How to Cite:
Kumar, G. K., Gopalachari, M. V., Mishra, R., Jayaram, D., Rakesh, S., & Rani, D. M. (2022). Enabling effective location-based services for road networks using spatial mining. *International Journal of Health Sciences*, *6*(S4), 5174–5188.
https://doi.org/10.53730/ijhs.v6nS4.9313

Enabling effective location-based services for road networks using spatial mining

G Kiran Kumar
Assistant Professor, CSE Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

M Venu Gopalachari
Associate Professor, IT Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India
Corresponding author email: mvenugopalachari_it@cbit.ac.in

Rupesh Mishra
Assistant Professor, CSE Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

D Jayaram
Assistant Professor, IT Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

S Rakesh
Assistant Professor, IT Department, Chaitanya Bharathi Institute of Technology, Hyderabad, India

D. Malathi Rani
Assistant Professor, ECE Department, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, India

***Abstract*---**A co-location pattern represents a subset of Boolean spatial attributes whose instances are located in a close geographic space. These patterns are important for location-based services. There are many methods for co-location pattern mining where the distance between the events in close geographic proximity is calculated using a straight-line distance called Euclidean distance. Since most of the real-time tasks are bounded to the road networks, the results computed using Euclidean distance is not appropriate. So to compute co-location patterns involving network we define a model where initially a network model is defined and the neighbourhood is obtained by using network distance. By comparing this approach with the previous Euclidean approach, the results

Manuscript submitted: 27 March 2022, Manuscript revised: 18 May 2022, Accepted for publication: 9 June 2022

obtained for co-location patterns on a road network are accurate. Our experimental results for synthetic and real data show that the proposed approach is efficient and accurate for identifying co-location patterns involving network entities.

***Keywords*---**spatial data, co-location patterns, road network distance, decision making, SVM, decision tree, random forest, Naïve Bayes.

Introduction

Co-location Pattern for Mining Spatial data (CPMS) is used to identify close proximity objects based on the geographical data Wang (2018). Complex associations are generally present in geographical data and knowledge extraction from both spatial and non-spatial attributes can be performed through CPMS. Location-based service (LBS) is a mobile service which uses wireless internet. These are capable of providing geographic information based on the current location. The information about surroundings with respect to a given position has become important to solve many real-time issues.

Any data is said to be spatial if it contains location entities. Extracting the close proximity geographic objects is possible through CPMS. The closeness between the objects can be calculated using various distance metrics like Euclidean distance and path distance Yao (2018). Usually, the distance between the objects is calculated using straight line distance called Euclidean distance and is ill-suited because human activities are mainly constrained to road networks. So the distance measure used to calculate the road networks is the network distance.

Co-location rule mining has got many applications across domains like location-sensitive advertisements, genetics, transportation, tourism, ecology, weather prediction, and crime prediction Lu (2018). For example, a mobile service provider who has data of service patterns used by geographically nearby users and can provide location sensitive offers which boost his sales. Of late, a genetic field of biology is using the co-location mining for predicting the structures of DNA.

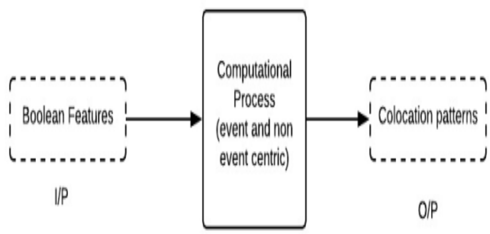

Figure 1 The high-level view of Spatial Mining

Other examples are the prediction of rainfall in a particular area based on climatic conditions of that area and other neighbouring areas of it and prediction of a possible crime in an area based on a series of events occurring in that area over a period of time. The high-level flow of CPMS is shown in Fig. 1.

The following terminology has been used throughout this paper.

- Network: A network is a weighted graph with a set of nodes N and a set of edges E. Represented as G=(N, E).
- Network Edge: Any two nodes which have direct communication can be expressed as network edge. It can be created between the start node and the ending node. Multiple nodes can even be configured on the network edge based on the object distance from the starting node.
- Network object: A network object is expressed as (d, k) where d is the nearest edge and k is the distance from the starting node.
- Network distance: Network distance is defined as the shortest path distance from one object to another. For example, if we consider the objects which lie on the same edge then the distance between the two objects can be calculated and identified as the shortest distance between the objects.
- Neighborhood: All the objects within the defined threshold are treated as the neighborhood to a target object. This measure is required to find all the co-located objects in the geographical area. Dijkstra's algorithm is one of the most popular methods for finding the shortest path. This gives the optimal solution for the given dataset.

Related Work

Based on the extensive literature, we prepared a taxonomy for identifying co-location patterns (Fig. 2.). Multiple researchers have worked on spatial co-location mining. Many of these past works are based on the Euclidean distance Yu (2017), Huang (2004), Cai (2018), Yoo (2004). However, most of the human activities depend on network distance Okabe (2009), Yamada (2007). The economic issues can be addressed using the network distance Lu (2017), Yu (2017).

In Agrawal (1994), the authors formulated the rules for generating advanced frequent patterns using the support and threshold. The authors in Koperski (1995) introduced the idea of spatial association mining in which they focused on certain spatial objects Flouvat (2015). But, these approaches may not contain the required reference feature. To address this, recent researches focused on CPMS.

CPMS is used for several applications especially in designing, natural investigations, business, and transportation Shekhar (2001). In another study, an object-centric model was proposed in Huang (2006) for co-location pattern mining which is based on clique instances rather than reference feature along with introducing a measure named participation index. But, this method has a lot of instances join operations, which increases the complexity.

To minimize this number of join operations, the authors in Yoo (2004) proposed different approaches. Firstly, they proposed an algorithm based on a plane dividing approach. Even though this approach performs better, still there are noticeable instance join operations. In recent studies, this algorithm is further extended to make it more efficient Flouvat (2015), Philips (2012).

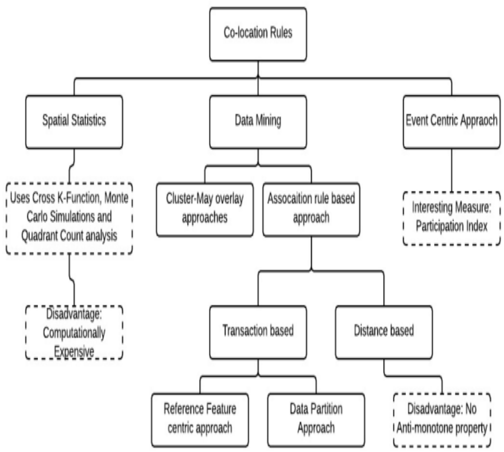

Figure 2. A proposed taxonomy for generating co-location patterns

In this paper, we propose to use a road network distance for CPMS. Instead of using the Dijkstra algorithm, we use A* search algorithm for finding shortest paths which improve the performance. A* algorithm is a widely used approach in path finding and graph traversal techniques.

The fundamental component of spatial data mining is based on how neighborhood relation is defined. Further, there are also many other approaches in the literature to discover the spatial co-location mining like buffer zones Appice (2003), topological connections Santos (2005), kNN algorithm Wan(2009). The utilization of network distance is not a new aspect of GIS. Many researchers Huang (2004), Cai (2018)Xie (2008), used network distance for CPMS but authors in Yu (2015) used a new interesting measure and their results are more convincing and accurate.

A format sheet with the margins and placement guides is available as both Word and PDF files as <format.docx> and <format.pdf>. It contains lines and boxes showing the margins and print areas. If you hold it and your printed page up to the light, you can easily check your margins to see if your print area fits within the space allowed.

Proposed Approach for Identifying Co-Location Patterns

The proposed design for identifying co-location patterns is shown in Fig. 3. In the previous studies, the distance between the two objects is calculated using Euclidean distance. Since most of the human activities are constrained only to the road networks, the Euclidean model is ill-suited. To overcome this drawback, a new distance metric called network distance was used. This distance metric is used to calculate the distance between the two objects through a network. However, there are very few studies which focused on using road network distance for CPMS. We use the existing framework of CPMS and customize it to use participation index as an interesting measure.

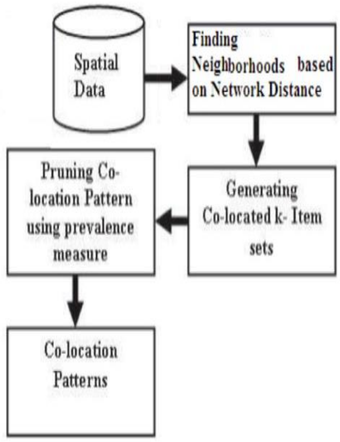

Figure 3 Proposed Architecture for CPMS (Co-location Patterns through Mining spatial data

In this method, the shortest distance between the two objects located in close geographic proximity is considered. The algorithm used for calculating the shortest path is A* search algorithm. In this context initially, the objects located in geographic proximity are identified. Later using the A* search algorithm the shortest path from each object to another object is obtained. Based on the obtained distances a threshold is fixed.

The objects within the given threshold tend to form a pattern called the co-location pattern. Based on the objects, co-location patterns of various sizes will be obtained. This obtained co-location patterns will be useful for users in their trip planning, business decisions etc. Generally, the co-location pattern mining in spatial data follows these steps: Initially, a neighborhood relation graph is built based on the predefined neighborhood. Next, the clique instances (co-location instances) will be generated. Finally, the relevant patterns are obtained (i.e. pruning) by using a threshold. The major variation in our paper is that we focused on creating the neighborhood relation graph based on network distance.

Experimental Results

Two different datasets have been considered in the experimentation, one is a synthetic data and other is a real-time data (Kukatpally area description). This approach analysed on a real data i.e., KPHB colony (Asia's largest colony) data to assess the performance and the details are as follows: In this, we considered all ATMs, shopping malls, hospitals, and banks. The data set consists of 107 banks, 238 ATMs, 49 Hospitals and 18 shopping malls (Fig. 4). We find latitude and longitude for each object and applied the concept of network distance where the equivalent weighted graph is constructed. We applied the A* algorithm for identifying the shortest distance to find co-location patterns.

In urban areas, the services and location of the objects are tied to road networks. So to represent this, we use nodes, edges, network object, network distance, neighborhood, participation ratio, and participation index. Initially, the network distance from each event to another event is calculated and stored. Now threshold is fixed up to certain units. Based on the threshold the co-location patterns based on road networks will be obtained. Based on the threshold the co-location patterns of various objects will be obtained. The participation ratio and participation index are calculated for the obtained co-location patterns. A higher value of participation ratio indicates that the identified co-location patterns are strongly correlated. Fig. 5. illustrates the object representation in the spatial reference system.

If we consider the objects which lie on the same edge then the distance between them is considered as the shortest one. On the other hand, in Fig. 5., objects A1 and D3 lie on different edges the network distance between them can be calculated by using the Dijkstra's algorithm i.e. Network distance (A1, D3) = A1.pos + networkdistance(n7, n7) + D3.pos. Substituting the corresponding values we get 7.1 (3.8+0+3.3). Fig. 6 shows the co-location instances obtained, participation ration and participation index which are obtained after the complete execution of the code. Higher participation index indicates that the co-located patterns occur frequently within the threshold.

The following measures have been considered to access the performance of the system.

- Participation Ratio (PR): Participation ratio can be defined as the ratio of the number of distinct objects that are included in co-location instances to the total number of objects of the type (Equation (1)).
  Participation Ratio = (number of distinct objects) / (total number of objects of the type)
- Participation Index (PI): Participation index can be defined as the minimum values of participation ratio.

We tested the proposed approach for multiple threshold values. Taking 100 meters as the threshold, the results obtained are shown from Fig. 7 to Fig. 11. We extended our approach for other threshold values and the same are shown at the end of the section.

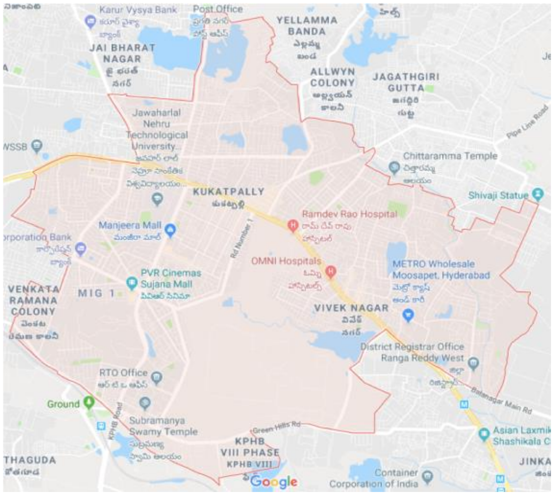

Figure 4 Map overview of the considered Real-Time Data Set

Applying the Proposed Approach on Synthetic Data set

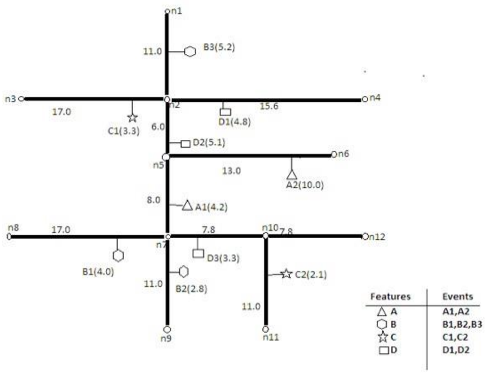

Figure 5 Linear referencing based Network Model

Fig 7 shows that the participation index for ‘C, D' co-location patterns is very high. This indicates that the co-location patterns occur frequently in an area of 9 units. Applying the Proposed Approach on Real-Time Data Set Fig. 8 shows that the co-location patterns for ATMs and Banks is very high, which indicates that ATMs and Banks occur frequently within an area of 100m in Kukatpally area.

Co-location instances with distance threshold=9units					
Network co-location	{A,B}	{A,D}	{B,D}	{C,D}	{A,B,D}
Co-location instances	{A1,B1}	{A1,D2}	{B1,D3}	{C1,D1}	{A1,B1,D3}
	{A1,B2}	{A1,D3}	{B2,D3}	{C1,D2}	
				{C2,D3}	
Participation Ratio	1/2,2/3	1/2,2/3	2/3,1/3	1,1	1/2,1/3,1/3
Participation index	1/2	1/2	1/3	1	1/3

Figure 6 Co-location patterns for sample data set in Fig.1.

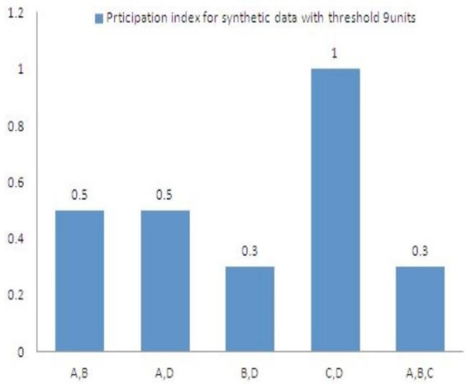

Figure 7 Participation index for co-location patterns in a synthetic dataset with threshold 9 units

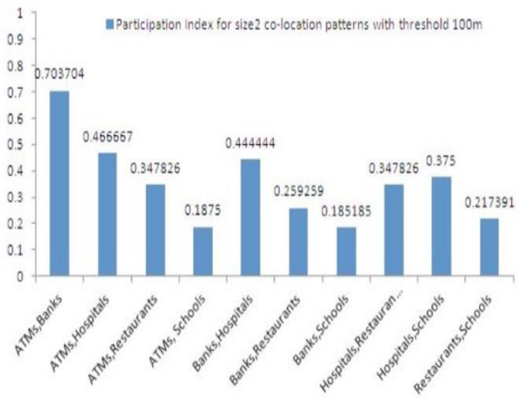

Figure 8 Participation index for size2 co-location patterns with threshold 100m

Fig. 9 shows that the co-location patterns for ATMs, Banks, and Hospitals are very high, which indicates that ATMs, Banks, and Hospitals occur frequently within an area of 100m in Kukatpally area.

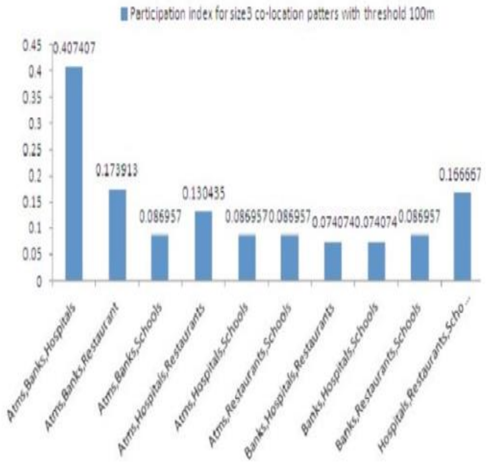

Figure 9 Participation index for size3 co-location patterns with threshold 100m

Fig. 10 shows that the co-location patterns for ATMs, Banks, Restaurants, and Schools is very high, which indicates that ATMs, Banks, Restaurants, and Schools occur frequently within an area of 100m in Kukatpally area. Fig. 11 shows that the co-location patterns for ATMs, Banks, Hospitals, Restaurants, and Schools are very high, which indicates that ATMs, Banks, Hospitals, Restaurants, and Schools occur frequently within an area of 100m in Kukatpally area.

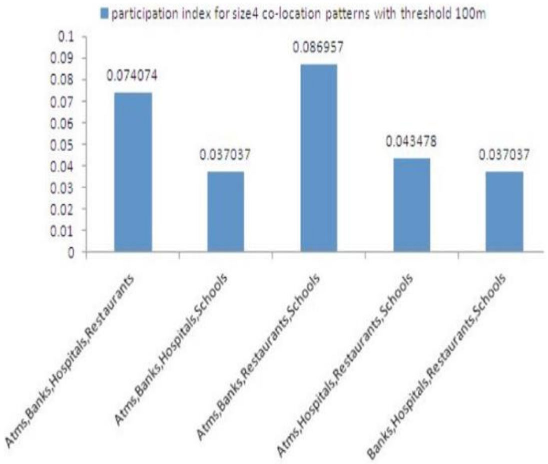

Figure 10 Participation index for size 4 co-location patterns with threshold 100m

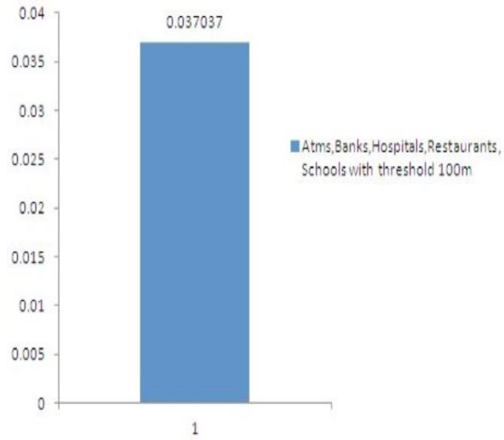

Figure 11 Participation index for size 5 co-location patterns with threshold 100m

We tested the proposed model using multiple spatial data sets with varying thresholds. We observed that when the threshold value is increasing, the probability of occurring the co-location patterns is increasing to a certain limit and later, it was almost constant (Table 1 and Table 2). The co-location patterns obtained for various threshold values are shown in Fig. 12.

Table 1 Co-location patterns with varying threshold values (Size=2)

Co-Location Patterns	Threshold=300m	Threshold = 500m
ATM, Banks	0.7407	0.777
ATM, Hospitals	0.8666	0.9333
ATM, Restaurants	0.8260	0.9565
ATM, Schools	0.562	0.8750
Banks, Hospitals	0.740	0.8518
Banks, Restaurants	0.5900	0.7777
Banks, Schools	0.5666	0.6255
Hospitals, Restaurants	0.800	0.933
Hospitals, Schools	0.8125	0.9375
Restaurants, Schools	0.500	0.625

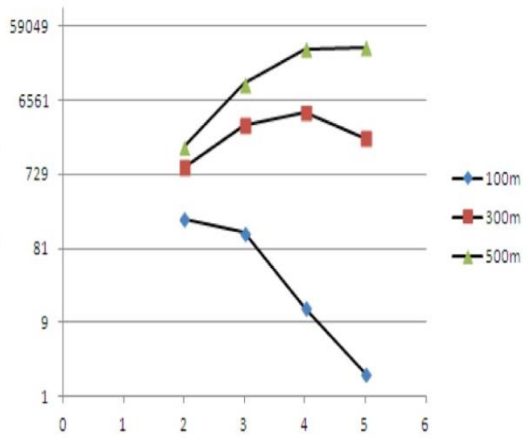

Figure 12 Various size co-location instances obtained for different threshold values

Table 2 Co-location patterns with varying threshold values (Size=3)

Co-Location Patterns	Threshold = 300m	Threshold = 500m
ATM, Bank, Hospitals	0.7077	0.7777
ATM, Bank, Restaurants	0.5925	0.7407
ATM, Bank, Shopping Malls	0.5625	0.625
ATM, Hospitals, Restaurants	0.7391	0.9333
ATM, Hospitals, Shopping Malls	0.5625	0.625
ATM, Restaurants, Shopping Malls	0.5	0.025
Bank, Hospitals, Restaurants	0.5925	0.7407
Bank, Hospitals, Shopping Malls	0.5625	0.625
Bank, Restaurants, Shopping Malls	0.500	0.625
Hospitals, Restaurants, Shopping Malls	0.500	0.625

We extend the proposed approach to identify the colocation patterns for each individual and aggregate them to predict the crowd in the target place. We consider each human activity on daily basis with respect to spatial and temporal aspects. We proposed an approach (Algorithm 1) to achieve the same.

Algorithm 1: Crowd prediction in the target locations from extracted individual spatiotemporal patterns
Input: User public activities recorded in the form of
f(x,y,t).
Output: Crowd prediction {High, Medium, Low} in the target location with respective to time limits.

1. For each user Ui
2. Identify the source of tracking the user activities
3. Store and pre-process the collected data in the Time Range Tw

4. Generate a graph with locations Li[] as nodes and edges represent the sequence followed to visit locations in Li[].
5. End For
6. Aggregate the individual user activity graphs using spatiotemporal clustering techniques.
7. For Each path Pi in Aggregated Graph in TW
8. For Each item Ix in Pi
9. Find the probability of Ix by considering Ix-1 node for all Pj
10. End For
11. If Prob(Ix) <Thd
12. Purge that node from Li[]
13. End If
14. End For
15. Apply supervised machine learning algorithms for predicting the crowd at the Target location(s)

We initially identify the source of tracking the user activities like cab trajectory, self-agreed geographic and mobile tracking IAdamu (2018). We then record the sequence of locations that every user visits and all this information is aggregated using spatiotemporal clustering techniques. We purge the nodes from the aggregated graph which are not meeting the defined threshold. Finally, we apply supervised machine learning algorithms to predict the crowd at a particular location.

Figure 13 Identified Hotspots for multiple users across multiple visits

We tracked multiple user visits for several locations over a period of time and identified hotspot locations if they are meeting a certain threshold (Figure 13). Then, we identified multiple users visiting sequence containing the hotspots (Figure 14). We then have taken a new user sequence to predict the crowd at the next location that he/she is supposed to visit based on the hotspot visiting sequence. The classification accuracy after applying various supervised learning algorithms is shown in Figure 15. We further improved the classification accuracy by considering the various constraints like less congestion path, good road path and shortest path (Figure 16). We identified that the user visiting sequence greatly varies based on the listed constraints by which crowd prediction accuracy at the target location is also improved.

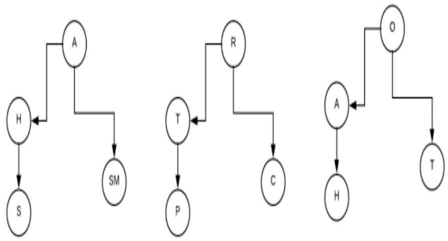

Figure 14 Identified user visiting sequence-based on the hot spots

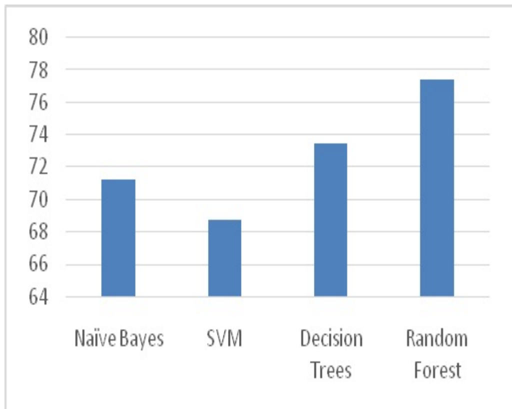

Figure 15 Crowd Prediction accuracy without constraints

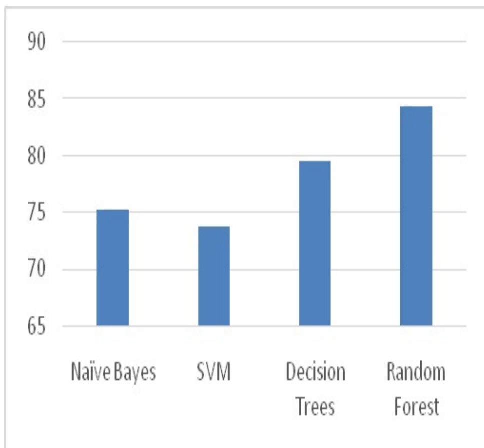

Figure 16 Crowd Prediction at the target location with constraints

Conclusion

Spatial Co-Location Mining is an emerging area and has various real-time applications. We identified that the traditional distance measures cannot be directly used for identifying frequent co-location instances in the target geographical data. We evaluate the effect of network distance to extract the co-location patterns for the spatial data. We customized the existing CPMS model and added a new interesting measure to it. We tested the approach on multiple data sets (synthetic and real) and observed that the observed results are accurate for various co-location patterns. Since most of the human activities are constrained to road networks, this model will be useful. We extended the proposed approach to predict the crowd in target location based on the users visiting sequence based on time. As future work, we can apply the proposed approach for larger areas and test the accuracy of it.

References

[1] Agrawal, R., and Srikant, R. (1994). Fast algorithms for mining association rules. In Proceedings of the 20th International conference on very large databases (VLDB'94), 635 (pp. 487–499).

[2] Appice, A., Ceci, M., Lanza, A., Lisi, F. A., and Malerba, D. (2003). Discovery of spatial association rules in geo-referenced census data: a relational mining approach. In Proceedings of the intelligent data analysis (pp. 541–566).

[3] Cai, Jiannan, et al. Adaptive detection of statistically significant regional spatial co-location patterns. Computers, Environment and Urban Systems 68 (2018): 53-63.

[4] Flouvat, F., Soc, J. N. V., Desmier, E., and Selmaoui-Folcher, N. (2015). Domain-driven co location mining. Geoinformatica, 19(1), 147–183.

[5] Huang, Y., Shekhar, S., and Xiong, H. (2004). Discovering colocation patterns from spatial data sets: a general approach. IEEE Transactions on Knowledge and Data Engineering, 16(12), 1472–1485.

[6] Zuang, Y., Pei, J., and Xiong, H. (2006). Mining co-location patterns with rare events from spatial data sets. Geoinformatica, 10(3), 239–260.

[7] Adamu, Joshua T. Jegede, Hilary I. Okagbue, and Pelumi E. Oguntunde. Shortest Path Planning Algorithm - A Particle Swarm Optimization (PSO) Approach Patience Proceedings of The World Congress on Engineering 2018, pp19-24.

[8] Koperski, K., and Han, J. (1995). Discovery of spatial association rules in geographic information databases. In Proceedings of 4th international symposium on large spatial databases (pp. 47–66). Portland, Maine Springer

[9] Lu, Junli, et al. "Mining strong symbiotic patterns hidden in spatial prevalent co-location patterns." Knowledge-Based Systems (2018): 190-202.

[10] Okabe, A., Boots, B., Sugihara K., and Chiu, S. N. (2009). Spatial tessellations: concepts and applications of Voronoi diagrams. Chichester, UK: Wiley.

[11] Phillips, P., and Lee, I. (2012). Mining co-distribution patterns for large crime datasets. Expert Systems with Applications, 39(14), 11556–11563.

[12] Wang, Lizhen, et al. "Effective lossless condensed representation and discovery of spatial co-location patterns." Information Sciences 436 (2018): 197-213.

[13] Wan, Y., and Zhou, J. (2009). KNFCOM-T: A k-nearest features-based co-location pattern mining algorithm for large spatial data sets by using t-trees. International Journal of Business Intelligence and Data Mining, 3(4), 375–389.
[14] Xie, Z., and Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. Computers, Environment and Urban Systems, 35(5), 396–406.
[15] Yao, Xiaojing, et al. "A spatial co-location mining algorithm that includes adaptive proximity improvements and distant instance references." International Journal of Geographical Information Science 32.5 (2018): 980-1005.
[16] Yu, Wenhao, et al. Spatial co-location pattern mining of facility points-of-interest improved by network neighborhood and distance decay effects. International Journal of Geographical Information Science 31.2 (2017): 280-296.
[17] Yamada, I., and Thill, J. C. (2007). Local indicators of network-constrained clusters in spatial point patterns. Geographical Analysis, 39(3), 268–292.
[18] Yu, W., Ai, T., and Shao, S. (2015). The analysis and delimitation of Central Business District using network kernel density estimation. Journal of Transport Geography, 45, 32–47.
[19] YYoo, J. S., and Shekhar, S. (2004). A partial join approach for mining co-location patterns. 712 In Proceedings of the 12th annual ACM international workshop on geographic information systems (pp. 241–249).
[20] Yao, Xiaojing, et al. "Efficiently mining maximal co-locations in a spatial continuous field under directed road networks." Information Sciences 542 (2021): 357-379.
[21] Wang, Lizhen, Yuan Fang, and Lihua Zhou. "Dominant Spatial Co-location Patterns." Preference-based Spatial Co-location Pattern Mining. Springer, Singapore, 2022. 167-199.
[22] Dong, Shangjia, et al. "Modest flooding can trigger catastrophic road network collapse due to compound failure." Communications Earth & Environment 3.1 (2022): 1-10.
[23] Rinartha, K., & Suryasa, W. (2017). Comparative study for better result on query suggestion of article searching with MySQL pattern matching and Jaccard similarity. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-4). IEEE.
[24] Rinartha, K., Suryasa, W., & Kartika, L. G. S. (2018). Comparative Analysis of String Similarity on Dynamic Query Suggestions. In *2018 Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)* (pp. 399-404). IEEE.
[25] Arianggara, A. W., Baso, Y. S., Ramadany, S., Manapa, E. S., & Usman, A. N. (2021). Web-based competency test model for midwifery students. *International Journal of Health & Medical Sciences*, *4*(1), 1-7. https://doi.org/10.31295/ijhms.v4n1.380