International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 22 (2018) pp. 15633-15637

An Improved Multi-Context Trajectory Embedding Model using Parameter Tuning Optimization for Human Trajectory Data Analysis

Suryakumar, B.
Ph.D. Research Scholar, Department of Computer Science, NGM College, Pollachi, Tamilnadu, India.

Dr. Ramadevi, E.
Associate Professor, Department of Computer Science, NGM College,Pollachi,Tamilnadu, India.

Abstract

As the explosion of location-based social networks such as Facebook, etc., a number of ways have been provided for tracing human mobility, including user-generated Geo-tagged contents, check-in-services and mobile applications. For mining human trajectory data, different techniques were proposed over the past decades. However, the issue in many applications was analyzing and mining trajectory data due to the complex characteristics reflected in human mobility which is affected by multiple contextual information. As a result, Multi-Context Trajectory Embedding Model using Convolutional Neural Network (MC-TEM-CNN) was proposed that reduces the computation time during the learning process of contextual features. However, it requires an optimization algorithm to enhance the tuning of parameters which are needed to model the contextual information. Hence in this article, an Improved Multi-Context Trajectory Embedding Model (IMC-TEM) is proposed based on the frog-leaping optimization algorithm. In this algorithm, the parameters are tuned according to the frog characteristics. In each iteration, the global best fitness is chosen to adjust the position of worst fitness frogs. Thus, the proposed IMC-TEM tunes parameters in a better manner. Finally, the experimental results are conducted based on three real-world datasets to observe the performance efficiency of the IMC-TEM than MC-TEM-CNN.

Keywords: location-based social networks, human trajectory, contextual information, MC-TEM-CNN, frog-leaping optimization.

INTRODUCTION

Due to the permeation of positioning-enabled mobile devices, the utilization of location-based social networks like Facebook, Foursquare, etc., are increased and they are providing a different way for recording human mobility with user-generated geo-tagged contents (i.e., tweets, photos), check-in services and mobile apps. Specifically, location-based social networks are known as the digital mirror to human mobility in a physical world, since it provides an opportunity for completely understanding the people's behavior i.e., users tend to have different spatial and temporal activity preference as their lifestyles [1]. Based on the human mobility which reflects by trajectory data, it can be used for location-based recommendation applications such as personalized location prediction [2], group-based location recommendation [3] and user mobility modelling [4].

Different methods and analysis have been investigated for spatiotemporal data like user's check-in records which refers to social strength among users and can be used for link prediction [5]. However, challenges are addressed for analyzing and mining the trajectory data due to highly complex characteristics in human mobility. Trajectory data itself is a type of sequential data and such surrounding contexts are significant for considering trajectory modeling. Moreover, the other issue is that the model complexity is high due to the incorporation of additional contexts i.e., tensor decomposition. Therefore, a more flexible method is required for characterizing the multiple types of contextual information to model the trajectory data.

In location-based social network systems [6], the human trajectory data was analyzed and mined by using a multi-context embedding model which is called as MC-TEM. This MC-TEM model was developed in the distributed representation learning method such as deep neural networks for exploring the contexts in a systematic way. This model was incorporated into user-level, trajectory-level, location-level and temporal contexts. In this method, the overall objective function for a given trajectory sequence was used to maximize the average log probability for each location. The softmax multi-class classifier was employed to generate the check-in-location conditioned on its contextual information. The contextual features were represented based on the hierarchy model. A three-level hierarchy was developed to organize the contextual features in a top-down fashion such as user level, trajectory level and location level. Also, temporal contexts were considered as a contextual feature. However, the computation time was high due to deep learning algorithm and additional parameters were required for parameter learning process. Therefore, MC-TEM was improved by CNN to learn the contextual features in an efficient manner [7]. However, the parameters required to model MC-TEM-CNN were fixed. Thus, it requires an efficient parameter tuning process to enhance the social link prediction performance.

Hence in this article, Improved Multi-Context Trajectory Embedding Model (IMC-TEM) is proposed to reduce the computational time and cost during parameter tuning process. In this method, the parameters are tuned based on the behavior of the frog. This algorithm is performed based on the evolution of memeplex which is carried by the number of interacting frogs that perform a global exchange of information among its population. Here, the global best fitness is selected and applied to improve the worst fitness frogs in each cycle. Based on the selected fitness value, the position of worst fitness frogs is adjusted. Thus, the parameters are tuned efficiently.

The rest of the article is structured as follows: Section II presents the works which are related to the trajectory data mining methods. Section III explains the proposed methodology. Section IV illustrates the performance evaluation of the proposed system. Section V concludes the research work.

LITERATURE SURVEY

Different methods and applications were discussed for trajectory data mining [8]. Initially, generic methods of trajectory mining and the relationships between them were discussed. Then, application problems were studied and classified for solving them based on the utilization of trajectory data. The trajectory-mining application problems were classified under major problem groups according to how they were related. This approach can be used by researchers in the identification of gaps between methods and inspiring them for developing the new methods. However, this approach requires customizing algorithms for performing a context-aware mining of trajectory data.

Geographical influence [9] was investigated for the collaborative point-of-interest recommendation. In this investigation, the Point-Of-Interests (POI) recommendation service was provided for the rapid growing Location-Based Social Networks (LBSN). The objective of this work was to explore the user preference, social influence and geographical influence for POI recommendations. The user preference was derived based on the user based collaborative filtering and social influence due to spatial clustering was explored. The collaborative recommendation algorithm was developed based on the geographical influence based on naive Bayesian. Unified POI recommendation framework was also proposed for fusing user preference to POI with social influence and geographical influence. The random walk and restart may not be suitable for POI recommendation in LBSN.

N-dimensional tensor factorization [10] was proposed for context-aware collaborative filtering. In this algorithm, an extension model was proposed to N-dimensions through the utilization of tensors. Here, a generic CF model was presented based on the generalization of matrix factorization for addressing the contextual recommendation problems. However, this model requires to be further exploring the temporal dependencies in standard CF settings and also how multidimensional tensor factorization can be used for modeling the non-contextual variables.

Points of Interest (POI) recommendation based context-aware [11] was investigated in mobile social networks. The POI recommendation model was developed based on context-aware built by combining spatial-temporal factors and behavior of users. Initially, the history of user activity was analyzed and the user's interests and preferences were mined by the region model based on location awareness. Spatial-temporal elements and user profiles from check-in time were extracted by a topic model based on location context-awareness. These two models were combined for POI recommendation based on context-aware and evaluated the recommendation satisfaction index (RSI). However, the complexity was increased due to the incorporation of more contexts.

Context-aware Location recommendation algorithm [12] was proposed by using Random Walk-based method (CLoRW) for LBSN. The current context such as preferences and location of the users were considered to provide personalized recommendations. The undirected and unweighted graph model of LBSN was developed to perform a random walk approach with the restart. This random walk was performed to calculate the recommendation probabilities of the nodes. A list of recommendation was recommended to users after ordering the nodes according to the estimated probabilities. However, the complexity of this algorithm depends on the iteration count which changes according to the graph size.

The investigation [13] on human mobility, social ties and link prediction methods was presented. In this study, the trajectories and communication patterns were utilized by using Call Detail Record (CDR) data from an anonymous country for measuring any pair of users. The user mobility was identified by introducing a series of co-location measures quantifying the similarity between their movement routines. The users were connected to the social network by adopting various well-established measures of network proximity based on common neighbors or structure of paths connecting the users in who-calls-whom network. The powerful interaction between users was achieved by the using number of calls between the users as a measure of the strength of their tie. The further improvement was required for link prediction by mixing mobility and network measures.

PROPOSED METHODOLOGY

In this section, the proposed IMC-TEM is explained in brief. Initially, the following preliminaries are described which are utilized to model the multi-context trajectory data.

- ***Check-in Record:*** When a user u checks in the location l with a category label c at the timestamp s, the check-in record is modeled as a quadruple $\langle u, l, c, s \rangle$.
- ***Trajectory:*** For a given user u, a trajectory t denotes a sequence of chronologically ordered check-in records associated to u: $\langle u, l_1, c_1, s_1 \rangle, \dots, \langle u, l_i, c_i, s_i \rangle, \dots, \langle u, l_N, c_N, s_N \rangle$, where N refers the sequence length and $s_i < s_{i+1}$ for $i \le N-1$.

For a given trajectory sequence, the overall objective function is maximizing the average log probability for each location given its corresponding contextual information and is given as follows:

$$\frac{1}{N}\sum_{j=1}^{N} \log P\left(l_j | x^{(l_j)}\right) \tag{1}$$

In equation (1), $x^{(l_j)}$ is the real-valued feature vector which consists of all contextual information for the target location l_j. Every dimension in $x^{(l_j)}$ corresponds to the unique contextual feature and $x_f^{(l_j)}$ is the weight of the f^{th} feature in $x^{(l_j)}$. Consider, $x^{(l_j)}$ is the nonnegative vector, each entry denotes the number of occurrences for the feature in the context respectively. For modeling trajectory data, MC-TEM-CNN is applied by using the distributed representation.

The MC-TEM-CNN model is performed based on the two significant parameters such as the vector size (V_S) and context window length (W_L) which are constant. In this proposed IMC-TEM, those parameters are optimally selected based on the Shuffled Frog-Leaping Optimization Algorithm (SFLOA) to enhance the social link prediction by contextual information.

Parameter Tuning

The parameters V_S and W_L are tuned by using SFLOA which performs a heuristic search according to the evolution of particles named memes carried by a number of frogs that perform a global exchange of information among the population. This algorithm tries to imitate the search for food by a group of frogs that exchange information among themselves. Each frog has a certain location within the search space (X^i). This vector represents a meme with various memotypes as decision variables (N_{VD}). Each memotype identifies the discrete value of each decision variable.

$$X^i = \{X_1^i, X_2^i, \dots, X_{N_{VD}}^i\} \quad (2)$$

The global exchange of information between the memes has a probabilistic component. The major parameters of SFLOA are the number of memeplexes (m), the number of frogs per memeplex (n), the ratio of frogs in the memeplex that will evolve (q), the number of memetic evolutions (N_S) within a sub-memeplex before shuffling and search-acceleration factor (C). Here, C is used for preventing premature convergence and balancing global and local searches. Based on this, the global search is accelerated by assigning high values to C at the initial stage of evolution process since larger changes in the frog's location will be allowed.

Consider the initial population (P) which is generated randomly by SFLOA. Each of these frogs with a possible solution is sorted based on the value of the objective function as follows:

$$f(i) = \frac{1}{N_t} max \sum_{j=1}^{N_t} log\, P\big(l_j | u, t, l_{j-K}: l_{j+K}, c_{j-K}: c_{j+K}, d, h\big) \quad (3)$$

In equation (3), u refers the user-level context, t refers to the trajectory-level context, $l_{j-K}: l_{j+K}, c_{j-K}: c_{j+K}$ denotes the location-level context, d and h are the temporal contexts. Also, N_t is the length of the trajectory t and the main aim of this objective function is maximizing the average log-probability for each location with its corresponding contextual information. To achieve this, the considered P is split into m number of memeplexes. Each memeplex contains n frogs and can be assumed as a different culture in which a local search is performed. After that, frogs are forwarded to different memeplexes based on their cost function. The global fitness is denoted as X_g and the best and worst solutions for each memeplex are denoted as X_b and X_w correspondingly.

Then, each memeplex is divided into sub-memeplex that represents the number of frogs entering memetic evolution. Within each sub-memeplex, frogs exchange information, thus the best informs to the worst which evolves in a process called an evolutionary leap. In this process, only the frog with the worst cost function in each iteration is updated as follows:

$$L_i = \delta \times C \times (X_b - X_w) \quad (3)$$

$$X_{w,1} = X_{w,0} + L_i(L_{max} \geq L_i \geq -L_{max}) \quad (4)$$

In above equations, L_i is the change in frog location, δ refers a random number between 0 and 1, $X_{w,0}$ is the current location of the frog, $X_{w,1}$ is the new location of the frog and L_{max} is the maximum allowed a change in a frog's location. If the evolution produces a better frog, it replaces the worst frog or else, X_b is replaced by X_g in (3) and the process is continued. If the fitness of the new frog is not better than the fitness of X_w, then a new frog is generated randomly for replacing the worst frog. This process is repeated for a particular number of iterations i.e., N_S within each sub-memeplex. Thus, the local search in each sub-memeplex is finished and the sub-memeplexes are returned to memeplexes.

The memeplexes are dissolved and the shuffling process is initiated where frogs are mixed again according to their cost function and re-sorted into new memeplexes. Based on this, a generation is completed. In the end, this algorithm has the ability to evolve a random initial population to the global minimum. The leaping and shuffling processes are continued until the condition of convergence is satisfied. Thus, the optimal solution is obtained based on this algorithm to set IMC-TEM efficiently

RESULTS AND DISCUSSIONS

In this section, the results of the experiments which are conducted on two application tasks namely location recommendation and social link prediction of both proposed and existing are illustrated. In this experiment, three public geo-social networking datasets are used such as $Foursquare_S$, $Foursquare_L$ and $Gowalla$. All datasets contain check-in records in the form of (User ID, Location ID, Location Category, Timestamp, City). Among these datasets, only $Foursquare_S$ and $Gowalla$ contain the social connection links among users. Moreover, $Foursquare_L$ and $Gowalla$ are utilized for location recommendation whereas $Foursquare_S$ and $Gowalla$ are utilized for link prediction. The detailed information of these three datasets is given in Table 1.

Table 1: Statistics of Datasets

Dataset	No. of Users	No. of Check-ins	No. of Links	No. of Locations
$Foursquare_S$	4163	483814	32512	121142
$Foursquare_L$	266909	33278683	-	3680126
$Gowalla$	216734	12846151	736778	1421262

Analysis on Location Recommendation

The effectiveness of location recommendation along both general location recommendation and time-aware location recommendation is evaluated by considering home-city recommendation settings. For a given user, the corresponding home-city is identified as the city with the most number of occurrences in his check-in records. The training and testing datasets are constructed by splitting the data based on the trajectories. Initially, 20% trajectories with only home-city locations are selected as a testing dataset and the remaining trajectories are chosen as training data.

In the testing dataset, 1000 locations which are not visited by the current user are selected randomly for each check-in record. In addition, a candidate list of 1001 locations combined with the target location is selected for recommendation. After that, the locations in the candidate list are ranked by using a recommender system. A ranked list is formed by sorting all the 1001 locations based on their ranking scores. Consider n is the rank of the target location within this list and the optimal result relates to the case where it precedes all the additional locations $(n = 1)$. Then, a top-k recommendation list is formed by collecting the top k ranked locations from the list.

Here, hit_k is defined for a single test case as either the value is 1 if the target location is identified in the top k results, otherwise, the value is 0. The overall $Recall_k$ refers to the ratio of hits in all the test check-in records.

$$Recall_k = \frac{Number\ of\ hit_k}{Number\ of\ all\ cases} \tag{5}$$

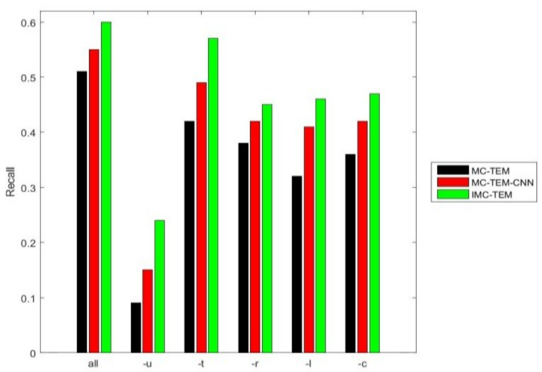

Figure 1: Comparison on General Location Recommendation

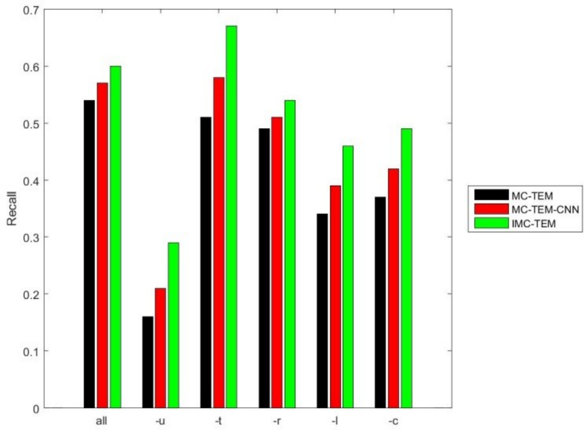

Figure 2: Comparison on Time-aware Location Recommendation

Figure 1 and Figure 2 shows that the impacts of contextual features for general and time-aware location recommendation on the $Foursquare_L$ dataset. In the graph, x-axis denotes the information about contextual factors where u denotes the user context, t denotes the trajectory context, r denotes the city/region context, l denotes the location context and c denotes the category context respectively. Additionally, in Figure 2, d denotes the day context and h denotes the hour context. Here, '-' indicates that the corresponding context type is not included. Also, y-axis refers the recall values. From the analysis, it is observed that all the considered contexts are useful for both recommendation tasks. Moreover, the user context is the most significant factor to be considered since the current task is essentially a personalized recommendation issue where user preference plays the key role in system performance. Finally, it is concluded that the proposed IMC-TEM achieves higher recall value while considering all the contextual factors compared to the MC-TEM-CNN and MC-TEM.

Analysis on Social Link Prediction

The objective of this analysis is predicting whether a social link between a pair of users exists or not only based on their trajectory data. All the trajectory information for a user is assumed as those are available for unsupervised feature extraction. Also, the social connection links are split into training and testing dataset. The user pairs without links are generated for both training and testing datasets by selecting non-linked user pairs randomly with the ratio of 1:1 compared to the social connection links.

Consider P_T is the set of all user pairs with real friend links and P_R is the number of all user pairs identified by a candidate as friends. Then, the effectiveness of social link prediction is measured based on the value of recall which is computed as follows:

$$Recall = \frac{|P_T \cap P_R|}{|P_T|} \tag{6}$$

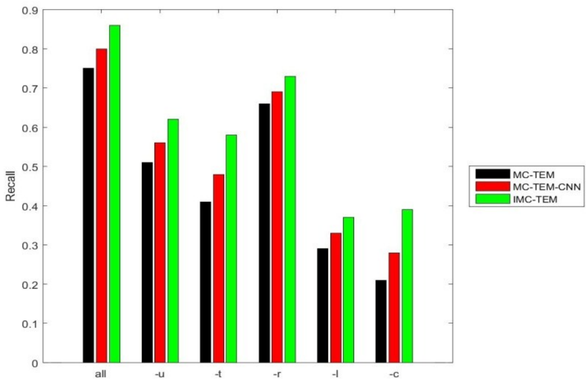

Figure 3: Comparison on Social Link Prediction

Figure 3 shows that the impacts of contextual features for social link prediction on the $Foursquare_5$ dataset. In the graph, x-axis denotes the information about contextual factors where u denotes the user context, t denotes the trajectory context, r denotes the city context, l denotes the location context and c denotes the category context respectively. Here, '-' indicates that the corresponding context type is not included. Also, y-axis refers the recall values. From the analysis, it is observed that all the considered contexts are useful for predicting the social links. User trajectory and location contexts have a less significant effect on recall while category context has a more significant effect. Finally, it is concluded that the proposed IMC-TEM achieves higher recall value while considering all the contextual factors compared to the MC-TEM-CNN and MC-TEM.

CONCLUSION

In this article, an Improved Multi-Context Trajectory Embedding Model (IMC-TEM) is proposed for exploring contexts in a systematic way incorporating user-level, trajectory-level, location-level and temporal contexts. The proposed model has more flexibility in characterizing the different types of contexts for various applications by utilizing the trajectory data. In this approach, the parameters required to model the contextual information are optimally learned based on the frog-leaping algorithm efficiently. Also, all the contextual information is represented in the equivalent embedding space for analyzing the association among different contexts with less complexity. Finally, the experiments are conducted based on the three datasets and the experimental results illustrated that the proposed IMC-TEM achieves higher performance than the existing model.

REFERENCES

[1] N.J. Yuan, F. Zhang, D. Lian, K. Zheng, S. Yu and X. Xie. "We know how you live: exploring the spectrum of urban lifestyles," in *Proceedings of the first ACM conference on Online social networks,* pp. 3-14, 2013.

[2] W. Mathew, R. Raposo and B. Martins. "Predicting future locations with hidden Markov models," in *Proceedings of the 2012 ACM conference on ubiquitous computing,* pp. 911-918, 2012.

[3] Q. Yuan, G. Cong and C.Y. Lin, "COM: a generative model for group recommendation," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining,* pp. 163-172, 2014.

[4] Q. Yuan, G. Cong, Z. Ma, A. Sun and N.M. Thalmann, "Who, where, when and what: discover spatio-temporal topics for twitter users," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining,* pp. 605-613, 2013.

[5] E. Cho, S.A. Myers and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining,* pp. 1082-1090, 2011.

[6] N. Zhou, W.X. Zhao, X. Zhang, J.R. Wen and S. Wang, "A general multi-context embedding model for mining human trajectory data," *IEEE Transactions on Knowledge and Data Engineering,* vol.**28**(8), pp. 1945-1958, 2016.

[7] B. Suryakumar and E. Ramadevi, "A multi context embedding model based on convolutional neural network for trajectory data mining," *International Journal of Computer Science and Mobile Applications,* vol.**5**(9), pp. 1-9, 2017.

[8] J.D. Mazimpaka and S. Timpf, "Trajectory data mining: a review of methods and applications," *Journal of Spatial Information Science,* vol.**2016**(13), pp. 61-99, 2016.

[9] M. Ye, P. Yin, W.C. Lee and D.L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval,* pp. 325-334, 2011.

[10] A. Karatzoglou, X. Amatriain, L. Baltrunas and N. Oliver, "Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering," in *Proceedings of the fourth ACM conference on Recommender systems,* pp. 79-86, 2010.

[11] F. Wong, S. Lee and Q. Wong, "Points of interest recommendation based on context-aware," *International Journal of Hybrid Information Technology*, vol.**8**(3), pp. 55-62, 2015.

[12] H. Bagci and P. Karagoz, "Context-aware location recommendation by using a random walk-based approach," *Knowledge and Information Systems*, vol.**47**(2), pp. 241-260, 2016.

[13] D. Wang, D. Pedreschi, C. Song, F. Giannotti and A.L. Barabasi, "Human mobility, social ties, and link prediction," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining,* pp. 1100-1108, 2011.