

# Context Co-occurrence Based Relationship Prediction in Spatiotemporal Data

Caixu Xu, Jianfeng Yan\*, Lu Yang, Guanggen Xu and Hongbin Shi  
Soochow University, Suzhou, China  
\*Corresponding author

**Abstract**—Recently, users’ relationship prediction in spatiotemporal data has attracted widespread attentions. Previous studies have focused on either co-occurrence or context in spatial aspect, where the context in time aspect is seldom considered. In this paper, considering co-occurrence, context, and mobility periodicity together, we propose a novel social relationship prediction approach named Multi-View Context Co-occurrence (MVCC) for this problem. The combination of context and co-occurrence is not simply merged together, specifically, we propose a method that artfully transfers user-pair relationship in spatiotemporal data to word-pair relationship in natural language processing domain. In our approach, the context sequences capturing spatiotemporal semantics information from multi-views are constructed and the multi-view context co-occurrence feature with different degree representation is extracted from them. These multi-view context co-occurrence features are used to train multiple classifiers. The outputs representing different degree spatiotemporal information are weighted and fused as the final relationship strength. The results show feasibility of our approach compared to the methods such as EBM and SCI.

**Keywords**—relationship prediction; spatiotemporal data; multi-view context sequence; context co-occurrence

## I. INTRODUCTION

Spatiotemporal data is so popular that they have attracted millions of users in the past decade. Because of the wide use of location-sensing mobile devices, users can share spatiotemporal information with their friends easily. People seem more willing to share their check-in information, which companies can consider to customize services. There are numerous relationship hidden in these geographic information, which requires a suitable method to mine. This social relationship information offers a groundbreaking opportunity to better understand and exploit users’ requirements, such as friend recommendations or targeted advertisements for Internet companies [1].

Intuitively, users who have higher relationship strength have a greater chance to appear together at the same occasions compared to strangers because of some same occupations or shared interests, such as colleagues meeting in workdays or friends spending time together at a coffee shop. Based on this intuition, the methods predicting relationship had been widely studied [2-5] through co-occurrence feature that means appear together. [6] proposed a model that considered current context of the user to provide relationship prediction. Different from methods based on either context or co-occurrence, we propose an innovative method to extract feature capturing context and co-occurrence information. None of these above methods [2-6] include temporal context. Our method better combines context

and co-occurrence together called context co-occurrence, different from literatures [2-6], and meanwhile the people’s periodic mobility also is considered.

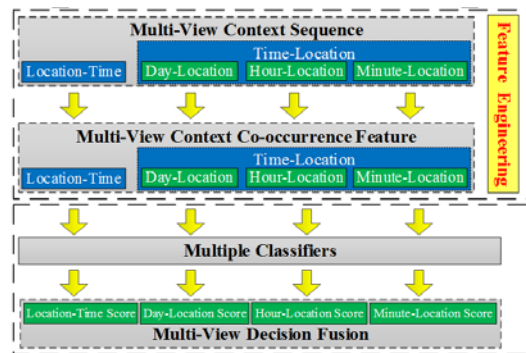


FIGURE 1. THE OVERVIEW OF MVCC APPROACH

In this paper, we propose Multi-View Context Co-occurrence (MVCC) approach. As shown in Fig.1, MVCC consists of four phases: 1) Multi-View Context Sequence Generation; 2) Multi-View Context Co-occurrence Feature Extraction; 3) Multiple Classifiers Learning; and 4) Multi-View Decision Fusion. The combination of the first and second phase is also called Feature Engineering. First, we analyze original data from multi-view and artfully construct multi-view context sequence. Second, a transferring method is proposed to artfully extract context co-occurrence feature based on context sequence. Moreover, user-pair feature combination is used to train multiple classifiers, the trained classifiers are used to predict scores based on multi-view. The multi-view scores are fused to achieve relationship strength.

In summary, the contributions to this paper are as follows:

- 1) We propose a multi-view context sequence method for spatiotemporal data. The multi-view context sequence considers co-occurrence, context and time periodicity simultaneously. The time is divided into three granularities and different granularities portray different degrees of co-occurrence and context.
- 2) We present a novel method to extract context co-occurrence feature based on the context sequence. Our method artfully transfers the user-pair relationship in spatiotemporal data to word-pair relationship in sequence.
- 3) We propose a decision fusion method based on multi-view to achieve a better result to offset the single view representation information deficiency.

The remainders of this paper are organized as follows. Section 2 discusses related work. Section 3 describes our methodology in detail. Section 4 reports our experiments. Section 5 presents results and discussion. Finally, we make conclusions in Section 6.

## II. RELATED WORK

We categorize the related works into three groups based on their focus: context based method [6], trajectory based methods [7-9] and co-occurrence based methods [2-5]. Our approach focuses on context and co-occurrence, and merge them together but different from traditional methods. The comparisons are in Table 1.

TABLE I. COMPARISON BETWEEN MVCC AND PRIOR WORKS

| Characteristics        | RWCFR [6] | EBM [3] | TAI [4] | SCI [5] | MVCC |
|------------------------|-----------|---------|---------|---------|------|
| location context       | √         |         |         |         | √    |
| temporal context       |           |         |         |         | √    |
| location co-occurrence |           | √       | √       | √       | √    |
| temporal co-occurrence |           | √       | √       | √       | √    |
| mobility periodicity   |           |         |         |         | √    |

In context based method, the context includes social context, personal preferences context, location context and temporal context. Bagci et al. [6] proposed a random walk based context-aware friend recommendation algorithm (RWCFR). They build a graph according to the current context (i.e. social relations, personal preference and location) of the user.

Trajectory based methods relaxed the concept of co-occurrence and use similarity in trajectory to measure likelihood of friendship [4]. Chen et al. [7] applied frequent sequential pattern mining technologies to extract the sequences of places that a user frequently visits, then use them to model his mobility profile combined with semantics of spatiotemporal information. [8-9] focused on measuring user similarity using trajectory patterns, and [8] provide a tool called MinUS which integrates the technologies trajectory pattern mining on discovering user similarity.

Co-occurrence based methods had been shown to improve accuracy of social relationship estimation than trajectory based methods because of exploitation feature of co-occurrence [4]. Grandall et al. [2] demonstrated that the co-occurrence feature contributes to inferring social ties by experiments over a dataset of 38 million geo-tagged photos from Flickr. Additionally, the probability of a social tie increases as the number of co-occurrence times increases and the temporal range decrease. [3-5] used co-occurrence feature and achieve good performance. Pham et al. [3] proposed an entropy-based model (EBM) that estimates the strength of social connections by analyzing people's co-occurrences in space and time through diversity and weighted frequency. Zhou et al. [4] proposed a Theme-Aware social strength Inference (TAI) approach that mines theme (also called the unit for co-occurrence) from co-occurrence behaviors, and then use the theme to measure the social strength of two persons. Njoo et al. [5] proposed a unified framework called SCI framework (Social Connection Inference framework), which quantified three key co-occurrence features (i.e. diversity,

stability and duration) and aggregated co-occurrence features using machine learning algorithms to predict the social ties.

Throughout the above methods, [2-5] shows the importance of co-occurrence feature which is also considered in our method. [6] shows that current context contribute to friend prediction, so our method also considers spatiotemporal context. The temporal context is included in our method which is not involved in [6]. Different from [3-6], the feature extraction method of spatiotemporal context co-occurrence has novelty, which is not a simple fusion. Peoples' mobility periodicity is also an import characteristic [10], and our method takes it into account through a special view. Generally, our method considers context and co-occurrence, combined with periodic mobility, shown in Table 1. Additionally, we extract context co-occurrence feature from different time views which captures different degree spatiotemporal representation.

## III. METHODOLOGY

In this section, we first give some important notations definition. Second, we describe how to generate multi-view context sequence and interpret each context sequence representation. Moreover, we propose a novel method to generate context co-occurrence feature based on above sequence. Finally, we describe decision fusion based on multi-view.

### A. Notations Definition

TABLE II. NOTATION OF PARAMETERS

| Variable         | Notation   |
|------------------|--|
| $u \in U$        | $u$ is a user id<br>$U$ is all different user id set $\{u_1, u_2, u_3, \dots\}$                  |
| $c$              | $\{u, t, l\}$ , a user check in at specific location $l$ at specific time $t$                    |
| $C_t$            | $\{c_1, c_2, \dots, c_n\}$ , all $c$ at the same time period                                     |
| $C_{l\&t}$       | elements in $C_t$ rank according shortest distance principle                                     |
| $C_t^{Sequence}$ | $\{C_t^1, C_t^2, \dots, C_t^N\}$ , the elements $C_t$ rank according time order                  |
| $C_l$            | $\{c_1, c_2, \dots, c_m\}$ , all $c$ at the same location  |
| $C_{l\&t}$       | elements in $C_l$ rank according time order  |
| $C_l^{Sequence}$ | $\{C_l^1, C_l^2, \dots, C_l^M\}$ , the elements $C_l$ rank according shortest distance principle |
| $\theta$         | a sequence that captures spatiotemporal context co-occurrence information                        |

In Table 2 we list the notations of parameters that we use. We denote  $u \in U$  as the user, and  $c \in C$  as the check-in data. Each  $c$  reflects the appearance of a user  $u$  at a specific location  $l$  at a specific time  $t$  with the form of  $\{u, t, l\}$ . Each user has many check-ins  $c$  information. The  $C_t = \{c_1, c_2, \dots, c_n\}$  represent all check-ins during the same time period.  $C_{l\&t}$  is a sequence that elements in  $C_t$  rank according location shortest distance principle.  $C_t^{Sequence} = \{C_t^1, C_t^2, \dots, C_t^N\}$  is a sequence that elements rank according time order.  $C_l = \{c_1, c_2, \dots, c_m\}$  represents all check-ins in the same location.  $C_{l\&t}$  is that elements in  $C_l$  rank according time order.  $C_l^{Sequence} = \{C_l^1, C_l^2, \dots, C_l^M\}$  is a sequence that elements rank according location shortest distance principle. The  $\theta$  is a

context sequence which only consists of user id. The parameters relationship is that  $\sum_{i=1}^M |C_i^i| = \sum_{j=1}^N |C_j^j| = |C| = |\theta|$  and the  $\theta$  usually includes many repetitive user id.

### B. Multi-View Context Sequence

We construct multi-view context sequence from two main views (location-time and time-location), shown in Fig. 2, to represent context and co-occurrence information. Based on granularities, the time is divided into three views.

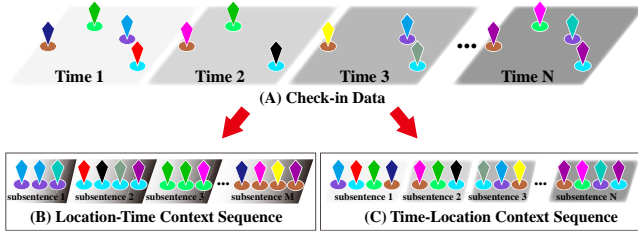


FIGURE II. THE RHOMBUSES WITH DIFFERENT COLORS DENOTE DIFFERENT USERS, THE ELLIPSES WITH DIFFERENT COLORS DENOTE DIFFERENT LOCATIONS, AND THE TIME CHANGE IS DENOTED BY THE SHADE OF THE BACKGROUND.

Fig. 2(A) shows that at a fixed time period, different users check-in at different locations. (B) shows location-time context sequence and (C) shows time-location context sequence. Note that the color of rhombuses, the color of ellipses and the change of shade.

#### Algorithm 1. Location-Time Context Sequence

```

Input: C
Output:  $\theta$ 
1 Define  $\theta \leftarrow \emptyset$ 
2  $C_i^{Sequence} = SortLocationByDistance(C)$ 
3 for all  $C_i$  in  $C_i^{Sequence}$  :
4    $C_{i\&t} = SortTime(C_i)$ 
5   for all  $c$  in  $C_{i\&t}$ :
6      $\theta.append(c.userid)$ 
7   end
8 end
9 return  $\theta$ 

```

**Location-Time Context Sequence.** The entire generation process of location-time context sequence is given in Algorithm 1. *SortLocationByDistance* function produce sequence  $C_i^{Sequence}$ . The elements in  $C_i^{Sequence}$  rank according distance nearest principle: if there is no location before, the first location is chosen randomly; otherwise, the location closest to the former location is assigned as the current location; the location closest to the  $(N-1)th$  location is assigned as the  $Nth$  location; and so on. *SortTime* function uses quick sort algorithm to rank according time order because time is one-dimensional information. The returned value of the Algorithm 1 is the location-time context sequence  $\theta$  consisted of user id, shown in Fig. 3 (B). Note that the same ellipse color represents the same location in Fig. 3 (B) and this context sequence capture strong location co-occurrence,

meanwhile including shortest location context and shortest time interval context.

**Time-Location Context Sequence.** The entire generation process of time-location context sequence is given in Algorithm 2. *SortTimeByGranularity* function uses mapreduce technique to produce  $C_i^{Sequence}$  quickly. The time parameter  $\tau$  can be accurate to day, hour or minute (24h, 1h, 1/60h). The elements in  $C_i$  rank according distance shortest principle. The returned value of the Algorithm 2 is the time-location context sequence  $\theta$ , shown in Fig. 3 (C). Note the distance between different ellipses in Fig. 3(C) and time-location context sequence capture time co-occurrence from different granularity, containing time order context and shortest location context.

#### Algorithm 2. Time-Location Context Sequence

```

Input: C,  $\tau$ 
Output:  $\theta$ 
1 Define  $\theta \leftarrow \emptyset$ 
2  $C_i^{Sequence} = SortTimeByGranularity(C, \tau)$ 
3 for all  $C_i$  in  $C_i^{Sequence}$  :
4    $C_{i\&t} = SortLocationByDistance(C_i)$ 
5   for all  $c$  in  $C_{i\&t}$ :
6      $\theta.append(c.userid)$ 
7   end
8 end
9 return  $\theta$ 

```

**Day-Location Context Sequence.** The time parameter  $\tau$  is set to day. As shown in Fig.2(A), every chunk is one day i.e. time 1 is the first day; time 2 is the second day and so on. The elements in the time chunk  $C_i$  rank according distance shortest principle. Day-location view is special time-location context sequence, because people's periodic movement is based on day [10] such as people commute on working day, people check in home at night. Therefore, this context sequence captures people's mobility periodicity, including day occurrence, time order context and shortest location context simultaneously.

**Hour-Location Context Sequence.** The generation processes are similar to day-location context sequence generation, but the granularity parameter  $\tau$  is set to the hour. This will generate more time chunks, but shorter sequence in each time chunk than day-location. The hour-location context sequence captures moderate time co-occurrence and spatiotemporal context.

**Minute-Location Context Sequence.** The generation processes are similar to hour-location context sequence generation, but the granularity parameter  $\tau$  is set to minute. This will generate more time chunks, but shorter sequence in each time chunk than hour-location view. The minute-location context sequence captures strong time co-occurrence and spatiotemporal context.

Different time-location context sequence portrays different degree spatiotemporal context and co-occurrence. However, the day-location context sequence includes other two time-location view information, meanwhile capturing the periodicity of people's movement. We do not adopt the second-location

context sequence because the check-in data in location-based social networks is very sparse in time. Of course, if the strong time co-occurrence is needed, the second-location view also can be considered. Through algorithm 1 and algorithm 2, four types of context sequences, including different degree spatiotemporal context and co-occurrence information, can be achieved from above four views.

### C. Multi-View Context Co-occurrence Feature

The context co-occurrence feature is not simple fusion between context and co-occurrence, different from traditional methods. We see each user id in spatiotemporal data as a word in the document sequence. We artfully use the toolkit word2vec to extract context co-occurrence feature through context sequence. The tool takes as its input a large corpus (corpus also can be seen as a sequence consist of words) and produces a dimensional space, with each unique word being assigned a vector in the space [11]. These word vectors are positioned in the vector space that words that share common contexts in the corpus are located in close proximity to one another in the space [11-12]. The word vectors is also context co-occurrence feature that we need. Through four types of user context sequences, we can obtain four types of user context co-occurrence feature, (i.e. location-time, day-location, hour-location and minute-location feature), shown in Fig. 2. These features includes spatiotemporal co-occurrence and context information.

More formally, given a context sequence  $\theta = \{u_1, u_2, u_3, u_4, \dots, u_T\}$  with spatiotemporal semantics information, our objective is to maximize the average log probability

$$\frac{1}{T} \sum_{t=1}^T \left[ \sum_{j=-k}^k \log p(u_{t+j} | u_t) \right] \quad (1)$$

where  $T$  is number of elements in  $\theta$  and  $k$  is the size of the window. The inner summation goes from  $-k$  to  $k$  to compute the log probability of correctly predicting the user  $u_{t+j}$  given the user in the middle  $u_t$ . The outer summation goes over all users in the context sequence. The values of the two ends of the window are filled by the boundary value. Every user  $u$  is associated with two learnable parameter vectors,  $w_u$  and  $v_u$ . They are the “input” and “output” vectors of  $u$  respectively which can be learned [12]. The probability of predicting the user  $u_i$  given the user  $u_j$  is defined as

$$p(u_i | u_j) = \frac{\exp(w_{u_i}^T v_{u_j})}{\sum_{l=1}^U \exp(w_l^T v_{u_j})} \quad (2)$$

where  $U$  is different users in the context sequence  $\theta$ . The optimization approach is using stochastic gradient descent and the gradient is computed using backpropagation rule [12]. Each user’s context semantic feature  $v_u$  (also called word vector in NLP domain) can be learned. The word vector  $v_u$  captures context and co-occurrence information, which is extremely useful for us.

The toolkit word2vec is usually used to find synonyms in document (a sequence consist of words), and we innovatively

apply it to finding user-pair relationship. The context feature captures all representation information that the context sequence has. Therefore, the context co-occurrence feature also includes spatiotemporal context and co-occurrence information, and different views has different degree of context co-occurrence, same to context sequence. The context co-occurrence feature is very different from above literatures [2-6], and not simply merges context and co-occurrence together.

### D. Multi-View Decision Fusion

The context co-occurrence feature with spatiotemporal semantics is combined to user-pair feature to train multiple classifiers from four views. The multiple classifiers are used to predict the relationship strength of user-pair in the test data, and four types of scores can be achieved i.e., location-time, day-location, hour-location and minute-location score, shown in Fig. 2. We consider every view characteristic and propose multi-view decision fusion.

Our decision fusion is a high-level fusion method, which can effectively reflect the different types of view information. When several view information is incorrect, the correct result can be achieved through high-level fusion. Specifically, our approach is calculated by the following formula:

$$RelationshipStrength = \sum_{d_i \in D} Score_{d_i} W(d_i) \quad (3)$$

where  $D$  is multi-views,  $d_i$  the single view, and  $W(d_i)$  the single view weight, and  $Score_{d_i}$  is the single view score. The value  $W(d_i)$  is calculated through two steps: we first calculate the value of  $W'(d_i)$  by the formula

$$W'(d_i) = \frac{Val\_AUC_i - Val\_AUC_{min}}{Val\_AUC_{max} - Val\_AUC_{min}} + \sigma \quad (4)$$

where  $Val\_AUC_{max}$  is the maximum  $AUC$  value of validation set from multi-view,  $Val\_AUC_{min}$  is minimum value, and the value  $\sigma$  is relatively small enough (such as the value  $\sigma$  set to 0.01). Then, the value of  $W'(d_i)$  is needed *L1 normalization* to achieve the value of  $W(d_i)$ . After obtaining  $Score_{d_i}$  and  $W(d_i)$ , The *RelationshipStrength* is calculated by using formula (3) and ranked in descending order for experiment evaluation. Although the decision fusion is seemingly simple, it is very effective because formula 4 gives more weight to the view that can better fit dataset distribution. The time granularity uncertainty of real data is adaptive to our strategy.

## IV. EXPERIMENTS

In this section, we first describe datasets and data imbalance solution. Second, we describe how to generate multi-view context co-occurrence feature. Moreover, we describe multiple classifiers learning and decision fusion based on multi-view.

### A. Datasets

Our experiments are based on two real datasets, the Brightkite and Gowalla [13]. The check-in data is handled to a triplet  $\langle u, t, l \rangle$ , where the  $l$  is represented by longitude and latitude. The user-pair data is handled to a triplet  $\langle u_1, u_2, label \rangle$

where the *label* is whether two users exist relationship. The overview is in Table 3.

TABLE III. EXPERIMENT DATASETS

| Datasets          | Brightkite | Gowalla   |
|-------------------|------------|-----------|
| Checkins          | 4,491,143  | 6,442,890 |
| Nodes (Users ID)  | 50,686     | 107,092   |
| Edges (User-Pair) | 194,090    | 456,830   |

The label of two datasets is very imbalanced [5, 13]. Up sampling method is widely used method to solve the imbalance, specifically we compound the minority instances to the same number of majority instances. This is because there are no negative instances in the original datasets that only offer positive instances (all labels are true). Random user-pair does not exist relationship other than in positive instances. Negative instances are compounded equal to the number of positive instances. The user-pair data is divided into train, validation, and test dataset.

### B. Context Co-occurrence Feature Generation

Location-time context sequence with spatiotemporal context and co-occurrence information can be generated through Algorithm 1. To achieve three types of time-location context sequences through Algorithm 2, the parameter  $\tau$  should be assigned to day, hour and minute (24h, 1h, 1/60h) respectively. Different views captures different degree context and co-occurrence. Then, four types of sequences with spatiotemporal information can be achieved.

The word2vec provides an implementation of the skip-gram architecture which is in accord with our objective function [12], so we choose the skip-gram architecture. The formula (2) is expensive because the cost of computing  $\nabla \log p(u_i|u_j)$  is proportional to the  $U$  [14]. An efficient alternative to the full softmax is the hierarchical softmax, which greatly reduces the complexity of computing  $\log p(u_i|u_j)$  (about logarithmically with respect to the  $U$ ) [14]. Therefore, we adopt the skip-gram and hierarchical softmax approach [12, 14] to extract context co-occurrence feature. The context co-occurrence feature size is set as 200 and the window of max skip length between users is set as 10 (the parameter  $k$  in Section 3.3 is 10). The learning rate is set as 0.01 and the value of the other parameters are default. After learning from context sequences, the feature with different degree spatiotemporal information (the vector parameter  $v_u$  in Section 3.3 called word vector in Natural Language Process domain) can be learned. Four different types of context co-occurrence features can be learned from multi-view and each user is mapped to four types of 200 dimensional context co-occurrence features. These context co-occurrence features represent different degree context and co-occurrence information in spatiotemporal data.

### C. Multi-View Classification and Decision Fusion

The multiple classifiers can be trained through four views with different degree context co-occurrence information. In our experiment, we choose the XGBoost classifier to make prediction. It is a supervised learning method that uses a tree boosting technique. For a given datasets with  $n$  examples and  $m$  features  $D = \{(x_i, y_i)\} (|D| = n, x_i \in R^m, y_i \in [0,1])$ , a tree ensemble model uses  $K$  additive functions to predict the output:

$$\bar{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (5)$$

where  $F$  is the space of the regression trees [15]. The output  $\bar{y}_i$  is the relative probability of a user-pair relationship strength. The user is represented with the context co-occurrence feature and then the user-pair features combination as the input. Considering relationship is bidirectional, a relatively larger vector position value is placed ahead of the corresponding position value. There are some primary parameters: the booster parameter is set as gbtree; the max depth is 3 which avoid overfitting; the boosting learning rate is 0.1; the objective function is binary logistic; the iteration time is 1000; and other parameters are set as default. The four validation AUC (the Val\_AUC in Section 3.4) can be achieved from four classifiers [15, 16]. Multiple classifiers are used to predict relationship score in the test data, and four types of scores can be achieved i.e., location-time, day-location, hour-location and minute-location score, shown in Fig. 1.

The  $D$  in Section 3.4 includes four views. We use the *Val\_AUC* values and formula (4) to calculate the  $W'(d_i)$ , and then use *L1 normalization* to get the  $W(d_i)$ . *RelationshipStrength* is the result of user-pair in test data which is calculated through formula (3), and ranked in descending order for experiment evaluation. Generally, the higher the score is, the higher strength the two users have relationship.

## V. RESULTS AND DISCUSSION

We use the AUC metric (the area under the *ROC* curve) [17] to evaluate. The higher AUC indicates the better prediction performance. The precision-recall curve (PR-Curve) [18] is another widely used metric.

We compare AUC metric (shown in Table 4) and PR-Curve (shown in Fig. 3) from four views and decision fusion in test data. Location-time view captures strong location co-occurrence, shortest location context and shortest time interval context. Day-location view is a special time-location view, because people check-in location usually is periodicity of the day [10] such as people commute on working day. The day-location view better portrays strong periodic time co-occurrence, spatiotemporal context. People's periodic check-in characteristic leads to day-location view better than other two time-location view. The hour-location view captures moderate time co-occurrence and spatiotemporal context. The minute-location view is worse than other two time-location view in current two dataset, because the experiment dataset is very sparse in time. Although minute-location view is not good in current datasets, it captures strong time co-occurrence and nearest spatiotemporal context. The minute-location view is very appropriate to the dense dataset that short time co-occurrence situation often appear, such as friends meet together in coffee shop during short time interval. By merging the four views' scores, the performance of decision fusion is higher than the single view result. In our approach, the four views are different which satisfy the diversity rule, and each view stress different degree spatiotemporal context and co-occurrence. Additionally, day-location view captures periodic co-occurrence which fitting peoples' periodic check-in. The formula (4) provides more weight to the good view and the

comparison shows that our strategy balances shortcomings of a single view.

We also compare MVCC approach with three representative approaches, which are the state-of-the-art approach (SCI) [5], the classical approach (EBM) [3] and the PGT approach [19]. The SCI method [5] is state-of-the-art method in recent literatures. The EBM [3] is classical method that acquires immense breakout at that time. The PGT method considered personal and global factors, combined with temporal factor, to measure the strength of the relationship between two given mobile users. We use our approach to make comparison with the PGT approach because it achieved decent performance. Fig. 4 shows our approach outperforms classical approach EBM by 2% to 8% in recall under the same precision. In Fig. 5, we observe that the AUC of our approach outperforms the SCI [5] and PGT [19] methods. The co-occurrence feature is important in relationship strength [2], and [3, 5] approaches take co-occurrence into account. Our method considers spatiotemporal co-occurrence, context and periodic mobility simultaneously.

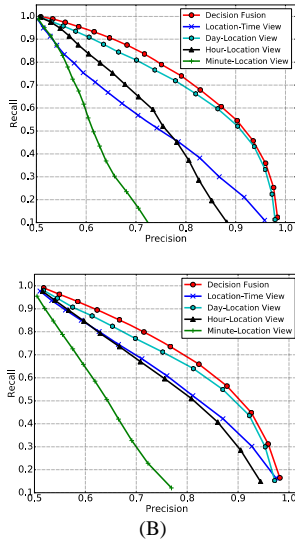


FIGURE III. (A)PR-CURVE ON BRIGHTKITE (B)PR-CURVE ON GOWALLA

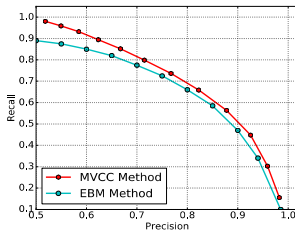


FIGURE IV. PR-CURVE ON GOWALLA

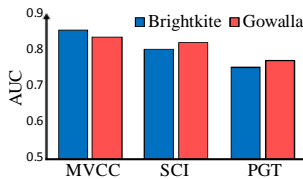


FIGURE V. AUC COMPARISON

TABLE IV. AUC OF DECISION FUSION AND OTHER VIEWS

| Testing Set            | Brightkite      | Gowalla         |
|------------------------|-----------------|-----------------|
| Location-Time View     | 0.730691        | 0.778668        |
| Day-Location View      | 0.829309        | 0.806326        |
| Hour-Location View     | 0.747866        | 0.761170        |
| Minute-Location View   | 0.635823        | 0.644667        |
| <b>Decision Fusion</b> | <b>0.851635</b> | <b>0.833099</b> |

## VI. CONCLUSIONS

In this paper, we propose a novel approach named Multi-View Context Co-occurrence to predict relationship in spatiotemporal data. We construct context sequences and then artfully extract context co-occurrence features, different from traditional methods. Our method is based on multi-view and different view represents different degree spatiotemporal context co-occurrence. The experiment results show the feasibility of our approach, and evaluation shows that our approach outperforms the baseline.

## ACKNOWLEDGEMENT

This work is supported by NSFC (Grant No. 61272449, 61373 092 and 61572339) and Science and Technology Support Pro-gram of Jiangsu (Grant No. BE2014005).

## REFERENCES

- [1] Machanavajjhala A, Korolova A, Sarma A D. Personalized social recommendations: accurate or private[J]. Proceedings of the VLDB Endowment, 2011, 4(7): 440-450.
- [2] Crandall D J, Backstrom L, Cosley D, et al. Inferring social ties from geographic coincidences[J]. Proceedings of the National Academy of Sciences, 2010, 107(52): 22436-22441.
- [3] Pham H, Shahabi C, Liu Y. Ebm: an entropy-based model to infer social strength from spatiotemporal data[C]. Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data. ACM, 2013: 265-276.
- [4] Zhou N, Zhang X, Wang S. Theme-aware social strength inference from spatiotemporal data[C]. International Conference on Web-Age Information Management. Springer, Cham, 2014: 498-509.
- [5] Njoo G S, Kao M C, Hsu K W, et al. Exploring check-in data to infer social ties in location based social networks[C]. Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Cham, 2017: 460-471.
- [6] Bagci H, Karagoz P. Context-aware friend recommendation for location based social networks using random walk[C]. Proceedings of the 25th international conference companion on world wide web. International World Wide Web Conferences Steering Committee, 2016: 531-536.
- [7] Chen X, Pang J, Xue R. Constructing and comparing user mobility profiles for location-based services[C]. Proceedings of the 28th Annual ACM Symposium on Applied Computing. ACM, 2013: 261-266.
- [8] Chen X, Kordy P, Lu R, et al. MinUS: Mining user similarity with trajectory patterns[C]. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2014: 436-439.
- [9] Chen X, Lu R, Ma X, et al. Measuring user similarity with trajectory patterns: Principles and new metrics[C]. Asia-Pacific Web Conference. Springer, Cham, 2014: 437-448.
- [10] Cho E, Myers S A, Leskovec J. Friendship and mobility: user movement in location-based social networks[C]. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2011: 1082-1090.
- [11] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.

- [12] Mikolov T, Le Q V, Sutskever I. Exploiting similarities among languages for machine translation[J]. arXiv preprint arXiv:1309.4168, 2013.
- [13] Leskovec J, Krevl A. {SNAP Datasets}:{Stanford} Large Network Dataset Collection[J]. 2015.
- [14] Morin F, Bengio Y. Hierarchical Probabilistic Neural Network Language Model[C]. Aistats. 2005, 5: 246-252.
- [15] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, 2016: 785-794.
- [16] <https://github.com/dmlc/xgboost>.
- [17] Bradley A P. The use of the area under the ROC curve in the evaluation of machine learning algorithms[J]. Pattern recognition, 1997, 30(7): 1145-1159.
- [18] Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves[C]. Proceedings of the 23rd international conference on Machine learning. ACM, 2006: 233-240.
- [19] Wang H, Li Z, Lee W C. PGT: Measuring mobility relationship using personal, global and temporal factors[C]. Data Mining (ICDM), 2014 IEEE International Conference on. IEEE, 2014: 570-579