Volume 4, Issue 11, November 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Co-Location & Segregation Pattern Mining by Means of Statistical Importance: An overview

Amrita Kulkarni
Department of CSE, G.H.R.A.E.T.,
RTMNU, Nagpur, India

Prof. Deepak Kapgate
Department of CSE, G.H.R.A.E.T.,
RTMNU, Nagpur, India

Abstract: A subset of features whose instances are commonly located together in spatial neighborhoods is co-location pattern and subsets of features whose instances are not commonly seen to be located in spatial neighborhoods are segregation pattern. The method of discovering interesting and unidentified but useful pattern in spatial database is spatial data mining. However, segregation pattern is not researched as co-location pattern. Techniques for mining co-location pattern are mainly divided into data mining and statistical approach, which is further classified on its use of transaction type of data. This paper presents overview on different approaches used to find co-location pattern, however segregation pattern mining is not yet developed area.

Keywords: spatial neighborhood, spatial data mining, co-location, segregation, statistical importance

I. INTRODUCTION

Spatial data can be said as the data that analyze the geographic location of features such as plants, animal species in forest areas, oceans, etc. Spatial data is generally stored as coordinates and topology and this data can be represented in form of maps. Spatial data can be accessed and managed through Geographic Information Systems (GIS). To search interesting and useful patterns from large spatial databases spatial data mining is performed.

Spatial data mining is complicated than traditional data mining because spatial data is generally autocorrelated, unlike non-spatial data which is independent. Hence it can be said that conventional data mining methods are not proper to spatial data because they do not support location data or the hidden relationships between objects.

Co-location can be defined as, subset of features whose instances are commonly located together in spatial neighborhoods which is presentation of positive interaction between features. Segregation pattern is negative interaction pattern representation can be defined as subsets of spatial features whose instances are rarely seen to be located in spatial neighborhoods. Spatial co-location and segregation patterns can give way to important insights for many applications.

This paper gives overview of the different approaches for finding co-location patterns. However research work on segregation pattern which is corresponds to negative pattern is not advanced like co-location pattern. Further paper is designed as Section 2 describes different approaches on co-location and segregation pattern mining, which is sub-divided into two sections data mining and Statistical approach. Table 1 gives summery of techniques or algorithms reviewed in this paper. Section 3 concludes the paper.

II. VARIOUS APPROACHES ON CO-LOCATION AND SEGREGATION PATTERN MINING

Techniques for finding co-location pattern are mostly based on data mining and spatial statistics. Different approaches in association analysis are classified on its transaction datasets i.e. transaction free and transaction based approaches. In statistical method spatial autocorrelation is taken into account.

A. Data mining

First type in this approach is transaction free approach [5] which does not produce its transaction type data. Mohammad Akbari et al. [3] transactions free approach that considers fuzzy definition of neighborhood. Regional co-location Pattern Mining algorithm which generates co-location candidates, calculates neighborhood value for each co-location instances based Fuzzy neighborhood. New participation ratio is also developed to find involvement of features. G. Kiran Kumar et al. [6] introduces new prevalence measure as mediod participation index. Firstly algorithm to generate co-location rules is given, later it calculates medoid Participation index which takes the medoid of the Participation ratios of the features in the feature set.

Jin Soung Yoo et al.[9], joinless co-location mining algorithm based on star neighborhood partition model, which separates neighboring objects using star relationship, is used. A partial join co-location mining algorithm modified as coarse filtering scheme introduced to decrease multiple join operation. Clique partition model which separates objects that have clique relationship is used in partial join co-location.

Yan Huang et al. [10], this paper focused on finding co-location pattern with rare event which is missed due to equal participation of features. A maximal participation ratio (maxPR) as measure which shows that if maxPR is high co-location pattern contains high rare events.

Yan Huang et al. [12], paper proposed concept of proximity neighborhood to mine co-location patterns which is transaction free approach. Prevalence-Based Pruning, Multi-resolution pruning techniques is also specified. Hui Xiong et al. [11], transaction free approach to mine extended spatial objects. A buffer based model for co-location pattern discovery used. EXCOM algorithm, will give set of co-location patterns, and apply pruning.

Shashi Shekhar et al., [13], this paper proposes concept of user specified neighborhood instead of transactions of specified group of items. Co-location miner algorithm to find co-location patterns is introduced. B. Arunasalam et al. [14] will mine positive and negative patterns in complex spatial relationship, later use statistical technique to determine if these relationships are substantial. To mine complex relationship NP_MaxPI algorithm is introduced.

Second approach is transaction based mining [5]; G. Kiran Kumar et al. [4], classification of instances of features according to its neighborhood has three types event-centric, Reference Feature Centric & window centric model. This paper proposed technique as Hierarchical Window Centric Model in which set of spatial features separated into four windows and window centric model applied on each window independently.

Seung Kwan Kim et al.,[5], this paper introduces framework which is transaction based approach i.e. generates its own transaction data type, this paper need two algorithms MAximal Clique Enumerator (MACE) algorithm used to find maximal cliques and Apriori algorithm to generate frequent patterns also Generate_Neighboring_Graph algorithm used to find the neighborhood of every spatial object.

Venkatesan M et al., [7], this paper process data in form of co-ordinates which generates instances of features. Later by calculating participation index and pruning technique co-location pattern is found. Event centric model focuses on subsets of spatial features probably to occur in a locality around instance. Xiangye Xiao et al., [8] this paper, in the test step instances of a candidate to obtain its prevalence are recognized. In order to reduce the computational cost of recognizing these instances, a density based approach is presented. The objects are divided into partitions and identifying instances. A dynamic upper limit of the prevalence for a candidate is maintained. If the current upper limit becomes less than a threshold, stop recognizing its instances in the remaining partitions.

B. Statistical approach:

Sajib Barua et al., [1] proposed method which takes spatial autocorrelation into consideration and extract co-location and segregation pattern. To find these patterns statistical significance test is used. The prevalence measure, participation index (PI), is calculated using null hypothesis and general observed data. Randomization test performed to generate positive and negative values. This paper also introduces a neighborhood sampling approach using a grid based space partitioning.

Jundong Li et al., [2] this paper proposes co-location mining algorithm which indicate statistical significant co-locations in datasets. At first buffer is used to demonstrate affected area near an object, second transaction dataset is generated, and at last statistical significant co-location is identified.

Table 1: Summery of various approach and their features

Sr. no.	Technique/ Algorithm Used	Features	Reference
I.	SSCSP: Statistically Significant Co-Location & Segregation Pattern	• Compute PI values of all interaction patterns • Generate a data set under the null model • Compare probability values i.e. p-pos, p-neg and detect pattern	[1]
II.	CMCStatapriori: Co-Location Mining Constrained Statapriori	• Generates transaction dataset • Efficiently detect more specific co-location rules	[2]
III.	Regional Co-Location Pattern Mining	• Based on definition of fuzzy neighborhood • With the help of lower and upper bound of neighborhood instances are generated	[3]
IV.	Apriori Algorithm	• New prevalence measure as mediod participation index	[6]
V.	Joinless Colocation Mining Algorithm	• Star neighborhood partition model, which separates neighboring objects using star relationship	[9]
VI.	Co-location Mining Algorithm: A General Approach	• Generation of candidate co-locations, • Generation of table instances of candidate • Generation of collocation rules	[12]
VII.	Co-Location Miner	• User specified neighborhood instead of transactions of specified group of items	[13]
VIII.	Hierarchical Window Centric	• Set of spatial features separated into four windows and window centric model applied on each window Hierarchically	[4]

IX.	Event Centric Modeling Approach	• Focuses on subsets of spatial features which occur in a close region around instance	[7]
X.	Density based co-location pattern mining algorithm	• Uses grid-based partitioning • Generates candidates • Density summation of the ratios of the number of instances of each size-k pattern	[8]

III. CONCLUSION

Hence, we have reviewed different techniques for co-location & segregation pattern mining. Many techniques are based on data mining which divides its sub-type as transaction based and transaction free types. A transaction based mining generates more co-locations than transaction free type. Similarly in statistical approach transaction datasets are generated. However, statistical approach is more accurate because, association analysis can miss some patterns if they are smaller in number than its prescribed limit. Future work includes, combining new prevalence measures and various sampling approaches.

REFERENCES

[1] Sajib barua et al., “Mining Statistically Significant Co-Location And Segregation Patterns”, IEEE transactions on knowledge and data engineering, vol. 26, no. 5, may 2014

[2] Jundong li et al., “Discovering Statistically Significant Co-Location Rules in Datasets with Extended Spatial Objects”, DaWaK 2014: 124-135

[3] Mohammad Akbari et al., “New Regional Co-Location Pattern Mining Method Using Fuzzy Definition of Neighborhood”, ACSIJ Advances in Computer Science: an International Journal, Vol. 3, Issue 3, No.9 , May 2014

[4] G. Kiran Kumar et al., “Hierarchical Window Centric Method of Modeling Spatial Co-Location Patterns on Spatial Database”, IJARCSSE Volume 3, Issue 11, November 2013

[5] Seung kwan kim et al., “A Framework Of Spatial Co-Location Pattern Mining For Ubiquitous GIS”, Multimedia Tools Application Springer Science+Business Media, LLC 2012

[6] G.Kiran Kumar Et al., “ Mining Of Spatial Co-Location Pattern From Spatial Datasets”, International Journal of Computer Application, Volume 42– No.2, 0975 – 88871, March 2012

[7] Venkatesan.m et al., “Event Centric Modeling Approach In Colocation Pattern Analysis From Spatial Data”, International Journal of Database Management Systems (IJDMS), Vol.3, No.3, 125-133, August 2011

[8] X. Xiao et al., “Density based co-location pattern discovery,” in Proceedings 16th ACM Int. Adv. GIS, Irvine, CA, USA, 2008, pp. 250–259

[9] J. S. Yoo et al., “A Joinless Approach For Mining Spatial Co-Location Patterns” IEEE transactions on Knowledge & Data Engineering, vol. 18, no. 10, pp. 1323–1337, Oct. 2006.

[10] Y. Huang, et al. “Mining Co-Location Patterns With Rare Events from Spatial Data Sets” Geo-Informatica, vol. 10, no. 3, pp. 239–260, 2006.

[11] Hui xiong et al., “A Framework For Discovering Colocation Patterns In Data Sets with Extended Spatial Objects”, Geoinformatica, 2006

[12] Y. Huang et al., “Discovering Co-location Patterns from Spatial Data Sets: A General Approach”, IEEE transactions on Knowledge and Data Engineering, vol. 16, no. 12, pp. 1472–1485, Dec. 2004.

[13] Shashi shekhar et al., “Discovering Spatial Colocation Patterns: A Summary Of Results”, Symposium on Large Spatial Databases - SSD , pp. 236-256, 2001

[14] B. Arunasalam, et al., “Striking two birds with one stone: Simultaneous mining of positive and negative spatial patterns,” in Proceeding of 5th SIAM ICDM, 2005, pp. 173–182.

[15] K. Koperski et al. “Discovery of spatial association rules in geographic information databases,” in Proceedings 4th International SSD, Portland, ME, USA, 1995, pp.47-66.

