Volume 3, Issue 4, April 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Hash-Based Collocation Rule Mining Approach for Spatial Environment

Karthika. A *
SCSE&VIT University
India.

Priya Pandey
SCSE&VIT University
India.

Amala Vetha Sophia.J
SCSE&VIT University
India.

***Abstract*— There are various approaches to predict the occurrence of spatial attribute in the region. The present class of algorithms are typically memory intensive, leading to prohibitive runtimes on large databases. To overcome this, we have proposed an improvised hash-based collocation rule mining which results in the increase of throughput. In this method, we form a hash table for the selected features which influences spatial attribute to occur and applied colocation rule mining to determine its relative occurrence.**

Keywords— ***memory intensive, hash based collocation rule mining***

I. Introduction

Spatial database is a database which stores information about spatial attribute(s).It has wide areas of application like geosciences, Soil and rock science, Climate study, Hydrology, Mining geology. Item type is considered Boolean spatial, there are some implicit finite set of transactions because to continuity of underlying space. The spatial database faces a challenge which is spatial join [1].Unlike relation database, spatial database has to undergo operations such as overlapping, disjointing, touching, intersecting and many more. When a Boolean spatial feature and their instances are provided, spatial collocation mining seeks to discover a set of features whose instances are frequently collocated in close proximity. Predicting the existence of a spatial attribute can be performed by examining its colocation with other spatial attribute. In order to find the colocation patterns, the spatial database must contain values of all the attributes which are related to any spatial attribute. With the values in database, colocation rule is applied. To reduce the number of candidate key generation, hash based technique is followed [2]. Computational cost, size and memory can be reduced by using this hash based colocation rule mining method [3]. In this paper, spatial colocation is generated by the spatial collocation rule mining algorithm with hash function.

The proposed algorithm will have following advantages:

- Less number of candidate key generations.
- Scanning time is reduced.
- Each value obtained in scanning database is stored in hash table and thus retrieval of those values can be easily made [3].
- Base area is efficiently utilized and the address of stored attribute in database can be calculated based on some simple arithmetic function content with the mapping function.

A spatial dataset has the following three properties:

1. A set of spatial feature types S={S1,S2,..Sn}, e.g., "park," "school," "zoo "and "library."

2. A set of feature instances C= {C1, Cm} such that $Ck \in C\ (1 \leq k \leq m)$ is an instance of type $Sl \in S, (1 \leq l \leq n)$. For example, "Zion National Park" is an instance of the type "park," and "Logan High" is an instance of the type "school."

3. A set of neighbourhood relationships whose elements are an unordered pair of instances in Cof different types, such that the two instances are neighbours to each other [3]. For example, ("Logan City Library," "Logan High") denotes that the two instances are neighbours. The neighbourhood relationships are usually defined according to Euclidean distances. However, since the neighbourhood relationships are given as input, the user can define other kinds of neighbourhood relationships according to the specific application, thus providing more flexibility.

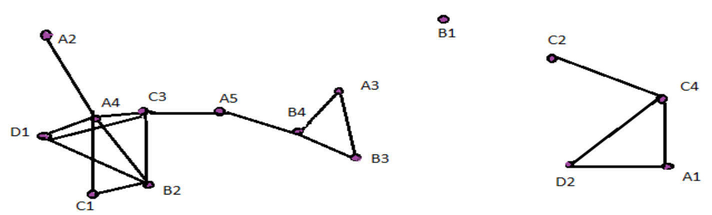

Figure 1 Spatial types A, B, C, D and their instances.

As an example, consider Figure 1 wherein each node represents an instance, and each edge denotes the neighbourhood relationship between the two corresponding instances at both ends. Hence, we identify a set of types $S = \{A, B, C, D\}$, a set of instances $C\{A1, A2, A3, A4, A5, B1, B2, B3, B4, C1, C2, C3, C4, D1, D2\}$

Definitions of the concepts used in our paper, which commonly appear in the spatial co-location literature are as follows:

Let f(Ck) be a function that returns type of instance Ck

II. CO-LOCATION INSTANCE LIST

Given a set of types S, a set of instances C defined over S, and a set of neighbourhood relationships R over C, a list of instances <C1,C2,..Cm>such that Ck$\in C$,is a called co-location instance list if

1. Any pair of instances in <C1, C2,..Cm>is an element of R, i.e., is a <C1,C2,..Cm> is a clique, and
2. $f(Ci) < f(Cj)$holds for every pair of instancesCi,Cj.

III.PARTICIPATION RATIO

Given a type set Sand a co-location type list C defined over F, let CL denote a co-location list set complying with C. Then, the participation ratio of C conditioned on type Si$\in$C is defined as

Pr(C/Si) = $\frac{\text{NumberofdistinctinstancesofSiinCL}}{\text{NumberofinstancesofSiglobally}}$

IV.CO-LOCATION RULE

Given a type set S, and a set of co-location type lists C$^{\tau}$, defined over F, a co-location rule is a rule of the form C1->C2 such that C1 ,C2$\in C\tau$ andC1$\cap$C2=$\emptyset$

With the help of Co-location Instance List, Participation Ratio, Co-location Rule spatial co-location rule mining in continuous data is to be found.

V. SPATIAL ASSOCIATION RULES MINING

The reference feature centric model [18] to discover spatial association rules.In this model, a spatial association rule is a rule of the form:

$$C1 \cap C2 \cap \dots Cm \xrightarrow{yields} E1 \cap E2 \dots En(s\%, c\%)$$

Where and $Ci(1 \leq i \leq m)$ and $Ei(1 \leq i \leq n)$are predicates and at least one of these Predicates should be spatial predicates. Spatial predicates can involve any spatial relationship supported by the spatial databases. The support of the rule, denoted as s% and confident of the rule, denoted as c%.

SQL-like spatial data mining query language is adopted for obtaining data from spatial database.

Suppose that a user wants to discover the association rules among houses in the state of Oregon and their physical proximity to roads, rivers, mountains, parks, and byways. Such a query can be written as:

Discover spatial association rules
Inside Oregon
From road R, river L, Mountain M, Park p
In relevance to house H
Where close to (H.geom, X.geom) and X$\in \{R, L, M, P, B\}$

The mining of spatial association rules takes the following steps:

1. Extract spatial objects from the spatial database.
2. For each object of the reference class, discover its specified spatial relationship with spatial objects of other classes.
3. Mine spatial association rules.

VI.PROPOSED ALGORITHM

Apriori algorithm with hash based approach is introduced in this paper in order to reduce the number of scans to determine the occurrence of spatial attribute. The values of spatial attributes are made to undergo discretization which then converted to Boolean attribute. Then the spatial Boolean attribute values are taken and stored in the database from which occurrence of some other spatial attribute can be determined. The values are made to store and retrieve in the form of hash table using some simple mathematical function.

The following steps are followed in hash based spatial rule mining algorithm:

INPUT: Spatial attributes, Values for each attribute, Final spatial attribute to determine colocation with other attribute.

OUTPUT: occurance of spatial attribute in colocation to other.

Steps:

1. The spatial attributes with its values has to be stored in the database (to predict rainfall, its attributes such as temperature, pressure and wind speed can be taken).
2. Preprocess the data with the application of discretization(Continuous to Boolean).
3. Select the occurrence of the spatial database with its respective values.

4. Generate association rule mining for Li level (initially i is 1).increment i as it reaches this step.
   i. $X \Rightarrow Y$
5. Calculate the support and confident for each association rule generated.
   a. Support=$n(X \cap Y)/ n(T)$
   b. Confidence=$n(X \cap Y)/n(X)$
6. With the help of calculated support and confident values, store it in hash table with attribute and its support count.
7. Scan the hash table again and find check for the support>SUPPORT_THRESOLD
8. When condition mentioned in 7 is satisfied, then add those attribute for next iteration.
9. Continue 5,6,7,8 until colocation is found.

VII. CONCLUSION

In this paper, we have presented a spatial co-location mining algorithm integrated with hash-based technique which can be used to collocate the occurrence of one spatial attribute to another. Using the hash-based technique, scanning database is greatly reduced and efficiency of generating colocation rules are also increased.

REFERENCES

[1] Agrawal, R. and Srikant, R. Fast Algorithms for mining association rules. In *Proceedings of 20th International Conference on Very Large Databases*, 1995.

[2] Arge, L., Procopiuc, O., Ramaswamy, S., Suel, T., and Vitter, J. S. Scalable sweeping-based spatial join. In *Proceedings of 24 International Conference on Very Large Databases*, 1998.

[3] Berge, C. *Graphs and Hypergraphs*. Elsevier, 1976.

[4] Celik, M., Kang, J. M., and Shekhar, S. Zonal co-location pattern discovery with dynamic parameters. In *Proceedings of 7th IEEE International Conference on Data Mining*, 2007.

[5] Celik, M., Shekhar, S., Rogers, J. P., Shine, J. A., and Yoo, J. S. Mixed-drove Spatio-temporal co-occurrence pattern mining: A summary of results. In *Proceedings 6th IEEE International Conference on Data Mining*, 2006.

[6] Estivill-Castro, V. and Lee, I. Data mining techniques for autonomous exploration of large volumes of geo-referenced crime data. In *Proceedings of 6th International Conference on Geocomputation*, 2001.

[7] Graefe, G. Sort-merge-join: An idea whose time has(h) passed? In *Proceedings of the 10th International Conference on Data Engineering*, 1994.

[8] Huang, Y., Pei, J. and Xiong, H. Mining Co-Location Patterns with Rare Events from Spatial Data Sets. *GeoInformatica 10*, 3 (2006), 239-260.

[9] Huang, Y. and Zhang, P. On the relationships between clustering and spatial colocation pattern mining. In *Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence*, 2006.

[10] Koperski, K. and Han, J. Discovery of spatial association rules in geographic information databases. *Lecture Notes in Computer Science 951* (1995), 47-66.

[11] Kuramochi, M. and Karypis, G. GREW-A scalable frequent subgraph discovery algorithm. In *Proceedings of the 4th IEEE International Conference on Data Mining*, 2004.

[12] Morimoto, Y. Mining frequent neighboring class sets in spatial databases. In *Proceedings of the 7th ACM Conference on Knowledge Discovery and Data Mining*, 2001.68

[13] Munro, R., Chawla, S., and Sun, P. Complex Spatial Relationships. In *Proceedings of the IEEE Bioinformatics Conference*, 2003.

[14] Patel, J. M. and DeWitt, D. J. Partition based spatial-merge join. In *ACM Conference on Management of Data*, 1996.

[15] PostGIS: Documentation. 2008. http://postgis.refractions.net/documentation/, Nov. 2008.

