 6). The test phase is designed to work incrementally using a sliding time-window and, hence, our framework can work on a streaming dataset.

Input: Training data, D_{tr} and G_{tr}, Convoy parameters, (*e*, *m*, *k*, *w*) and GBM learning rate *delta* and Testing data D_{ts}
Output: a set of social groups
1. Convoys in Training Data C_{tr} ← *Convoy* (D_{tr}, *e*, *m*, *k*, *w*)
2. Features *X* ← *Features* (C_{tr}), *clf* ←GBM (*delta*, *X*, G_{tr})
3. **for each** length *w* sliding time-window in D_{ts} **do**
4. Potential social groups G′ ← *Convoy* (D_{ts}, *e*, *m*, *k*, *w*)
5. Features *X'* ← *Features* (G′))
6. **Output** social groups *clf* (*X'*)

Box 1. Framework to Discover Social Groups

We are going to illustrate how our framework works using the datasets in Fig. 1 and Fig. 2 as training and test datasets. In the training phase, convoys formed by {*a*,*b*,*c*},{*p*,*q*,*r*} and {*x*,*y*} will be mined and features will be calculated for them. Based on the features and labelled data G_{tr} = {{*p*,*q*,*r*},{*x*,*y*}}, the classifier *clf* is built. Once the classifier is trained, the test phase begins. In the first step of test phase, potential social groups G′ is mined to find convoys formed by {*d*,*e*,*f*}, {*s*,*t*,*u*} and {*i*,*j*}. In the second step, features of these three groups are calculated. Notice that in contrast to applying a machine learning classifier directly, where features of 247 groups needed to be calculated, our framework only calculates features of 3 groups. These features will be used to identify {*s*,*t*,*u*} and {*i*, *j*} as social groups while {*d*,*e*,*f*} will not be reported.

Mining Potential Social Groups (MPSG). The first step in our proposed framework is to mine convoys to reduce the number of groups, for which feature computation to be performed. We adapt the S^3 proposed in [7] to discover all the convoys as the S^3 algorithm is capable to mine convoys in incremental manner. An alternative candidate is the online flock mining algorithm proposed in [6] but this algorithm has lossy-flock problem and, thus, will introduce false-negatives.

S^3 algorithm requires four parameters, ε, *m*, *k* and w_c to control the model or the query of the convoy mining process. All these parameters are intuitive and easy to set.

Identifying Social Groups (ISG). The second step in our proposed framework is to identify genuine social groups from the potential social groups. We choose Gradient Boosting Machine (GBM) for its simple regularization strategy and its ability to build good classifiers from weak classifiers. The regularization of a GBM learning process can be tuned through the shrinkage parameter. Other parameters include number of estimators and minimum sample split to split an internal node.

Features Selection. We calculated 57 features in total reflecting group movement (such as mean group velocity) and relative position/movement angle of group members. Most relative features are based on the movement angle and frame of the group, which are defined as follows.

Let $\mathcal{L} = \{\alpha_{1,1}, \alpha_{1,2}, ..., \alpha_{1,m}, ..., \alpha_{n,1}, \alpha_{n,2}, ..., \alpha_{n,m}\}$ represent the movement angles of a group of *n* users over *m* time-stamps. The movement angle of a group is *mean_ma* = *arctan*(*mean*(*sin*(α)),*mean*(*cos*(α))).

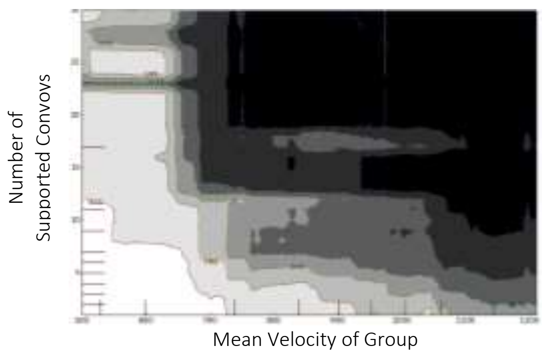
Let θ be the movement angle of a group of users *G* for a period *P*, the group-frame of *G* at time-stamp $t \in P$ is the bounding rectangle covering all the locations of the group members at time *t* and having sides parallel and orthogonal to the x' axis of the coordinate system rotated by θ from the default (*x*,*y*) coordinate system.

From movement angle and group frames of a user group, we compute features. Table 2 tabulates the three most important features we obtained from the model.

Fig. 3 shows the likelihood of a group being a social group given the feature values (darker means higher likelihood). We observe that faster movement and more occurrence of convoys formed by the group also indicate high chance of being a social group. Likewise, groups elongated in the direction orthogonal to the group movement angle is more likely to be a social group Since these observations agree with our observation in the dataset, we deduce that the features we select can differentiate a genuine social group from a mere convoy.

Table 2. Three Most Important Features

No.	Feature	Description
1	Mean_Velocity	Mean velocity of the group
2	Num_Convoys	Number of convoys supporting this group
3	Group_Length	Length of the group-frame parallel to group's mean velocity



Figue 3. Partial Dependence Plot for Features, Mean_Velocity and Num_Supported_Convoys

6. Experiments

We conducted experiments to assess the performance (run-time and accuracy) of our proposed framework against other methods based on Flock and Convoy models. Since our framework can discover social groups in real-time, we tried to compare it against Flock and Convoy models specifically because they have incremental algorithms to mine them in real-time. Flock represents all traditional convoy models while Convoy represents the dynamic convoy model (defined in Sect 4).

We used five sets of human mobility data [11] collected in five different days. Each day of datasets consists of four hours (10:00-11:00, 12:00-13:00, 15:00-16:00, 19:00-20:00) of movement data along with social groups labelled by a human coder [3]. We used two datasets (0424 and 0508) for performance comparison while the rest were used to train the GBM model for our framework. Details of the datasets are given in Tab. 3.

We define the step size of the sliding window for all three methods as 5 second. We chose parameters for Flock and Convoy methods to obtain the best accuracy. For convoy, we set the parameters m = 2, w_c = 6 and k = 5 while for flock, we set similarly as m = 2 and w_c = 6 — a social group should stay together for 6 time-stamps, which are 5 second apart, i.e. 30 second. Clustering parameter for convoys is given as e =1.5m while the disc size for flocks is r = 0.75m. The difference in clustering/disc-size parameters is fair as they reported similar results.

Table 3. A Summary of the Datasets

Name	Num. Objects	Covers	Num. Records	Num. Social Groups
0109	3,774	09/01/13	210, 079	317
0217	7,465	17/02/13	449, 201	1, 221
0324	7,472	24/03/13	449, 804	1, 242
0424	2,750	24/04/13	140, 867	326
0508	2,858	08/05/13	156, 913	384

For the convoy sub-routine in MPSG step in our proposed framework, we chose to set w_c = 3 and k = 2 respectively. This relaxes the convoy selection criteria and introduces a very large number of false positives while reducing a few false negatives. Although this trade-off reduces accuracy of the MPSG step drastically, it increases the overall accuracy of our framework because MPSG step will report less false negatives while the ISG step will discard the false positives MPSG step introduces.

We performed 5-fold cross validation in a small parameter space to find satisfactory parameters for the GBM learner in the initialization phase of our framework. We found that learning-rate g = 0.05 and number of estimators = 150 gives good accuracy performance.

In measuring accuracy of each method, we counted a group reported by a method as a true positive if and only if there is an exact match with a social group in the ground-truth data, i.e. if {p, q, r} is a social group in ground truth data, reporting {p, q, g} or reporting {p, q, r, s} will not be counted as true positive. Instead each of these instances will be counted as a false positive.

First, we measured the run-time performance of each method on two test datasets, 0424 and 0508. Fig. 4 shows the run-time of each method and dataset pair. Approximating with Convoys take the least amount of time while approximating with Flocks took the longest to report the results because flocks of single group can be detected multiple times in a time-window, but a group can form a single convoy for each time-window (by its definition). Our framework takes longer than approximating with convoys since, for each potential social group outputted from MPSG step, ISG step needs to compute its features and identify if it is a social group. Feature computation time dominates the ISG step.

Next, we measured the accuracy performance of the three methods. Analysis of the results for the datasets, 0424 and 0508, we learned that that the proposed framework gives higher accuracy (found 230/326 and 282/384 social groups for 0424 and 0508 datasets) than Convoys (found 221 and 267) and Flocks (found 67 and 78) as the approximating methods report many false positives (126 and 161 false positives were reported by Convoy while our framework reports only 71 and 91). Flock method performs the worst as also misses many of the social groups (259 and 306 false negatives from Flock method as compared to our method's 96 and 179 false negatives).

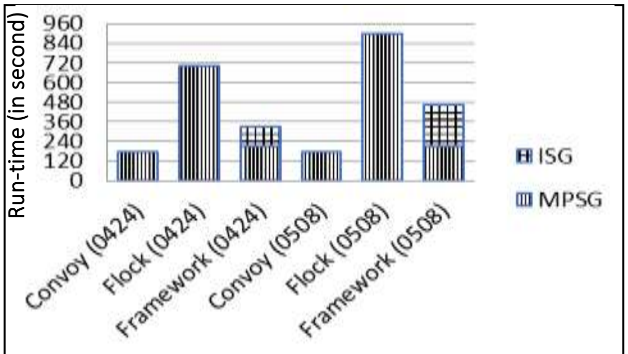


Figure 4. Run-time of the methods on each data

Table 4 compares the accuracy of the three methods in comparison in term of commonly used metrics, Precision, Recall, AUC (Area Under the Curve of ROC) and F1. Our proposed Framework yields far better AUC than Convoy and Flock, which performs worse than random chance due to the false positives. F1 values reflect a similar trend.

Table 4. Summary of the Datasets

Dataset	Method	Precision	Recall	AUC	F1
0424	Convoy	0.64	0.67	0.33	0.65
	Flock	0.32	0.21	0.10	0.25
	Framework	0.76	0.71	0.73	0.74
0508	Convoy	0.62	0.70	0.35	0.66
	Flock	0.29	0.20	0.10	0.24
	Framework	0.76	0.61	0.68	0.76

From the experiments, we concluded that approximating social groups using traditional convoy methods (Flock) cannot produce accurate results. Both Flock and Convoy methods cannot produce results better than random chances (ROC AUC of 0.1x and 0.3x respectively). Our proposed framework can produce more accurate results (ROC AUC of 0.7x) than simple convoy mining methods at a reasonable run-time, which even is faster than Flock by several minutes.

7. Conclusion

In this paper, we studied how to extract social groups from human mobility datasets. Social groups do not necessarily translate to a convoy. We reported that approximating social groups using traditional convoys does not yield a satisfactory result in real-life datasets. We proposed a two-step framework to discover genuine social groups from streaming human mobility data in real-time. The first step of our framework narrows down the search space while the second step employs machine learning techniques to model and correctly output social groups accurately. In experiments using real-life datasets, our proposed framework outperforms other approaches.

8. References

[1] Dražen Brščić , T. Kanda, T. Ikeda and T. Miyashita, "Person Tracking in Large Public Spaces Using 3-D Range Sensors," in IEEE Transactions on Human-Machine Systems, vol. 43, no. 6, pp. 522-534, Nov. 2013.

[2] H.H. Aung, L. Guo, K-L. Tan. "Mining Sub-trajectory Cliques to Find Frequent Routes." In Proc. Advances in Spatial and Temporal Databases. SSTD 2013. Lecture Notes in Computer Science, vol 8098. Pp. 92-109.

[3] F. Zanlungo, T. Ikeda and T. Kanda. "Potential for the dynamics of pedestrians in a socially interacting group." Physical review. E, Statistical, nonlinear, and soft matter physics vol 89 1 (2014). pp 012811.

[4] J. Gudmundsson and M. Kreveld. "Computing longest duration flocks in trajectory data." In Proc. the 14th annual ACM international symposium on Advances in geographic information systems (GIS '06). pp 35-42.

[5] M. Vieira, P. Bakalov, and V. Tsotras. "On-line discovery of flock patterns in spatio-temporal data." In Proc. the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '09). pp 286-295.

[6] H. Jeung, H. T. Shen and X. Zhou, "Convoy Queries in Spatio-Temporal Databases," In Proc. IEEE 24th International Conference on Data Engineering, 2008, pp. 1457-1459.

[7] H.H. Aung and KL. Tan. "Discovery of Evolving Convoys." In: Proc. Scientific and Statistical Database Management. SSDBM 2010. pp. 196–213.

[8] M. Ester, H-P. Kriegel, J. Sander, and X. Xu. "A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise." In Proc. Second International Conference on Knowledge Discovery and Data Mining (KDD'96). pp 226-231.

[9] D. Mccaffrey, G. Ridgeway and A. Morral. "Propensity Score Estimation With Boosted Regression for Evaluating Causal Effects in Observational Studies". Psychological methods. 9. pp 403-25.

[10] Jerome H. Friedman, "Stochastic gradient boosting", Computational Statistics and Data Analysis, Vol 38. Issue 4, 2002, pp 367-378.

[11] D. Brscic, T. Kanda, T. Ikeda, T. Miyashita, "Person position and body direction tracking in large public spaces using 3D range sensors", IEEE Trans. on Human-Machine Systems, Vol. 43, No. 6, pp. 522-534, 2013.