International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958, Volume-8, Issue-3S, February 2019

An Effective Identification of Human Trajectory Data Using Parameter Tuning Optimization Technique

B. Suryakumar, E. Ramadevi

Abstract--- ***As the tremendous growth of location enabled social networks sites such as Facebook, Foursquare etc., a number of ways have been provided for tracing human movements from one location to another location, containing user-created contents like geographically tagged records, mobile technology embedded services and applications. For the datamining purpose human trajectory data, various type of techniques was proposed over the past decades. However, the main issue in many applications was analysing process of data mining trajectory data owing to the composite features mirrored in human mobility which is directly connected by different contextual information. So that the Multi-Context Trajectory Embedding Model using Convolutional Neural Network (MC-TEM-CNN) was proposed that reduces the computation time during the process of learning contextual features. However, it requires an optimization of algorithm to enhance the tuning of parameters which are needed to model the different contextual information. So 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 main priority is for the parameters tuning process. The parameters are tuned according to the frog characteristics. For attaining this in each iteration, the universally best fitness is chosen to adjust the location of worst fitness frogs. Thus, the proposed IMC-TEM tunes parameters in an efficient manner. In the last step, the experimental results are conducted based on three real-world datasets to observe the performance efficiency of the IMC-TEM than MC-TEM-CNN***

1. INTRODUCTION

Due to the acceptability of location-oriented mobile devices, the utilization of location based social networks like Facebook, Foursquare, etc., are increased day by day and they are focussed a different way for recording human movement with user-created geo-tagged contents (i.e., tweets, photos and survey results), check-in services and mobile applications. Exactly, the location-oriented social networks are known as the digital mirror to human mobility in a IT era. It enables an opening for fully understanding or tracing the people's behaviour i.e., users have various spatial and temporal activity based on their lifestyles [1]. It is based on the human mobility which reflects by trajectory data, it can be used for location-based service recommendation applications such as customised location prediction [2], based on group-based, location recommendation [3] and the end user mobility and 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 analysing and mining the human trajectory data because of the complex characteristics in human movements. Trajectory contains sequential data and surrounding its contexts are significant for considering trajectory modelling process. 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 term known as multi-context embedding model which is named 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, location-level, trajectory-level, and finally temporal contexts. In this method, the overall objective function for a given trajectory sequence was employed to enhance the average log probability based on the location. Here the softmax multi-class classifier was engaged to generate the check-in-location conditioned on different contextual information. The contextual features were represented based on the hierarchy model. For establish the contextual features in top-down fashion, a three-level hierarchy was developed which include 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.

To achieve fine tuning of parameters we introduced, a new method named Improved Multi-Context Trajectory Embedding Model (IMC-TEM) is proposed to reduce the computational time and cost during parameter tuning process. By using this method, the parameters are tuned based on the behaviour of the frog. This algorithm is executed based on the advancement of memeplex. Memeplex is maintained by the number of interacting frogs that perform a global exchange of information among its population. In our experiment, the global best fitness is selected and applied to improve the worst fitness frogs in each cycle.

Revised Version Manuscript Received on 14 February, 2019.
B. Suryakumar*, Ph.D Research Scholar, Department of Computer Science, NGM College, Pollachi. (e-mail: mailtoyeske@gmail.com)
Dr.E. Ramadevi, Associate Professor, Department of Computer Science, NGM College, Pollachi. (e-mail: ramajus@hotmail.com)

Based on the selected fitness value, the position of worst fitness frogs is fine-tuned and adjusted. Thus, the parameters are tuned efficiently.

The remaining work is framed as follows: Section II deals the works which are related to the trajectory data mining methods. Section III presents the proposed methodology. Section IV explains the performance evaluation of the proposed system. Finally, the Section V concludes the research work.

2. 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.

The Geographical influence [9] was investigated for the point-of-interest recommendation. This investigation, the Point-Of-Interests (POI) service was provided for the fastest developing Location-Based Social Networks (LBSN). The objective of this work was to explore the preference of user, social sway and geographical impact for POI recommendations. The user preference was derived on the user based collaborative filtering and social influence due to spatial clustering was explored. The collaborative recommendation based 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. The algorithms like random walk and restart may not be suitable for POI recommendation in LBSN.

The N-dimensional factorization of tensor[10] was proposed for context-aware collaborative filtering process. 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 modelling 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 behaviour of users. Initially, the history of user activity was analysed 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 the preferences of location based users were considered to provide personalized recommendations. The undirected and unweighted graph model of LBSN was developed to accomplish a random walk approach with the restart. This random walk algorithm was accomplished to calculate the recommendation based 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 with respect to the graph size.

The investigation [13] on, social ties, human mobility 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 unknown country for measuring any pair of active users. The user mobility was identified by introducing a series of methods co-location measures for measuring the similarity between their movement procedures. The users were connected to the social network by adopting various well-established measures of network proximity based on similar neighbours or structure of paths connecting the users in who calls whom which network. The powerful interaction between users was achieved by the using number of rings or pings 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 area, the proposed method IMC-TEM is explained in brief. Initially, the following preliminaries are described which are utilized to model the multi-context trajectory data.

Based on Check-in Record: Here a user u track the location name l with a category of label c at the time stamp s, the check-in record is modelled as a quadruple $\langle u, l, c, s \rangle$.

Based on Trajectory: In this method for a given user u, a trajectory t denotes an order of occurrence based check-in records associated to u: $\langle u, l_1, c_1, s_1 \rangle, \ldots, \langle u, l_i, c_i, s_i \rangle, \ldots, \langle u, l_N, c_N, s_N \rangle$, where N refers the sequence length and $s_i < s_{i+1}$ for $i \leq N-1$.

For a given trajectory sequence, the overall objective of this function is to maximize the average log probability for each different location. The corresponding contextual based information is given as follows:

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

In the above equation (1), $x^{(l_j)}$ is the real-valued feature enabled vector which consists of all the relative information for the target location l_j. Every dimension in $x^{(l_j)}$ corresponds to the 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 represents the number of occurrences for

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

the above feature in the context respectively. For modelling trajectory data, in method MC-TEM-CNN is applied by using the in the distributed representation.

The MC-TEM-CNN model is performed based on the two significant parameters. The parameters are 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.

3. PARAMETER TUNING PROCESS

The two parameters V_S and W_L are tuned by using SFLOA which performs the method of searching called heuristic search according to the evolution of particles. This particles named memes that conceded by a number of frogs that implement a global exchange of information among the population. Each frog has a specific location within the search space (X^i). The above vector represents a meme with various memotypes as decision variables (N_{VD}). Each memotype identifies the discrete value of each decision variable.

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

The interchange 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 early convergence and corresponding 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 variations 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:

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

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. This function is maximizing the average probability for each location is taken with its corresponding contextual information. For accomplish this, the considered P is split into m number of different memeplexes. Each memeplex contains n frogs and can be assumed as a different culture in which a local search is executed. Finally, the frogs are forwarded to different memeplexes based on their cost function. The global fitness is represeted as X_g and fiind the best and worst solutions for every memeplex are represented as X_b and X_w correspondingly.

Then, each memeplex is divided into sub-memeplex that represents the number of frogs entering memetic evolution.. 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 statements, L_i is the change in frog location, δ denotes a random number between 0 and 1, $X_{w,0}$ is the current position of the frog, $X_{w,1}$ is the new position of the frog and L_{max} is the maximum allowed a change in a frog's position. If the evolution produces a better frog. This frog replaces the worst frog or else, X_b is replaced by X_g in (3) . The process is continued till evolution completes. 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 procedure is repeated for a particular number of iterations i.e., N_S within each sub-memeplex. After that, the local searching in each sub-memeplex is over, 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. It will 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 universal minimum. The leaping and shuffling processes are continued until convergence is fulfilled. Thus, the optimal solution is obtained based on this algorithm to set IMC-TEM efficiently.

4. RESULTS AND DISCUSSIONS

In this area, the results obtained from our experiments which are conducted on two application tasks named social link prediction and location recommendation. Both proposed method and existing are illustrated. In this experiment, we used three most accepted geographically oriented social networking datasets such as $Foursquare_S$, $Foursquare_L$ and $Gowalla$. Our selected datasets contains check-in records in the format of (User ID, Location Category, Timestamp, Location ID, City). Among these datasets, only $Foursquare_S$ and $Gowalla$ are contains the social connection links between users. Moreover, $Foursquare_L$ and $Gowalla$ are utilized for location based recommendation purpose. In addition this, the $Foursquare_S$ and $Gowalla$ are utilized for link prediction process. The comprehensive information of these three datasets we used is given in Table 1.

Table 1: Datasets Analysis

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 based on Location Recommendation

The efficiency of location recommendation includes both time-aware location recommendation and general location recommendation is evaluated by considering home-city recommendation settings. Consider a given user, the respective home-city is identified. The city with the most number of occurrences in the datasets particularly based on as per the check-in records.

The training and testing datasets are constructed by splitting the data based on the trajectories. Initially, 20% trajectories are with only found 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. Next step, 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. For a given n is the rank of the target location inside this list and the optimum value of result relates to the case where it precedes all of the additional locations ($n = 1$). Then, a top-k recommendation list is formed by collecting the top k is ranked locations from the list of elements.

Here, hit_k is denoted for a single test case in the datasets. It is as either the value is 1 if the objective location is located in the top k results, otherwise, the value is 0. The overall $Recall_k$ denotes to the ratio of the hits based on all the test based check-in records from datasets.

$$Recall_k = \frac{Number\ of\ hit_k}{Number\ of\ all\ cases} \quad (5)$$

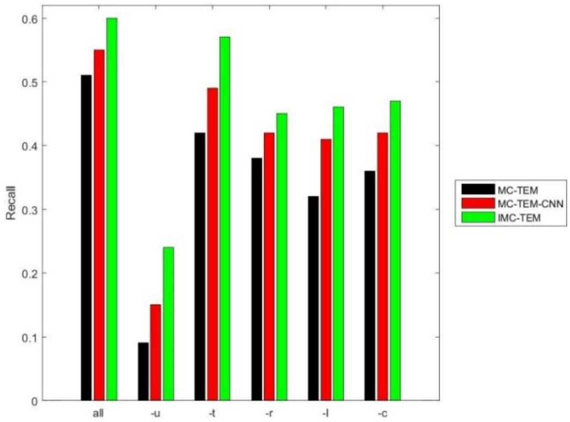

Figure 1: Comparison chart on General Location Based Recommendation

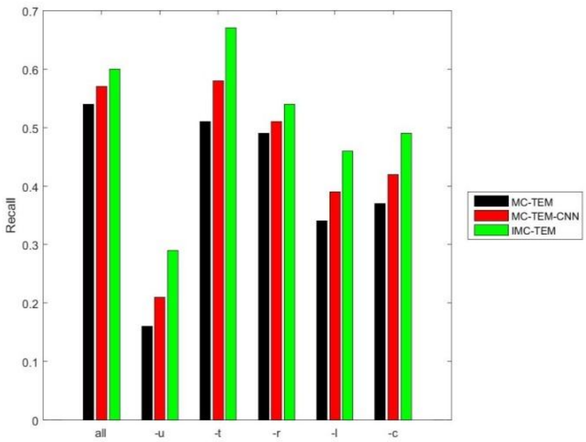

Figure 2: Comparison chart on Time-aware Location Recommendation

Figure 1 and Figure 2 shows that the impacts of contextual features for generalised and time-aware location recommendation on the data set of $Foursquare_L$. 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-context of the user is the most significant factor to be considered since the active task is essentially a custom-made recommendation issue where the preference of users plays the important part 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.

Social Link Prediction Analysis

The objective of this analysis is predicting the existence of a social link between a pair of users. This prediction based on only the analysis of their trajectory data. All the trajectory information for a user is assumed as those are available for unsupervised feature extraction. Also, the connection of links are split into training and testing the dataset. The pairs of users without links are generated for both training and testing datasets by choosing non-linked user pairs randomly with the ratio of 1:1 compared to the social connection links.

Consider P_T is collection of all pairs of users with actual friend links and P_R is the number of all pairs of users recognized 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:

$$Recall = \frac{|P_T \cap P_R|}{|P_T|} \quad (6)$$

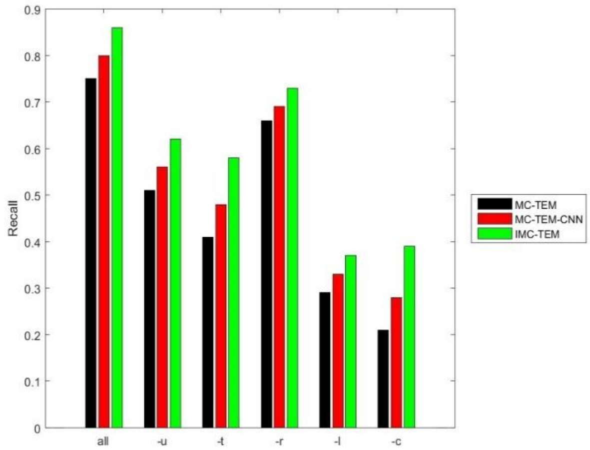

Figure 3: Comparison chart on Social Link Prediction

Figure 3 shows that the impacts of contextual features for social link prediction on the $Foursquare_S$ dataset.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

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, '-' designates that the matching 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.

5. CONCLUSION

In this article, an "Improved Multi-Context Trajectory Embedding Model (IMC-TEM)" is proposed for discovering the contexts in a organized 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 effectively. Also, all the contextual information is represented in the equivalent embedding space for analyzing the association among different contexts with less complexity. At the final stage, the experiments are done 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 *Proc. 1st ACM conf. Online Soc. Netw.*, 2013, pp. 3-14.
2. W. Mathew, R. Raposo and B. Martins, "Predicting future locations with hidden Markov models," in *Proc. 2012 ACM conf. Ubiquitous Comput.*, pp. 911-918.
3. Q. Yuan, G. Cong and C. Y. Lin, "COM: a generative model for group recommendation," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2014, pp. 163-172.
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 *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2013, pp. 605-613.
5. E. Cho, S. A. Myers and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2011, pp. 1082-1090.
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 Trans. Knowl. Data Eng.*, vol. *28, no.* 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," *Int. J. Comput. Sci. Mob. Appl., vol. 5, no.* 9, pp. 1-9, 2017.
8. J. D. Mazimpaka and S. Timpf, "Trajectory data mining: a review of methods and applications," *J. Spat. Inf. Sci.*, vol. *2016, no.*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 *Proc. 34th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, 2011, pp. 325-334.
10. A. Karatzoglou, X. Amatriain, L. Baltrunas and N. Oliver, "Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering," in *Proc. 4th ACM Conf. Recomm. Syst.*, 2010, pp. 79-86.
11. F. Wong, S. Lee and Q. Wong, "Points of interest recommendation based on context-aware," *Int. J. Hybrid Inf. Technol.*, vol. *8, no.* 3, pp. 55-62, 2015.
12. H. Bagci and P. Karagoz, "Context-aware location recommendation by using a random walk-based approach," *Knowl. Inf. Syst.*, vol. *47, no.* 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 *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2011, pp. 1100-1108.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication