REAL TIME EVENT LOCATION DETECTION BASED MOBILITY PATTERN MODELLING FOR SOCIAL MEDIA USER MOBILITY ANALYSIS

AMJAD JUMAAH FRHAN

PhD Student, Department of Telecommunication and Information Technology,

University Politehnica of Bucharest, Romania.

Abstract:-Detection of important events in a region through social media has been a recent development with scope for multiple applications. One of the applications is the analysis of the mobility of the users in the region to adapt the energy and resources as per their movement. For this purpose, the travel or movement behaviour of the social media users is analysed through their posts and related messages. An efficient method of mobility pattern modelling named as Event Location based Mobility Pattern Modelling WebClickviz (ELMPM-WebClickviz) is proposed in this paper based on the event location detection. Initially the social media data are collected and pre-processed. Then the geographical as well as temporal information are extracted along with the time and distance parameters. Then the pattern modelling is initiated using Sequential Hierarchical pattern clustering which detects the continuous events from the user data along with the location of occurrence. Based on these results, the mobility behaviour can be modelled with higher accuracy. The evaluation results prove that the proposed model provides efficient mobility patter modelling to be utilized for the organizations and official concerns in fulfilling the resources and needs of the users of that location.

Key-Words: Social media, Travel behaviour, Event location detection, Mobility pattern modelling, WebClickviz, Sequential Hierarchical pattern clustering

1. Introduction

Handheld and wearable devices have turned out to be instrumental devices for the greater part of our day by day assignments. These devices have been consistently improved with more exact situating sensors, for example, GPS, which have permitted gathering a lot of high-determination digital follows and have facilitated the advancement of the mobility mining discipline [1]. Such teach concentrates on giving knowledge into the fundamental spatio-temporal directions of the assembled follows. Subsequently, creative area based administrations have been produced like personal advertisement campaigns [2] or pervasive navigation systems [3]. In the meantime, social networking has turned into an exceptionally prominent action in most created and creating social orders enabling individuals to remain socially associated with their companions, relatives and partners in a simple way. The expectation of the mobility of a population has turned into a vital point in the field of asset sparing. Specifically, one of the territories in which the expectation of mobility has had the best effect is in the vehicle segment, which speaks to one of the significant energy utilization of a population [4]. The main aim is to understand how the population of the cities move in order to use the transport resources efficiently based on real demands of the users, knowing where the population flows, at what time and its intensity.

Although this subject has been examined extensively, those techniques have a few limitations. They concentrate on extracting general mobility data identified with a specific urban region or region without recognizing the time of the day in which the data was created (time slots). Therefore, existing arrangements don't consider the connection between the snapshot of the day when social-media documents are posted and its related spatial place. This missing data could give a worldwide vision of the movement of a population along the day. In this way, these works are not taking full progress of social-media datasets. Consequently, the outcomes are not as exact as they could be. In addition, most works don't consider the activity level of the users inside each distinguished region [5]. This permits setting up a connection between classes of users and their movement over the region under investigation. These limitations have created a void in the mobility pattern modelling.

In this paper, an efficient user mobility model based on the event location detection is developed using the WebClickviz model. WebClickviz model [6] has shown better performance for user clickstream analysis which was extended to the use in wireless multimedia sensor network [7] based application in the successive researches. This proposed model named as ELMPM-WebClickviz, after pre-processing of the social data, extracts the geographical and temporal features of the data. It also extracts the time and distance measures followed by which the clustering is performed for event and its location detection. Detection of series of events occurred in the daily life of users analysed as a collateral process helps in identifying the moving pattern of them. Thus the proposed model improves the resource usage efficiently based on time and intensity of travel. The remainder of the article is organized as follows: Section 2 discusses some the most related research works. The proposed methodologies are discussed in Section 3. Section 4 focuses on the performance analysis results of ELMPM-WebClickviz. Finally, Section 5 makes a conclusion about the proposed model while also suggesting future directions of research.

2. Related Works

Travel behaviour studies based on social media is emerging. The aggregated discoveries coincide with accord by customary travel overview. For example, Cheng et al [8] found that the impressions (registration) left by the social media users take after a Levy Flight mobility design. Additionally, different investigations demonstrated that social media is a significant supplement to conventional travel conduct thinks about. For example, Rashidi et al [9] checked on the present cutting edge techniques for social media thinks about and presumed that it had a colossal capability of enhancing our knowledge in activity interest conduct; Zhu et al [10] explored the location-based social networks and accomplished more than 75% accuracy in anticipating travel purposes joining with the customary travel overview; Zheng et al [11] even joined the social media with skimming sensors and occurrence answer to anticipate the human mobility and controlled the movement in both the physical and cyber spaces.

Various investigations have been proposed to show user mobility practices. Brockmann et al. [12] discover human mobility conduct can be approximated by a constant time irregular walk demonstrate with long-tail conveyances. Gonzalez et al. [13] find that users occasionally come back to a couple of beforehand went to locations, and the mobility of every user can be demonstrated by a stochastic procedure focused at a settled point. Cho et al. [14] watch that the mobility of every user is focused at a few regions, and the likelihood that a user remains at a region is affected by time. They propose a generative model, Periodic Mobility Model (PMM), which predicts a user's location by assessing the regions in which an objective user in all probability remains at an objective time. Tarasov et al. [15] take after this paper and model a region by radiation display [16]. Isaacman et al. [17] look at spatiotemporal dispersions of individuals' call records information to consider individuals' mobility at a metropolitan scale. Deb et al. [18] and Zhang et al. [19] utilize the Hidden Markov Model to remove inactive semantic locations. Wang et al. [20] propose a half and half mobility display that consolidates normality and similarity of human mobility. Jiang et al. display human dynamic practices with tensor technique [21] and Minimum Description Length standard [22]. None of these investigations are equipped for recognizing genuine periods. Other important researches related to social media user mobility detection in literature are given in [23], [24], [25]. None of these studies are capable of detecting the exact time slots of the user movements. Hence the development of the proposed mobility pattern modelling approach is required.

3. Event Location Based Mobility Pattern Modeling WEBCLICKVIZ

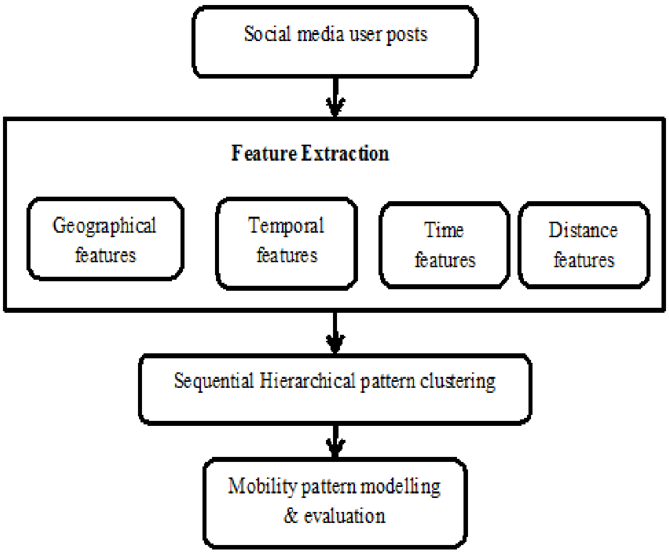

Figure.1. Proposed ELMPM-WebClickviz model

The textual data from the social media is collected and it is pre-processed followed by the transformation into suitable format. Then the features are extracted using extraction techniques. For time features extraction, utilize discrete Fourier transform to locate the spectral with the highest power, and utilize its relating period as the outcome. At first compute the auto-correlation of the time series [26], and afterward utilize FFT to choose the period with the most elevated power as the outcome. At that point apply Fourier Transform to the time arrangement and distinguish the scope of applicant periods relating to the best power. At that point for every period run given by the periodogram, test whether there is a peak value inside it. On the off chance that there is a peak, restore the location of the peak as the outcome. At that point ascertain the disparity score for every competitor period, which measures to which surviving the records focus on an arrangement of time points on the off chance that portion the records based on the period and overlay them together. The period with the biggest error score is yield as the outcome. To separate the temporal features, viability of mutually displaying spatial and temporal information, another standard which initially extricates regions and after that recognizes the period is utilized. For location feature extraction, periodic mobility model [14] is utilized based on the fact that it is the most important technique that can anticipate location whenever on GPS organizes information without requiring social information. In periodic mobility model, two GMMs are worked for weekdays and ends of the week individually rather than one GMM for every day, in light of the fact that the execution of the latter approach provides worst outcomes. Thus the features can be extracted efficiently for the given social data and then the clustering technique is used.

The proposed model uses Sequential Hierarchical Pattern Clustering for sequentially updating a hierarchical tree which in turn enables the detection of event location. An underlying hierarchical tree is developed by figuring all pairwise similarities between a small subset of the data, and after that passing these to the Single-Round-MC-UPGMA [27]. Following the development of the underlying tree utilizing Single-Round-MC-UPGMA, the rest of the data is consecutively prepared. Whenever a new pattern (x_i) lands for clustering, its similarity distance d to the root of the current hierarchical tree is registered. If d is more prominent than a predefined limit (θ), a new root is made having the present pattern (x) and the past root as its children, and as a result the depth of the tree increments by one. The value of the new root is assigned with the arithmetic mean of all the leaf nodes. In any case, if d is not exactly θ, the closest offspring of the present node is recovered. If the distance of x_i to this child node is also less than θ then repeating finding the nearest child node until either the distance to the present node is higher than θ or achieve a leaf node. In both of the two cases, x_i is made as a child to the node under consideration, and x_i is propagated up the tree to refresh its ancestor nodes. By changing θ, one can acquire diverse quantities of clusters at various levels of granularity. The leaf nodes of the hierarchical tree are the input patterns and each intermediate node contains the arithmetic mean of every leaf node it represents. Along these lines, it is not required to traverse the whole tree amid the update procedure. The overall process performed in the Sequential Hierarchical Pattern Clustering is given in the algorithm.

3.1Algorithm: Sequential Hierarchical Pattern Clustering:

Input: Root of initial tree R-node; the new pattern N-node, novelty threshold θ
Begin
Compute similarity distance $\mathrm{Sim}-\mathrm{Dist}$ between R-node and N-node
If ($\mathrm{Sim}-\mathrm{Dist} \leq \theta$) then
    $\text{Child node} \leftarrow \text{getchild node}\ (\mathrm{R}-\text{node})$
    If ($\text{Child node} == \text{null}$) then
        Make N-node as Child node
        Update nodes
    Else
      R-node has child nodes
     Nearest node
        $\leftarrow \min \mathrm{Sim}-\mathrm{Dist}(\text{Child node}, \mathrm{N}-\text{node})$
        If ($\text{Nearest node} \leq \theta$) then
        $\mathrm{R}-\text{node} \leftarrow \text{Nearest node}$
        Else
        Make N-node as Child node
        Update nodes
        End if
    End if
Else
    Make N-node as Child node using new root
End if
End

4. ELMPM-WEBCLICKVIZ Performance Analysis

In the experiments, a social media dataset from the FACEBOOK website is employed as the experimental data. There were 5999 records in the raw file from a period of January 2014 to December 2016. After data cleaning, there were 3222 records left from 243 user sessions. There were 8 different kinds of activities during the user sessions in these data. As the primary work of WebClickviz is to visualize the data, the process begins from ClickStream data visualization, followed by event detection and mobility pattern modelling. The performance of the ELMPM-WebClickviz is compared with that of the WebClickviz, Pattern WebClickviz, Social pattern-WebClickviz and Event WebClickviz, to showcase its efficiency.

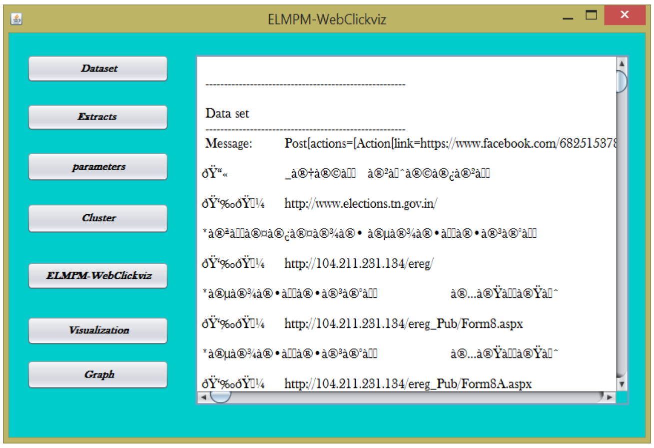

Figure.2. Dataset representation

Figure 2 shows the loading process of the data into the proposed model. The data are represented as texts. Figure 3 shows the extraction results after extracting the geographical, temporal and time features. It can be seen that the proposed approach effectively extracts the features from the given dataset.

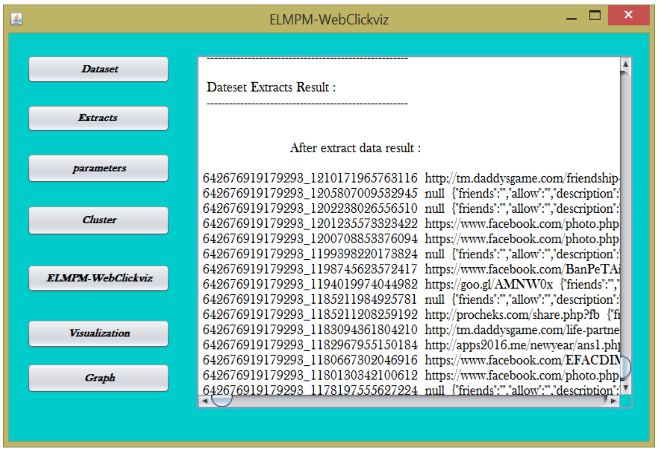

Figure.3. Feature extraction

Figure 4 shows the post detection from the social media carried out using feature extraction. This method is much efficient for the social media data because of its ability extract the different kinds of features especially the geographical and the time features.

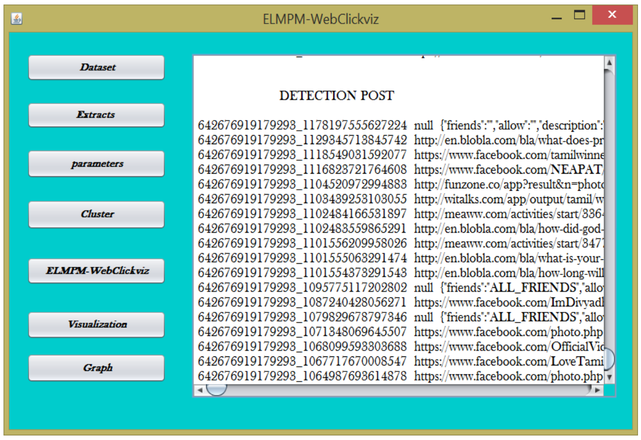

Figure.4. Feature Posts detection

Figure 5a) & b) shows the clustering results. It can be seen that the proposed approach has segmented the posts based on the features and utilized them to predict the possible event occurred. The feature posts are those that have all the features available for the pattern modelling. The proposed approach efficiently detects these posts unlike other existing methods.

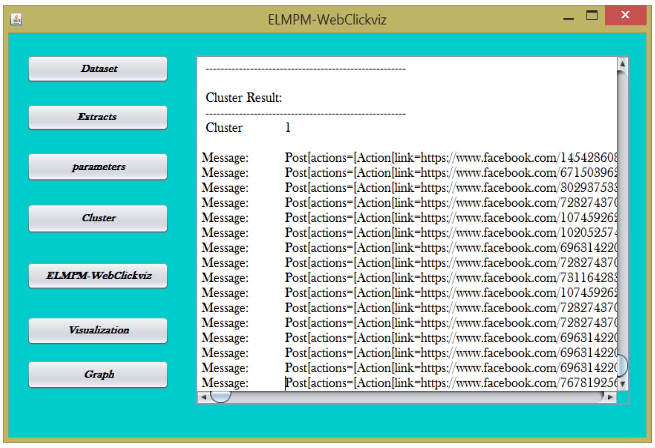

Figure.5a). Clustering results

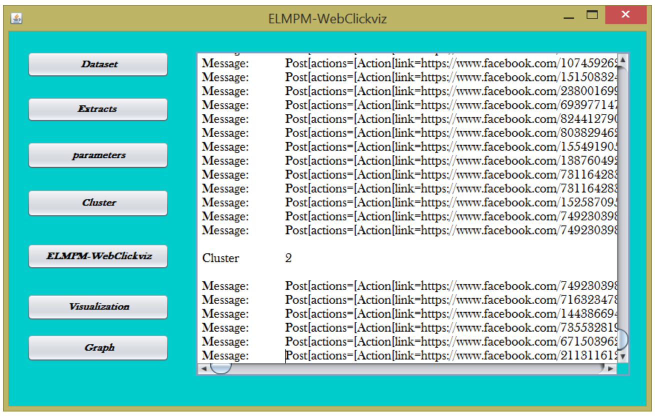

Figure.5b). Clustering results

Based on these results the different events are detected and interlinked to form the mobility model. The refined list of events identified by the proposed model is shown in Figure.6. The proposed model detects the events which has the highest probability of occurrence which improves the overall accuracy of performance.

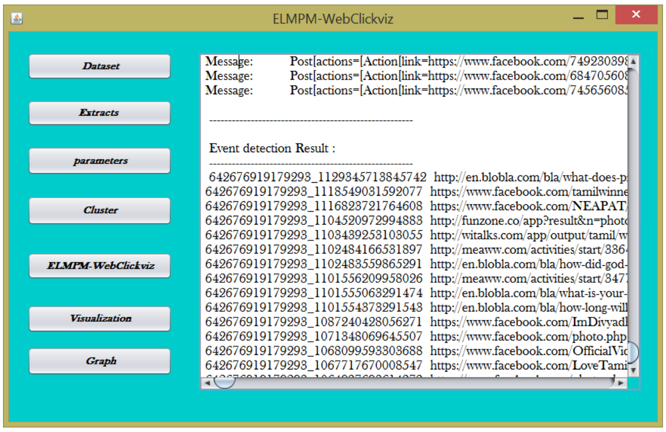

Figure.6. Refined event detection

Figure 7 shows the results of ELMPM-WebClickviz model. These results indicate that using the events identified and clustered, a mobility pattern can be laid out. The events are clustered in the order of increasing probability of occurrence.

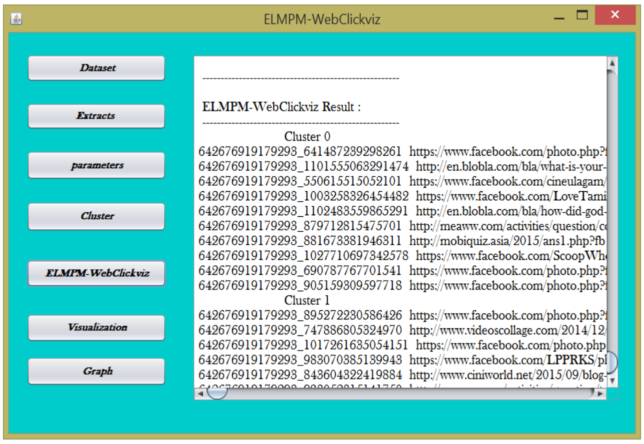

Figure.7. ELMPM-WebClickviz results

Figure 8 shows the final visualization results which express the time duration of each activity of event occurrence. Combining this information will be helpful in finalizing the pattern of the users individually as well as collectively. From this figure, it can be seen that the use of event location based model improves the mobility prediction and offers immense applications.

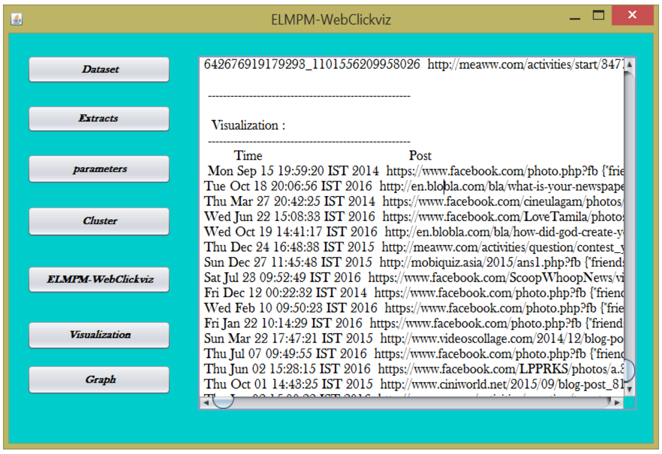

Figure.8. final visualization result

Performance comparison

Figure 9 shows the clickstream value comparison of the proposed ELMPM-WebClickviz with the other WebClickviz based models. It is see that the proposed model improves the detection process by identifying the larger actions in the same dataset and increases the mobility pattern modelling performance.

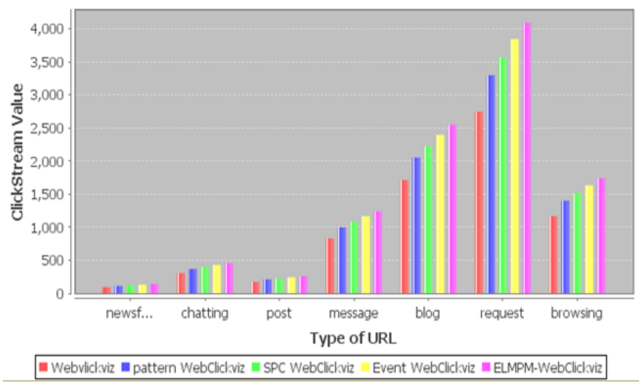

Figure.9. Clickstream value

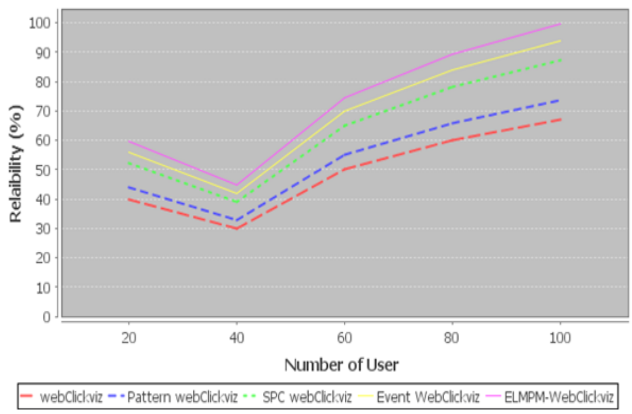

Figure.10. Reliability

Similar to clickstream value, the reliability and consistency score comparisons are shown in Figure 10 and 11 respectively. It is found that the proposed ELMPM-WebClickviz has better values in both the cases, thus justifying its performance efficiency. It is due to the fact that the involvement of the geographical features considerably improves the reliability and consistency of the event detection.

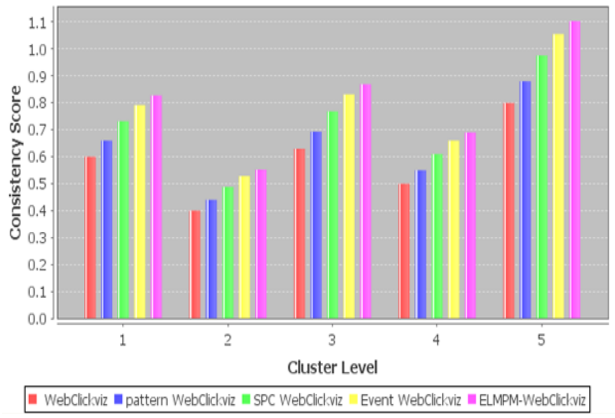

Figure.11. Consistency score

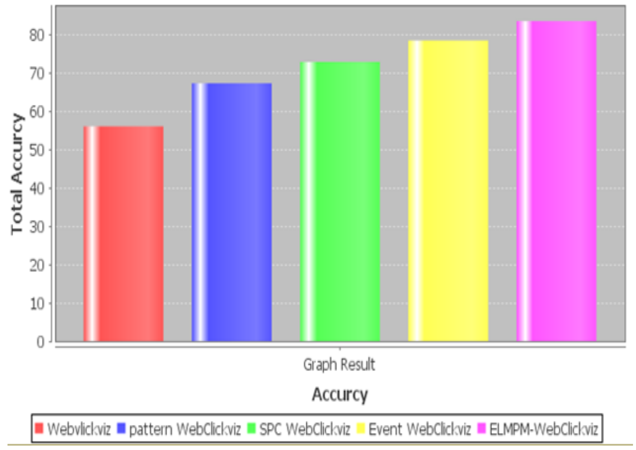

Figure.12. a) Accuracy

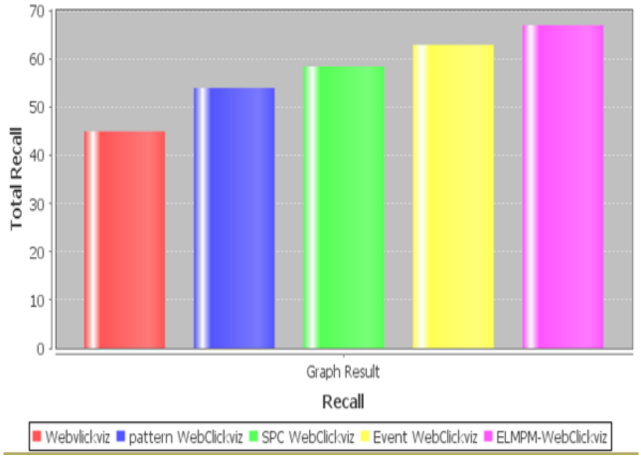

Figure.12. b) Recall

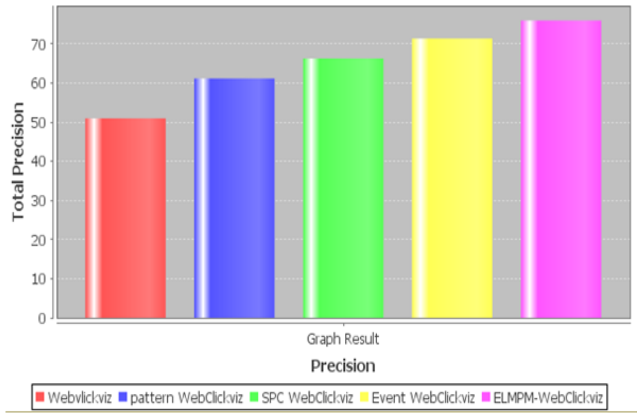

Figure.12. c) Precision

Figure 12 shows the a) accuracy, b) recall and c) precision comparison of the ELMPM-WebClickviz with the other WebClickviz based models. ELMPM-WebClickviz outperforms the other models with higher values of accuracy, recall and precision. This is due to the novel process of the proposed approach in extracting the features from the data corpus and also due to inclusion of geographic and time features for efficient clustering.

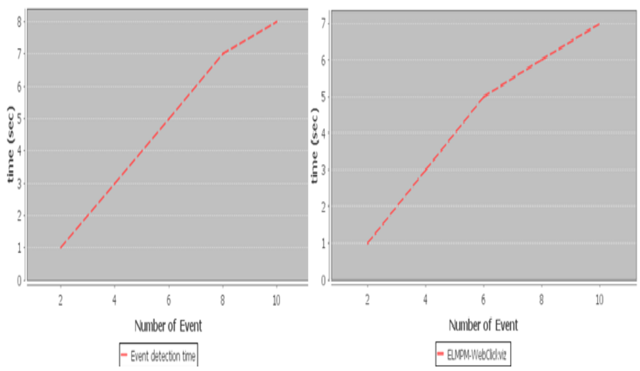

Figure.13. Event detection time

Figure 13 shows the event detection time of the ELMPM-WebClickviz and Event WebClickviz models. It is evident that the ELMPM-WebClickviz model has less time for detecting the events than the other model due to the inclusion of time features as a predominant feature. Thus the proposed approach is not only accurate but it is also very faster in performance.

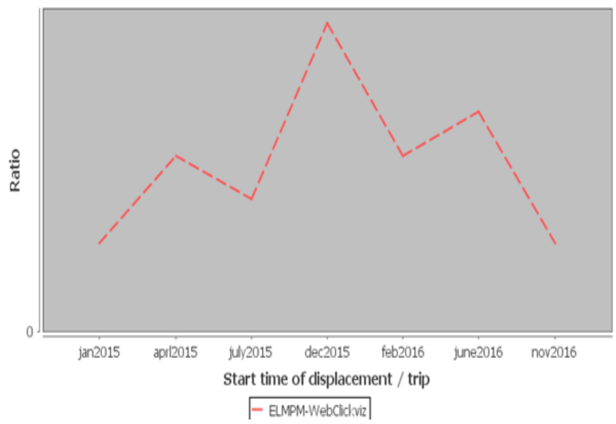

Figure.14. Time displacement vs. user ratio

Figure 14 shows the time displacement/trip vs. user ratio. The proposed ELMPM-WebClickviz has efficient values due to the efficient clustering. The mobility pattern can be drawn out from this result with higher accuracy.

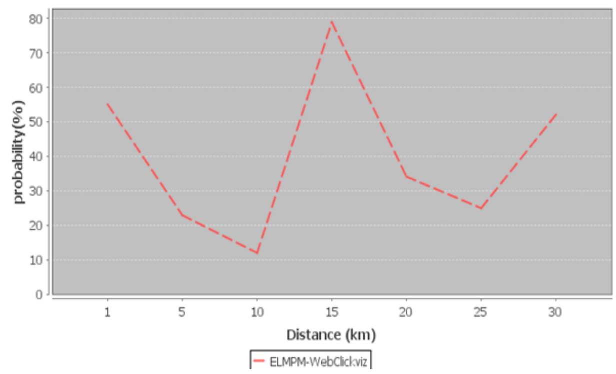

Figure.15. Distance vs. probability

Figure 15 shows the distance vs. probability graph. The proposed ELMPM-WebClickviz has shown immense responses towards the detection of users through various distances. Though the results seem fluctuating it is the actual route of travel by the users at a specified duration.

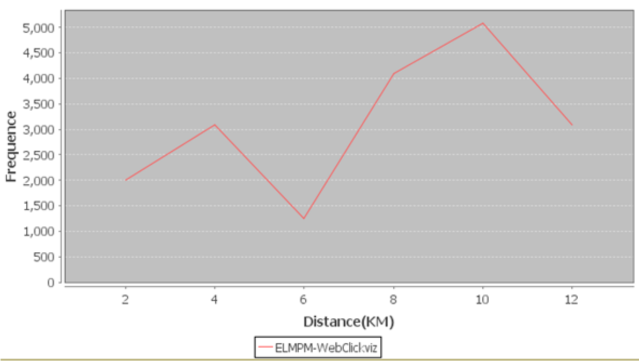

Figure.16. Distance vs. frequency of occurrence

Figure 16 shows the distance vs. freuency of occurence. The proposed ELMPM-WebClickviz detects the users' locations based on their posts in social media. Using the posts, the users' favourite or routine locations in a day is plotted. The frequency of occurrence denotes a place like home, school or favourite hanging out location visited more often by the user. Thus the proposed model fulfils its purpose of establishment through efficient performance.

5. CONCLUSION

In this paper, an efficient user mobility pattern modelling approach called ELMPM-WebClickviz has been developed and evaluated. The use of Sequential Hierarchical pattern clustering approach increases the clustering performance of events. The event detection performance is improved by the inclusion of the geographical, temporal, distance and time features. By the use of these features the vents are detected and the continuous event detection of individual users and community users helps in modelling their mobility pattern. Thus the proposed model has significant place in the user mobility pattern modelling with higher values of performance values. In future, it has been planned to use the extracted mobility patterns to improve other important location-based applications, such as location recommendation and local event detection. Similarly the enrichment of the obtained results with the use of other semantic features will also be considered.

References

[1] Yan, Z., & Chakraborty, D. (2014). Semantics in mobile sensing. Synthesis Lectures on the Semantic Web: Theory and Technology, 4(1), 1-143.

[2] Wu, F., Li, Z., Lee, W. C., Wang, H., & Huang, Z. (2015). Semantic annotation of mobility data using social media. In Proceedings of the 24th International Conference on World Wide Web (pp. 1253-1263). International World Wide Web Conferences Steering Committee.

[3] Gabrielli, L., Rinzivillo, S., Ronzano, F., & Villatoro, D. (2014). From tweets to semantic trajectories: mining anomalous urban mobility patterns. In Citizen in Sensor Networks (pp. 26-35). Springer, Cham.

[4] Hasan, S., Zhan, X., & Ukkusuri, S. V. (2013). Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In Proceedings of the 2nd ACM SIGKDD international workshop on urban computing (p. 6). ACM.

[5] Thilakarathna, K., Seneviratne, S., Gupta, K., Kaafar, M. A., & Seneviratne, A. (2017). A deep dive into location-based communities in social discovery networks. Computer Communications, 100, 78-90.

[6] Frhan, A. J. (2017). Website Clickstream Data Visualization Using Improved Markov Chain Modelling In Apache Flume. In MATEC Web of Conferences (Vol. 125, p. 04025). EDP Sciences.

[7] Al-Ariki, H. D. E., & Swamy, M. S. (2017). A survey and analysis of multipath routing protocols in wireless multimedia sensor networks. Wireless Networks, 23(6), 1823-1835.

[8] Cheng, Z., Caverlee, J., Lee, K., & Sui, D. Z. (2011). Exploring millions of footprints in location sharing services. ICWSM, 2011, 81-88.

[9] Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S., & Waller, T. S. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. Transportation Research Part C: Emerging Technologies, 75, 197-211.

[10] Zhu, Z., Blanke, U., & Tröster, G. (2014). Inferring travel purpose from crowd-augmented human mobility data. In Proceedings of the First International Conference on IoT in Urban Space(pp. 44-49). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[11] Zheng, X., Chen, W., Wang, P., Shen, D., Chen, S., Wang, X., & Yang, L. (2016). Big data for social transportation. IEEE Transactions on Intelligent Transportation Systems, 17(3), 620-630.

[12] Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. Nature, 439(7075), 462-465.

[13] Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. Nature, 453(7196), 779-782.

[14] Cho, E., Myers, S. A., & Leskovec, J. (2011). 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). ACM.

[15] Tarasov, A., Kling, F., & Pozdnoukhov, A. (2013). Prediction of user location using the radiation model and social check-ins. In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing (p. 8). ACM.

[16] Simini, F., González, M. C., Maritan, A., & Barabási, A. L. (2012). A universal model for mobility and migration patterns. Nature, 484(7392), 96-100.

[17] Isaacman, S., Becker, R., Cáceres, R., Martonosi, M., Rowland, J., Varshavsky, A., & Willinger, W. (2012). Human mobility modeling at metropolitan scales. In Proceedings of the 10th international conference on Mobile systems, applications, and services (pp. 239-252).

[18] Deb, B., & Basu, P. (2015). Discovering latent semantic structure in human mobility traces. In European Conference on Wireless Sensor Networks (pp. 84-103). Springer, Cham.

[19] Zhang, C., Zhang, K., Yuan, Q., Zhang, L., Hanratty, T., & Han, J. (2016). Gmove: Group-level mobility modeling using geo-tagged social media. In KDD: proceedings. International Conference on Knowledge Discovery & Data Mining (Vol. 2016, p. 1305). NIH Public Access.

[20] Wang, Y., Yuan, N. J., Lian, D., Xu, L., Xie, X., Chen, E., & Rui, Y. (2015). Regularity and conformity: Location prediction using heterogeneous mobility data. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1275-1284). ACM.

[21] Jiang, M., Cui, P., Wang, F., Xu, X., Zhu, W., & Yang, S. (2014). Fema: flexible evolutionary multi-faceted analysis for dynamic behavioral pattern discovery. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1186-1195). ACM.

[22] Jiang, M., Faloutsos, C., & Han, J. (2016). Catchtartan: Representing and summarizing dynamic multicontextual behaviors. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 945-954). ACM.

[23] Yuan, Q., Zhang, W., Zhang, C., Geng, X., Cong, G., & Han, J. (2017, February). Pred: Periodic region detection for mobility modeling of social media users. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (pp. 263-272). ACM.

[24] Zhang, Z., He, Q., & Zhu, S. (2017). Potentials of using social media to infer the longitudinal travel behavior: A sequential model-based clustering method. Transportation Research Part C: Emerging Technologies, 85, 396-414.

[25] Cuenca-Jara, J., Terroso-Saenz, F., Valdes-Vela, M., Gonzalez-Vidal, A., & Skarmeta, A. F. (2017). Human mobility analysis based on social media and fuzzy clustering. In Global Internet of Things Summit (GIoTS), 2017 (pp. 1-6). IEEE.

[26] Li, Z., Ding, B., Han, J., Kays, R., & Nye, P. (2010). Mining periodic behaviors for moving objects. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1099-1108). ACM.

[27] Loewenstein, Y., Portugaly, E., Fromer, M., & Linial, M. (2008). Efficient algorithms for accurate hierarchical clustering of huge datasets: tackling the entire protein space. Bioinformatics, 24(13), i41-i49.

[28]