A Parallel Approach for Finding Spatial Colocation Patterns

[1] M.Sheshikala,[2] D.Rajeshwara Rao,[3] P.Manasa

[1]Assistant professor, Dept.of CSE, SR Engineering college, Warangal, Telangana, India, marthakala08@gmail.com
[2]Professor, Dept.of CSE, KL University, Guntur, Andhra Pradesh, India, rajeshduvvada@kluniversity.in
[3]PG Scholar, Dept.of CSE, SR Engineering college, Warangal, Telangana, India, pittalamanasa@gmail.com

ABSTRACT:Spatial data mining come to be one of the vital essential areas due to the fact that of the fast evolution in science which leads in enormous spatial data.Co-location data mining is an interesting and most important predicament in spatial data mining discipline which discovers the subsets of points whose events are usually placed together in geographic area.Spatial proximity is the major inspiration to verify the colocation patterns from huge data.The computation of co-location data discovery is very high-priced with enormous data volume and nearby existence of neighborhoods. So there is number of spatial co-loaction mining algorithms had been developed to overcome these drawbacks. In this paper, a new co-location data mining framework has been proposed that benefits from the power of parallel processing,in particular, the Map Reduce to obtain higher spatial mining processing effectivity. Map Reduce model have been proven to be an efficient framework solution for big dataprocessing on clusters of commodity machines, and for big data evaluation and lots of functions. The experimental result of the proposed framework indicates scalable and efficient computational efficiency.

KEYWORDS-DataMining,Colocation Pattern,MapReduce,Constraint Neighborhood.

I. INTRODUCTION

Data mining quite often is looking for hidden and exciting patterns that may exist in frequent data. Spatial data mining in detailed is discovering the fascinating relationships and traits that can exist implicitly in spatial data [1].Spatial data mining is a new and rapidly establishing discipline of data mining, concerned with the identification of intriguing spatial patterns from data stored in spatial datasets and geographic information systems.GIS are utilized in various areas such as environmental influence comparison, city planning, cartography, criminology,traffic evaluation, and so on. Assortment of information is enabled by global positioning techniques(GPS) and sensor networks, even as laptop storage technology allows the storage of huge quantities of accumulated knowledge.These advanced technologies are the rationale for the existence of a growing quantity of spatial datasets.The size of spatial datasets and the complexity of dealing with spatial attributes require the use of specialized data mining techniques[2].Spatial co-place pattern mining, is likely one of the primary subject in spatial data mining, has been researched in spatial data mining techniques. Spatial co-place pattern describes “a suite of spatial activities which can be mostly located collectively in a spatial proximity”[3].Also a co-vicinity sample determines what these colocated objects,each and every probably occur in geographical proximity. Recognized co-vicinity patterns are exciting and useful for a lot of functions comparable to region-established services,public well being,climatology,sicknessmanage,Transportation,trade,Social Science,Geology, and mobile computing [3],[4].Co-location rules are items to seek out the presence of Boolean spatial sides within the nearby of instances of different Boolean spatial aspects. Mining Co-loction pattern is the approach to determine co-location patterns from huge spatial datasets with a number of Boolean elements.The spatial colocation rule discovery concern looks like

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 18
December 2016

the organization rule mining problem,but,correctly, it is rather distinct from the organization rule mining quandary. These variations have been made in view that of the lack of transactions.It uses spatial predicate as object varieties.So utilizing of co-location pattern mining enhancing the efficiency of detecting fascinating patterns from the very significant spatial data.Co-location patterns are learned with the aid of using local definition and spatial joins. These definitions and algorithms shall be mentioned the detection of co-location pattern from the huge spatial datasets [5].Significant data is one of the hotspot in technological discipline and brings not only colossal amounts of data but additionally quite a lot of data types that would no longer been regarded [10].The evolution of location sensing, mobile computing, and scientific simulation is producing big quantities of rich spatial data.Finding the solution that is capable to translate the considerable quantity of spatial data that surrounds us into meaningful andvaluable data has ended in the upward push of spatial data mining[20]. So Spatial data mining has been popularly studied for detecting a designated association relationships between a suite of spatial attributes and a few of them is also non-spatial attributes.But dealing with tremendous-scale spatial data mining isn't effortless for the reason that of problematic spatial information forms, neighbor relationships [19].

II. RELATED WORKS

The problem of mining association rules based on spatial relationships (e.g., proximity, adjacency) was first discussed in [6]. The work discovers the subsets of spatial features frequently associated with a specific feature, e.g., cancer. Fig.1(a) shows an example dataset with three spatial features, A,B and C. Each object is represented by its feature type and unique instance id, e.g., A.1.Identified neighbor objects are connected by solid lines. Fig. 1 (b) shows the neighbor objects near the objects of a specific reference feature type, A. A setofneighboring objects of each reference object is converted to a transaction. All rules generated from the transactions are related to the specific feature. However, directly applying this approach to the co-location mining may not capture our colocation meaning with no specific reference feature.

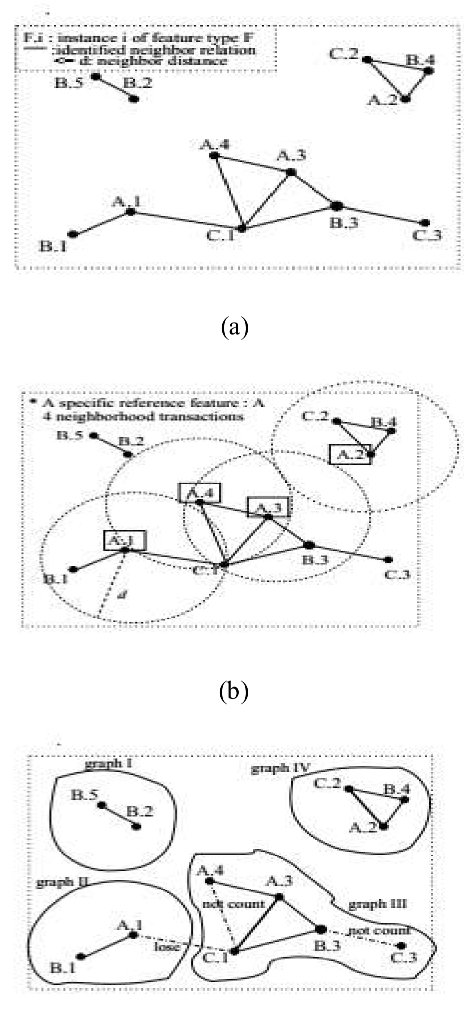
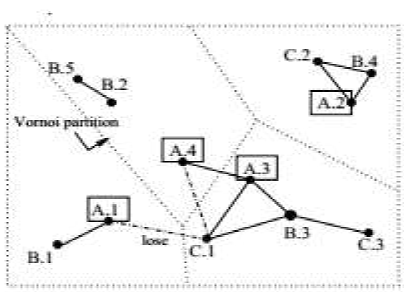


Fig. 1. (a) An example spatial dataset (b) A reference feature based spatial association rule mining (c) Subgraph mining

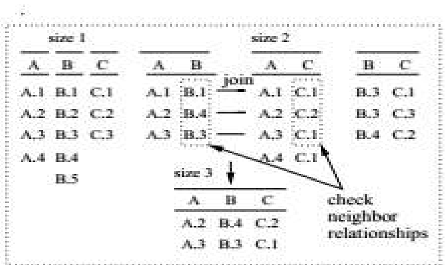
Previous works on co-location pattern mining have presented different approaches for identifying co-location instances and choosing the interest measures

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 18
December 2016

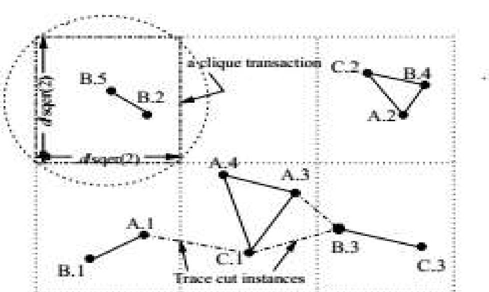
of co-location patterns [8] discovers frequent neighboring class (e.g., colocated features) sets using a support count measure. This approach uses a space partitioning and non overlap grouping scheme for identifying neighboring objects.However, the explicit space partitioning approach may miss co-location instances across partitions like fA.4, C.1g in Fig.2(a). [7] proposed statistically meaningful interest measures for co-location patterns and a join-based co-location mining algorithm.



(a)



(b)



(c)

Fig. 2. Different approaches for finding co-location instances (a) Space partition (b) Instance join (c) Clique partition and residual instances

The instance join operation for generating co-location instances is similar to apriori gen [5]. First, after finding all neighbor object pairs (size 2 co-location instances), the approach finds size k(> 2) co-location instances by joining the instances ofits size k 1 subset co-locations where the first k 2 objects are common, and checking the neighbor relationship betweenk 1th objects. Fig. 2 (b) shows generating the instances of co-location fA, B, Cg. This approach finds correct and complete co-location instance sets. However,apriori-join operation is computationally expensive with the increase of colocation instances. In the spatial database literature, multi-way join techniques using R-trees for multiple spatial features [9],[10] can be used as alternatives to the apriori-join. However,we assume our spatial dataset has no spatial index, and thus it is difficult to apply these techniques to our co-location mining directly. In our previous work [11],we proposed a partial join algorithm to reduce the number of expensive join operations in finding co-location instances. This approach transactionizes objects using clique neighbor relationships,e.g., posing simple grids in Fig. 2(c), and uses the apriori join only for the residual instances not modeled by the explicit transactionization. This method reduces the number of join operations significantly. However, its performance depends on the number of cut instances by explicit partitioning.

Co-location pattern mining may appear to be similar with subgraph mining [12], [13], which finds frequent subgraphs in a large graph database. The prevalent interest measure used is support, the ratio of graphs which include a subgraph pattern. For example, in Fig. 1(c), the support of subgraph AC is 24. Subgraph mining is a different problem from our colocation mining, which finds all frequent subsets(i.e., clique subgraphs) from a spatial dataset (i.e., conceptually a single graph which represents input spatial objects and their neighbor relationships). A spatial dataset can be represented to a set of disjoint graphs, and a subgraph mining technique might be applied to it. However, some neighbor relationships can be lost by the distinct partition, e.g., {fA.1, C.1} in Fig. 1(c), and the support measure may count one instance in a graph, e.g., {fA.3, C.1}, but not other instances, e.g., {fA.4, C.1}.

III. THE PROPOSED APPROACHES

In this section, we present our parallel colocation pattern mining model based on constraint neighborhood approach and MapReduce model, which starts by Identifying spatial input dataset reordering the instances records rendering their types then according to ID'S. COUNT find total number of instances for each feature type Find CN for each object in the dataset according to spatial relationship between objects such that R(oi,oj) calculated according to constraint neighbor approach. Find list of constraint neighbor for each object.Eliminate remove object that are inappropriate. If object has no neighbors and doesn't included in a list of constraint neighbor. Finally generate colocation patterns.Our proposed system find prevalent co-location patterns and realized across three map/reduce jobs.

A. Phase 1: Preparation of Spatial Files

Our Spatial Dataset obtained from more than one file. So there is need to combine these files into one file using MapReduce model.First the map job assigns each object oito its corresponding feature type fi. Then applying the reduce job for sorting these objects according to their features types and within the same type according to the objects ID'S. Then counting and saving the number of object instances per feature type for future prevalence calculation.The ordering task used for eliminating the duplications and missed instances. If we have D spatial dataset consists of no of bjects D = {o1,o2, ..., on} , and F set of m features F ={f1,f2, ..., fm}, (m << n), i.e., f1< f2<... <fm, id1< id2< ... <idn .

B. Phase 2: Generate the List of Constrain Neighbors (CN)For Each Object

In this phase the main job is to generate the list of neighbors of each object according to the definition of the constraint neighborhood approach CN by check each object with other to find the constraint neighbor list according to the following:For clique colocation:CCN({oi}):= sort<({oj |(oi,oj) ∈ R ∧ ((oi.type<oj .type)∨ (oi.type = oj .type∧ oi.id <oj .id)),(j ≠ i)}). For star colocation patterns:SCN ({oi}):=sort<({oj|(oi,oj) ∈ R, (j ≠i)}).Then builds a set of single feature (size-1) co-location candidates.

C. Phase 3: Colocation Patterns Generation

In this phase the algorithm applies the level-wise approach togenerate size-k pattern candidates from size-(k−1) prevalent patterns and checks whether all subsets of the new candidate are prevalent. The pattern instances of the new candidate are discovered. The algorithm uses the participation index that has the anti-monotonic property to measure the prevalence of new candidates.

IV. CONCLUSION

In this paper, we advise the materialization of nearby data for efficient co-location data mining.We present two local materialization items, a star partition model and a clique partition model,with out a loss ofco-vicinity occasions.Also an efficient parallel co-location data mining strategy proposed that easily discovers colocation patterns and self co-location patterns based on constraint regional method. Additionally the drawbacks in previous methods had been superior by means of using hadoop-MapReduce model which enable us with the parallel and allotted processing manner.

REFERENCES

[1]Banalata Sarangi and LaxmanSahoo,"A Survey onSpatial Association Rule Mining Technique and Algorithms for mining spatial data",International

Journal of Scientific & Engineering Research, Vol.4, Issue 12, pp.1664- 1670, 2013.

[2]N.Lavrac,D.Jesenovec,N.TrdinandN.M.Kosta,"Mining spatiotemporal data of traffic accidents and spatial pattern visualization",Metodoloskizvezki, Vol. 5, No. 1, pp. 45-63, 2008.

[3]Tran Van Canh and Michael Gertz,"A Constraint neighborhood based approach for co-location pattern mining",Fourth International Conference on Knowledge and System Engineering, pp. 128-135,2012.

[4]H.Yang, S.Parthasarathy and S.Mehta, "A generalized framework for mining spatio-temporal patterns in scientific data",In Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,ACM,pp. 716–721, 2005.

[5]VenkatesanMeenakshiSundarama,Arunkumarthnagavelub and Prabhavathy Paneerc,"Discovering Co-locationPatterns from Spatial Domain using a Delaunay Approach",International Conference On Modelling Optimization And Computing,Vol. 38, pp. 2832–2845, 2012

[6]Sonali Satija and Rajender Nath,"Performance Improvement of Apriori Algorithm Using Hadoop",International Journal of AdvancedResearch in Computer Science and Software Engineering ,Vol. 5, Issue 6, pp. 765-768, June2015.

[7]Gy oz o Gid ofalvi,"Spatio–Temporal DataMining for Location–Based Services", Ph.D.Thesis, Faculties of Engineering, Science andMedicine at Aalborg University, Denmark, 2007.

[8]G.KiranKumar, P.Premchand and T.VenuGopal, "Mining Of Spatial Co-location Pattern from Spatial Datasets",International Journal of Computer Applications,Vol. 42– No.21, pp. 25-30, March2012.

[9]Lizhen Wang,YuzhenBaoa and ZhongyuLub, "Efficient Discovery of Spatial CoLocation Patterns Using the iCPI-tree",The Open Information Systems Journal,Vol. 3, pp. 69-80,2009.

[10]Jin SoungYoo, Douglas Boulware and David Kimmey,"A Parallel Spatial Co-location Mining Algorithm Based on MapReduce",IEEE International Congress on Big Data,Anchorage,AK,pp. 25 – 31, 2014.

[11]S.Shekhar and Y. Huang,"Co-location Rules Mining:A Summary of Results",in Proceedings of International Symposium on Spatio and Temporal Database, pp. 236–256, 2001.

[12]Rushirajsinh L. Zala, Brijesh B. Mehta and Mahipalsinh R. Zala,"A Survey on Spatial Colocation Patterns Discovery from Spatial Datasets", International Journal of Computer Trends and Technology (IJCTT),Vol. 7, No. 3, pp.137-142,2014.

[13][https://github.com/hadoop illuminated/hadoopbook] Accessed on 10/9/2015

[14]Othman Yahya, Osman Hegazy and EhabEzat, "An Efficient Implementation Of Apriori Algorithm Based On Hadoop Mapreduce Model",International Journal of Reviews in Computing,Vol. 12,pp.59-67, 2012.

[15][https://hadoop.apache.org/docs/r1.0.4/mapred_tutorial.pdf] accessed on 2/8/2015

[16]Jin SoungYoo and ShashiShekhar,"A Partial Join Approach for Mining Co-location Patterns",GIS'04, Washington,DC, USA, ACM, November12–13, 2004

[17]T. Elsayed, J. Lin, and D. W. Oard,"Pairwise Document Similarity in Large Collections with MapReduce",in Annual Meeting of the Association for Computational Linguistics on Human Language Technologies,2008, pp. 265–268.

[18]Y.Morimoto,"Mining Frequent Neighboring ClassSets in Spatial Databases",In Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2001.

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 18
December 2016

[19]Jin SoungYoo ,Shashi Shekhar and Mete Celik, "A Join-less Approach for Co-location Pattern Mining: A Summary of Results",5th IEEE International Conference on Data Mining, 2005.

[20]Jin SoungYoo and Douglas Boulware,"A framework of spatial co-location mining on MapReduce", IEEE International Conference on Big Data, pp.44., 2013.-945，2009