IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-5, Issue-11 E-ISSN: 2347-2693

Social link prediction using category based location history in trajectory data

Revathy S B[1*], Remya R[2]

[1*]Dept Of CSE&IT, College of Engineering Perumon, APJ Abdul Kalam Technological University, Kerala ,India
[2]Dept Of CSE&IT, College of Engineering Perumon, APJ Abdul Kalam Technological University, Kerala ,India

**Corresponding Author: revathysinoj244@gmail.com*

Available online at: www.ijcseonline.org

Received: 10/Oct/2017, Revised: 27/Oct/2017, Accepted: 19/Nov/2017, Published: 30/Nov/2017

Abstract— With the rising popularity of location-based services, trajectory data mining became an important research topic. There exists many data mining algorithms for systematic processing, managing and mining of trajectory data. Trajectory data mining has many applications such as location recommandations, social link prediction, movement behaviour analysis etc. Here proposes a contextual trajectory analysis model which provides a flexible way to characterize the complex moving nature of humans. It embed multiple contextual information for efficiently modeling data. It includes user-level, trajectory-level, location-level, and temporal-level contexts. It can be used to predict the future location of a user based on the previous travelling pattern. Social link prediction aims to find out whether there exists reciprocal link between two users. Here also propose a method for social link prediction from trajectory data by analyzing the nearest neighbour. This method considers the tf-idf metrics as the baseline.

Keywords—Trajectory,contextual information,social link prediction

I. INTRODUCTION

Rapid growth on wireless technologies and services has made it possible to continuously track movements of objects. The location- based social networks always provides an opportunity to enormously analyze behavior of humans. People have different patterns of movement and activity preference according to their lifestyles. Not only limited to trajectory data, social link prediction has attracted much attention in various social networking websites. The check-in data provides the information about the nature of humans. For a single check-in record of a user, there exists several important factors to be considered including user's regional and categorical preference.There also exists different algorithms for finding location recommendation, social link prediction, movement behavior analysis etc. Trajectory data itself is a kind of sequential data, and for that surrounding contexts are compulsory to consider for modeling of trajectory data . Another issue with previous studies is that the addition of more new contexts will degrade the essential features and also increase model complexity.

Trajectory data of a user is a sequence of check-ins. The data contains user id, location, time and category information. A general model which consider multiple contexts called multi context trajectory embedding model provides a adaptable way to distinguish the traces of moving objects . The model includes four contexts namely user, trajectory, location and temporal contexts. All the contextual information needed for analysis are represented in an embedding manner. A neural network approach called distributed representation method is used for analyzing the trajectory pattern. It forecast the future location of a user based on their historical trajectory pattern. It also forecast the social link among users by checking their feature vectors. The key point of this approach becomes how to derive effective feature representation for link prediction. The need for effective feature representation make the model complex. Location prediction of a user can be carried out by using the contextual information .Here we propose a framework for predicting social link among users based on the tf-idf metrics. Tf-idf is used as the weighting factor. It takes the location entropy and location frequency for each location visited by a particular user. It find out the nearest neighbour of a particular user among a set of users by calculating the Euclidean distance .It reduces the computation cost to a great extent. Jaccard index is used as a metrics for social strength estimation.

The remaining part of the paper is organized as follows, section II contains the related work, section III contains the methodology, section IV contains results and discussion, and finally section V contains conclusion.

II. RELATED WORK

A distributed representation method[1] was proposed for inferring the social connections among users by comparing the feature vectors. It first extracts features from a single user and construct feature vector for the user pair. Since different users have a varying number of trajectories, it adopts a technique known as max pooling to derive a single vector with fixed length. It then perform the Hadamard product between user vectors to derive the final feature vector for classification. The need for effective feature representation make the model complex.

An entropy based method [2] for deducing the social inter connection among users was proposed which analyzes location entropy and its frequency. The model is applicable to a variety of fields of trajectory data mining. Weighted frequency and location diversity are the two factors considered in this model. The model quantify social strength between two users from their co-occurrence vectors. In cases, where only limited amount of location information is available, the characteristics of each location is also taken into consideration. EBM also alleviated the problem of data sparseness by incorporating the location characteristics into the model when estimating the strength of social connections.

A graphical method, referred to as hierarchical-graph-based similarity measurement (HGSM)[3] was proposed which helps the geographical systems to continuously monitor each people location history and exactly measure the similarity among the users. This hierarchical graph based model takes into account both the sequential property of people's movement behaviour and the hierarchy property of geographic spaces. The main steps involved in the model are location history representation and user similarity exploration. Similarity measurement is carried out by sequence matching algorithm.

A periodic mobility model (PMM)[4] was build on the basis that the majority of human movement is periodic between a small set of locations. The PMM model includes both spatial and temporal content of movement. The model combines the periodic day-to-day movement patterns with the social movement effects coming from the friendship network. The model reliably captures and predicts human mobility patterns.

Another method for finding similar users based on semantic location history [5] was proposed which takes into account more semantic meanings involved in user interests rather than simple low-level positions in geographical spaces. This semantic location history approach can clearly estimate the similarity between two individuals without making overlaps in the geographic spaces. For example, two people living in different cities. Later on, this semantic similarity score can be used as a distance function by some existing algorithms for clustering such as K-means to cluster users into groups. POI database is also needed to translate a users semantic location history from geographic spaces into the semantic spaces for pattern matching.

III. METHODOLOGY

A.Overview

When a user *u* checks in a location *l* with a category label *c* at the timestamp *s*, the check-in record can be modeled as a quadruple <u,l,c,s>. An example check-in record can be (UID2, Lakeshore@LH Point, Restaurant, 2014-01-13/1:00pm),which tells that a user with the ID of 2 has visited the restaurant "Lakeshore @LH Point" at 1:00 pm on January 13, 2014. Given a user *u*, a trajectory *t* is a sequence of chronologically ordered check-in records related to u: $<u,l_1,c_1,s_1>,\ldots,<u,l_i,c_i,s_i>,\ldots<u,l_N,c_N,s_N>$ where N is the sequence length and $s_i < s_{i+1}$ for i <= N -1. Given a trajectory sequence data , the objective function is to maximize the log probability function for every location given its corresponding four contextual information. Here we take into account that the generation of a check-in location of a particular user is always associated with their four contextual information. The contexts involved are user level, trajectory level, location level and temporal level contexts which are represented in an embedded manner.

B.Social link prediction

Besides the direct applications to improve geo-oriented methods, human trajectory similarity has been shown to be an important evidence for predicting social links. Especially, the geo-social network sites provide us with both trajectory data and social link information. Given the trajectory data from two users namely u and v, link prediction aims to predict whether there exists a reciprocal link between u and v. Link prediction rely on the explicit co-occurrence of check-in locations. Social strength is a measure that tells how socially close two users are. Given two users *u* and *v*, their social strength is characterized as a real number *r(u,v)* falling in the interval [0, 1], where a larger value indicates stronger strength between two users.

- Intensity and duration: It quantify the number of observations, number of co-location observations.
- Location diversity: It includes location frequency and location entropy[6].
- Specificity: It measures whether two persons meet at locations where less frequently visited by the public. Tf-idf score penalizes popular places that many people frequently visit.

Tf-idf is term frequency- inverse document frequency, which is the product of term frequency and inverse document frequency. Tf-idf is calculated for each location. It is used as the weighting factor.

Algorithm:

- Create user files

- Calculate tf-idf matrix. It is the product of term frequency and inverse document frequency.
- Penalizes the popular places by using the tf-idf score.
- Calculate the Euclidean distance

$$dist = \sqrt{\sum_{k=1}^{n}(p_k - q_k)^2}$$

- Find the nearest neighbour having minimum distance.
- From the social network graph, calculate the Jaccard's index. Jaccard's index calculation returns the user with the high social strength.

- Compare the nearest neighbour from the proposed method with the jaccard index value[7].

In public places, there may have greater tf-idf score. The people may not have real friend links. So, here we penalizes the weights. Location entropy associated with such places is greater compared with others. There are also other social strength estimation can be used such as adamic/adar similarity, katz score etc. If the estimated value is near to 1, it means that there exists high friendship among the users. It can be used as the ground truth for our approach.

III. Results and Discussion

The public Geo-social networking dataset namely weeplaces is used in this work.The datasets contain check-in records in the form of (User ID, Location ID, Location Category, Timestamp, City). It also contains social connection links among users. The trajectory data of the users are first preprocessed to eliminate unwanted details. The different levels are then processed. To measure the effectiveness on social link prediction, we adopt the commonly used metrics called recall. Recall is the ratio of the hit cases to the total number of cases.

Jaccard index is used for checking the accuracy of the proposed method. The predicted values are compared with the index value. The dataset contains the connection links among users and jaccard index is applied. Jaccard index outputs the social strength among users from the social networking graph. The value will be between 0 and 1. The value nearer to 1 means high social strength. If same user names are outputted from the proposed method and jaccard index estimation , it is said to be a hit. When the number of hits increases, the recall value grows and hence the prediction accuracy increases.

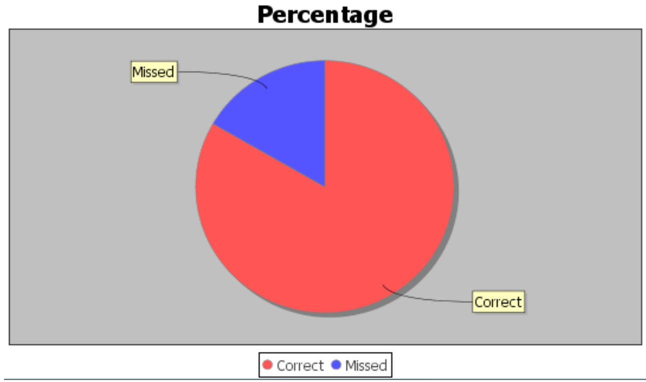

Fig 4.1 Recall graph

Fig 4.1 shows the recall graph. The graph sows the accuracy of our proposed approach. It shows high degree of recall value. Most of the cases are hits by using this approach.

V Conclusion

With the high growth of location-based services provides valuable data resources to continuously trace moving characteristics of people.The trajectory data is a very useful tool to analyze the moving pattern and behaviour of any situation. There exists different applications of trajectory data mining which includes pattern detection, location recommendation, link prediction etc. Social link prediction application aims to predict whether there exists social connections between users. The proposed system predicts the links among users based on their trajectory information history. It uses tf-idf as the baseline. The predicted results are compared with the jaccard index value of social strength estimation. The proposed link prediction method outperforms in terms of accuracy. Currently only a small tf-idf metrics is used. In future, deep learning methods can be explored for feature learning.

References

[1]Ningnan Zhou, Wayne Xin Zhao, Xiao Zhang, Ji-RongWen, Shan Wang, "*A General Multi-Context Embedding Model for Mining Human Trajectory Data*" *IEEE Trans on Knowledge and Data Engineering* , Vol. 28, No. 8, pp.1945-1958, 2016.

[2] H. Pham, C. Shahabi, and Y. Liu, "*Ebm: An entropy-based model to infer social strength from spatiotemporal data*," in Proc. ACM SIGMOD Int. Conf. Manage. Data, pp. 265–276, 2013.

[3] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma, "*Mining user similarity based on location history*," in Proc. Annu. ACM Int. Symp. Adv. Geographic Inf. Syst., pp. 34:1–34:10, 2008.

[4] X. Xiao, Y. Zheng, Q. Luo, and X. Xie, "*Finding similar users using category-based location history*," in Proc. Annu. ACM Int. Symp. Adv. Geographic Inf. Syst., pp. 442–445, 2010.

[5]E. Cho, S. A. Myers, and J. Leskovec ," *Friendship and mobility: User movement in location-based social networks*", in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 1082–1090, 2011.

[6] B. Perozzi, R. A.-Rfou, and S. Skiena, "*Deepwalk: Online learning of social representations*," in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 701–710, 2014.

[7] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi, "*Human mobility, social ties, and link prediction*," in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 1100–1108, 2011.
[8] W. Mathew, R. Raposo, and B. Martins, " *Predicting future locations with hidden markov models*," in Proc. Int. Joint Conf. Pervasive Ubiquitous Comput., pp. 911–918, 2012.
[9] G.Sivaiah, P.K.Rao,"*A comprehensive survey on providing efficient directions using GPS and driver's ability*", International Journal of Computer Sciences and Engineering, Volume-2, pp. 79-82, 2014.

