MINING OF SPATIAL CO-LOCATION PATTERN IMPLEMENTATION BY FP GROWTH

G.Manikandan

Research Scholar,
Sathyabama University, Chennai,
Tamil Nadu, India, mani4876@gmail.com.

Dr. S.Srinivasan

Director of Affiliations /
Head of Computer Science and Engineering & Master of Computer Application
Anna University of Technology Madurai, Madurai, Tamil Nadu, India.

Abstract

Mining co-location patterns from spatial databases may disclose the types of spatial features which are likely located as neighbours in space. Accordingly, we presented an algorithm previously for mining spatially co-located moving objects using spatial data mining techniques and Prim's Algorithm. In the previous technique, the scanning of database to mine the spatial co-location patterns took much computational cost. In order to reduce the computation time, in this work, we make use of R-tree that is spatial data structure to mine the spatial co-location patterns. The important step presented in the approach is that the transformation of spatial data into the compact format that is well-suitable to mine the patterns. According to, we have adapted the R-tree structure that converts the spatial data with the feature into the transactional data format. Then, the prominent pattern mining algorithm, FP growth is used to mine the spatial co-location patterns from the converted format of data. Finally, the performance of the proposed technique is compared with the previous technique in terms of time and memory usage. From the results, we can ensure that the proposed technique outperformed of about more than 50 % of previous algorithm in time and memory usage.

Keywords:- *Spatial data mining, Co-location patterns, minimum support, minimum bounding rectangle, FP tree, Vehicle movement data.*

1. Introduction

An automated discovery of spatial knowledge is essential because of the rapid development of spatial data and extensive use of spatial databases. Since more spatial data have been stored in spatial databases, the spatial data mining becomes more significant and interesting. Spatial data mining is defined as the process of mining the interesting and previously unknown, but potentially useful patterns from spatial databases [10, 11]. Due to the intricacy of spatial data types, spatial relationships and spatial autocorrelation, the mining of interesting patterns from spatial datasets is more complex than extracting the corresponding patterns from conventional numeric and categorical data [12]. Spatial patterns have great values in several applications. The significant task of spatial data mining is extracting the spatial co-location patterns in wide applications [9]. Spatial co-location and de-location patterns are same for both positive and non positive association rules. The subsets of Boolean spatial feature types are represented by spatial co-location patterns and its instances are usually situated in close geographic proximity [7]. Conventional spatial associations are extended by spatial de-location patterns in order to include an association rules in the form of $A \rightarrow \neg B$, which represent that B will not exist nearby A. These association rules are competent in mining some useful and previously unknown hidden information and also very useful in some spatial problems [8].

In several applications, the main function of spatial data mining is to extract the spatial co-location patterns [3, 4, 5]. Spatial co-location patterns describe subsets of the spatial features and its objects are usually placed in close geographic proximity. Some examples of co-location patterns are symbiotic species in ecology such as the Nile crocodile and Egyptian plover, frontage roads [31] and highways in metropolitan road maps, and co-located [27] services often requested and located together from mobile devices (e.g., PDAs and cellular phones) in location-based services [6]. The subsets of spatial Boolean events are represented by co-location and its instances are frequently located in a neighborhood. The presence or absence of geographic object types at diverse locations in a 2D or 3D metric space is described by Boolean spatial events, for instance, the surface of the Earth. Business types, mobile service request, disease, crime, climate, plant species and more are the examples of Boolean spatial events [1]. Several existing co-location pattern discovery techniques perform generate-and-test methods, i.e., generating and testing each candidate to find whether it is a co-location pattern [2].

In spatial database systems, spatial data mining and spatial analysis techniques are playing a major role to discover interesting but implicit patterns in spatial datasets of ever increasing size and intricacy. Co-location rule discovery presents challenges because of the following reasons [15, 16]: 1) as the instances of spatial features are enclosed in a continuous space and share neighbor relationships, it is hard to identify co-location instances. So, a large part of the computation time is used to find the co-location instances and 2) as there are no pre-defined transactions in numerous spatial datasets of ever increasing size and intricacy, it is important to reuse association rule mining algorithms for co-location pattern mining [13, 14]. However, for spatial datasets, the similar shift of paradigm in spatial co-location mining becomes very complex due to the lack of a transaction concept, which is significant in frequent pattern definition and its mining algorithms. Neighborhood that is, co-location row instance, enumeration is a major challenge and a key part of any co-location mining algorithm [22]. This problem can be addressed by using a combinatorial technique like apriori or a geometric approach such as spatial-self-join. Based on the definition of neighbors, the combinatorial technique formularizes the problem as a smart clique enumeration problem from a graph. A geometric spatial join approach using a plane sweep method scans the original space and stops at anchor points to gather neighborhood information. Using the spatial database techniques such as spatial indexes, both methods may utilize optimizations at system level [17].

In this paper, we have presented a novel algorithm for effectively mining of spatially co-located moving objects [28] from the spatial databases. The designed statistic measure can help to mine the significant information rather than the measure described in the literature such as, popularity measure and conditional threshold. The input for the proposed algorithm is a spatial database that contains three fields such as, instance id, the spatial information and moving object id. The co-located patterns are mined for the particular area which consist number of vehicles that must satisfy the minimum bounding rectangle value as given. If the number of vehicles is not support the minimum bounding rectangle value means subsequently the area is elaborate in all sides in a unit level to increase the number of vehicles for support the minimum bounding rectangle. This process is repeated up to 3 times whenever the number of vehicle is not supported to minimum bounding rectangle. This process leads to get number of vehicles in each area after that by using the FP growth algorithm the co-located patterns are mined.

The organization of the paper is as follows: The problem statement is described in section 2 and the proposed algorithm for mining of spatially co-located patterns is given in section 3 and the conclusions are summed up in section 4.

2. Problem description and definition

Given a set of features $F = \{f_1, ..., f_n\}$, $1 \le n \le k$ where k is the maximum number of features. The each features f_n consists of some objects O $f_n = \{o_m\}$, $1 \le m \le l$ where l is the maximum number of objects in corresponding features. The each object having the value of x coordinate value and y coordinate value $o_m = \{x, y\}$. Based on the x and y coordinate values, the spatial objects are plotted on their places.

Converting the database format into FP tree input format

The total spatial is divided into grids $G = x/p, y/p$ where p is value which is used to divide the spatial data. Each cell of the grid contain some number of patterns if the number of pattern R_i is below minimum bounding rectangle $R_i < mbr$ then the cell size get increased by single unit in all direction $G_i = x+1, y+1, -x-1, -y-1$. The each of the object is repeated in cell to cell, the number of object $N(o_m)$ is less than the minimum support $N(o_m) < M_s$ then remove the corresponding object. The finalized number of objects is input to the FP tree.

Definition 1 (data). Given a set of features $F = \{f_1, ..., f_n\}$ each of the feature consist of object $O\ f_n = \{o_m\}$ and the each object having the location of the spatial value called x coordinate value and y coordinate value $o_m = \{x, y\}$

Definition 2 (Co-location pattern). A co-location pattern is a set of spatial features with the neighborhood and the node membership function of this set greater than the user specified minimum conditional threshold.

Definition 3 (mbr). The minimum bounding rectangle (MBR), also known as bounding box or envelope, is defined as that the maximum objects object within its 2-D (x, y) coordinate system belonging to defined conditions. Here, the condition to extent the rectangular box is defined based on the minimum number of events bounded within the rectangle and the number of extension carried out to satisfy the minimum bounding condition.

3. Proposed approach

With the wide availability of GPS, wireless, telecommunication, and Web technologies, massive amounts of object movement data have been collected from various moving object targets, such as animals, mobile devices, vehicles, and climate radars. Analyzing such data has deep implications in many applications, e.g., ecological study, traffic control [29][32], mobile communication management, and climatologically forecast. In this paper, we focus our study on vehicle movement data analysis and examined the mining methods for discovery of various vehicle movement patterns. It is common that objects follow some regular movement patterns. For example, vehicles could have some daily behaviors between one area and its destination. With these types of vehicles, one might want to know the relationships among the individuals in particular area. One of the most useful tasks is to find groups of objects that move together in particular location. By discovering such type of clusters [30], one can detect the vehicle movements by area wise. For finding the co-located patterns there is two major steps are required they are

1) Conversion of the spatial data
2) Mining of the co-located patterns

1) Converting the spatial data
    a. Step 1: Input representation
    b. Step 2: The spatial data information
    c. Step 3: Finding the number of vehicles in each area
2) Mining of co-location pattern
    a. Step1: FP tree construction
    b. Step2: Mining from FP-Tree

Once we construct the FP tree, the mining of frequent patterns from the compact tree arrangement is done by FP tree mining procedure defined in [26]. FP-growth functions in a *divide-and-conquer* way. The first scan of the database extracts a list of frequent items in which items are sorted by frequency- descending order. According to the frequency-descending list, the database is packed into a frequent-pattern tree, or *FP-tree*, which retrieves the itemset association information. The FP-tree is mined by initially from every frequent length-1 pattern (as an initial suffix pattern), building its *conditional pattern base* (a “sub database”, which includes of the set of prefix paths in the FP-tree co-occurring with the suffix pattern), then building its conditional FP-tree, and executing mining recursively on such a tree. The pattern growth is obtained by the concatenation of the suffix pattern with the frequent patterns obtained from a conditional FP-tree. The FP-growth algorithm converts the problem of finding long frequent patterns to probing for shorter ones recursively and then concatenating the suffix. It uses the slightest frequent items as a suffix, offering good selectivity. Performance studies show that the method substantially decrease search time.

4. Conclusion

We have presented an algorithm for mining spatially co-located moving objects using spatial data mining techniques. We have presented an efficient algorithm for mining spatially co-located moving objects which materializes spatial neighbor relationship and reduces the computational cost extremely with aid of the well known FP Tree mining algorithm. The spatially co-location mining algorithm efficient since the frequent pattern mining is used here. Finally we have generated the candidate co-location patterns which satisfy the minimum bounding rectangle. We have carried out the experimental evaluation using the synthetic datasets and obtain the computation time is very less when compared with previous algorithm and our proposed method leads to reduce the memory usage extremely when compared with the previous algorithm. From the results, we ensured that the proposed technique outperformed of about more than 50 % of previous algorithm in time and memory usage.

References

[1]. Jin Soung Yoo, Shashi Shekhar, "A Partial Join Approach for Mining Co-location Patterns", Proceedings of the 12th annual ACM international workshop on Geographic information systems, 2004

[2]. Xiangye Xiao, Xing Xie, Qiong Luo, Wei-Ying Ma, "Density Based Co-Location Pattern Discovery", Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems, 2008.

[3]. S. Shekhar and Y. Huang. “Co-location Rules Mining: A Summary of Results”. In Proc. 7th Intl. Symposium on Spatiotemporal Databases, 2001.

[4]. N. Cressie. “Statistics for Spatial Data”. Wiley and Sons, ISBN:0471843369, 1991.

[5]. Y. Morimoto. “Mining Frequent Neighboring Class Sets in Spatial Databases”. In Proc. ACM SIGKDD, 2001.

[6]. Yan Huang, Yan Huang, "On the Relationships between Clustering and Spatial Co-location Pattern Mining", in proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence, Arlington, VA, pp: 513 - 522, 2006.

[7]. S. Shekhar and Y. Huang. “Co-location Rules Mining: A Summary of Results”. In Proc. of Symposium on Spatio and Temporal Database (SSTD), pp. 236- 256, 2001.

[8]. You Wan, Jiaogen Zhou, Fuling Bian, "CODEM: A Novel Spatial Co-location and De-location Patterns Mining Algorithm", Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008.

[9]. Yan Huang, Hui Xiong, Shashi Shekhar, Jian Pei, "Mining confident co-location rules without a support threshold", Proceedings of the 2003 ACM symposium on Applied computing, 2003.

[10]. J. Roddick and M. Spiliopoulou, “A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research”, In Proc. SIGKDD Explorations, Vol: 1, No: 1, pp: 34-38, 1999.

[11]. S. Shekhar and S. Chawla. Spatial Databases: A Tour. Prentice Hall, 2003.

[12]. S. Shekhar, P. Zhang, Y. Huang, and R. Vatsavai. “Trends in Spatial Data Mining, as a book chapter in Data Mining: Next Generation Challenges and Future Directions”. H. Kargupta, A. Joshi, K. Sivakumar and Y. Yesha (editors), AAAI/MIT Press, 2004

[13]. R. Agarwal and R. Srikant. “Fast algorithms for mining association rules”. In Proc. of Int’l Conference on Very Large Databases (VLDB), 1994.

[14]. K. Koperski and J. Han. “Discovery of Spatial Association Rules in Geographic Information Databases”. In Proceedings of Int’l Symposium on Large Spatial Data bases, Maine. pp: 47-66, 1995.

[15]. Jin Soung Yoo, Shashi Shekhar, Mete Celik, "A Join-less Approach for Co-location Pattern Mining: A Summary of Results", in proceedings of Fifth IEEE International Conference on Data Mining, pp.813-816, 2005.

[16]. Shekhar, S., and Huang, Y., “Discovering Spatial Co-location Patterns: A Summary of Results”, In Proceedings of the 7th International Symposium on Advances in Spatial and Temporal Databases, Springer-Verlag, London, pp. 236-256. 2001.

[17]. Mete Celik, James M. Kang, Shashi Shekhar, "Zonal Co-location Pattern Discovery with Dynamic Parameters", Seventh IEEE International Conference on Data Mining,

[18]. You Wan, Jiaogen Zhou, Fuling Bian, "CODEM: A Novel Spatial Co-location and De-location Patterns Mining Algorithm", in proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery, Shandong, Vol: 2, pp: 576, 2008.

[19]. Al-Naymat, G., "Enumeration of maximal clique for mining spatial co-location patterns", in proceedings of the IEEE/ACS International Conference on Computer Systems and Applications, pp: 126, Doha, 2008.

[20]. Lizhen Wang, Yuzhen Bao, Lu, J., Yip, J., "A new join-less approach for co-location pattern mining", n proceedings of the 8th IEEE International Conference on Computer and Information Technology, Sydney, NSW, pp: 197, 2008.

[21]. Yue Jiang, Lizhen Wang, Ye Lu, Hongmei Chen, "Discovering both positive and negative co-location rules from spatial data sets", in proceedings of the 2nd International Conference on Software Engineering and Data Mining, Chengdu, pp: 398, 2010.

[22]. Yoo, J.S.; Shekhar, S.; "A Joinless Approach for Mining Spatial Co-location Patterns", IEEE Transactions on Knowledge and Data Engineering, Vol: 18, No: 10, pp: 1323 - 1337, 2006.

[23]. Lizhen Wang, Lihua Zhou, Joan Lu and Jim Yip, "An order-clique-based approach for mining maximal co-locations", Information Sciences, Vol: 179, No: 19, pp: 3370-3382, 2009.

[24]. Feng Qian, Qinming He and Jiangfeng He, "Mining Spatial Co-location Patterns with Dynamic Neighborhood Constraint", Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science, Vol: 5782, pp: 238-253, 2009.

[25]. You Wan, Jiaogen Zhou, "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, Vol. 3, No.4, pp. 375 - 389, 2008.
[26]. Jiawei Han,Jian Pei, Yiwen Yin and Runying Mao," Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach", vol.8, no.1, 2004.
[27]. G.Manikandan and S.Srinivasan, "Mining Spatially Co-Located Objects from Vehicle Moving Data", International Journal on "European Journal of Scientific Research – EJSR", ISSN 1450-216x Vol. 68 No. 3 PP. 352 - 366, 2012.
[28]. G.Manikandan and S.Srinivasan "A Novel Approach for effectively mining for spatially co-located moving objects from the spatial data base", International Journal on "CiiT International Journal of Data Mining and Knowledge Engineering", PP 816 - 821, ISSN: 0974 – 9683 Volume 3,No 13, October 2011.
[29]. G.Manikandan and S.Srinivasan, "Traffic Control by Bluetooth Enabled Mobile Phone", proceedings of International Conference on "Database and Data Mining – ICDDM 2012", Chengadu, China, May 5 – 6 2012. Co-Sponsored by the International Association of Computer Science and Information Technology (IACSIT) and Sichuan Institute of Electronics.
[30]. G.Manikandan and S.Srinivasan, " SRC in web Mining ", proceedings of International Conference on " Global Computing and Communication – ICOGCAC 09", Hindustan University, Chennai, PP 179, 9th – 11th December 2009, ISBN 978 – 81 – 8424 – 543 – 1.
[31]. G.Manikandan and S.Srinivasan "Traffic Management on Road Section", Proceedings of the National Conference on "Recent Trends in Computer Science and Engineering – NCRTCSE 2010" Sathyabama University, Chennai, PP 117 - 120, on 5th & 16th March 2010.
[32]. G.Manikandan and S.Srinivasan "A Study of Data Mining application by using GPS and GIS Based Traffic Information Systems", Proceedings of the National Conference on "Computer Science and Informatics – CSI'12" Anna University of Technology Madurai, Madurai, on 23rd & 24th April 2012.PP 83 – 87, ISBN 978 – 1 – 62050 – 345 – 4.

Mr. G. MANIKANDAN received the B.E degree in Computer Science & Engineering from Madras University, M.E degree in Computer Science & Engineering from Sathyabama University, Post Graduate Diploma in Geo Spatial Information Technology from Periyar Maniammai University and the M.B.A degree in HR from Tamil Nadu Open University, Where he is currently working towards the Ph.D degree in Computer Science & Engineering at Sathyabama University. His area of interests includes Data mining, Spatial Databases, Geographic information System and intelligent Transportation Systems.

Dr. S. SRINIVASAN received the Ph.D degree in Computer Science & Engineering from the Sathyabama University, Chennai, M.Tech. in Computer Science & Engineering from the Indian Institute of Technology, Chennai, M.B.A. in Systems & Finance from Sathyabama Engineering College, University of Madras, Chennai, and the M.Sc. in Mathematics from the Gobi Arts College, Gobichettipalayam, Bharathiar University, Coimbatore. He is presently working as Director of Affiliation and Head / Department of Computer Science and Engineering, Anna University of Technology Madurai. He has rich experience in Teaching and Research. He has also held various positions and responsibilities in Technical Institutions. He is currently serving as an expert in various capacities at different levels. He has published more number of research papers in journals, books, conferences, and workshops. His research interest include text mining, data mining & data warehouse, and image database.