Mining Co-Location Patterns in Spatial Dataset – A Review

Swati Meshram,
Department of Computer Engineering,
Government College of Engineering, Amravati, Maharashtra, India.
Email- swati.meshram@computersc.sndt.ac.in

Kishor P. Wagh,
Department of Information Technology,
Government College of Engineering, Amravati, Maharashtra, India
Email- kishorpwagh2000@gmail.com

ABSTRACT

Co-location pattern discovery in the spatial environment determines a spatial prevalent pattern over a region. It identifies the instances of spatial features, often co-occurring based on neighbourhood relationship. The participation index is used to determine if two features co-locate together with conditional probability. Co-location pattern mining has applications in biology, criminology, ecology, epidemiology, earth science, transportation, urbanization, health, recommendation etc. The aim of the review is to present the co-location pattern mining conception, approaches, types of co-location patterns and its applications along with future scope of work.

Keywords: Pattern mining, co-location, participation index, symbiotic patterns, interaction patterns, contrast and common sets.

I. INTRODUCTION

Spatial dataset has an interesting pattern Co-location which has drawn a wide interest among researchers. Co-location pattern discovery locates objects which coexist as entities globally or are prevalent locally in geographical neighbourhood. [1].

Co-location pattern mining in spatial environment leads us to find relevant and significant domain related insights in the areas such as, biology, criminology, ecology, epidemiology, earth science, transportation etc. Co-location patterns have been successfully applied to understand the urban function of cities and prevent urban problems to offer better urban infrastructure to its citizen. Colocation mining has been experimented to discover relationship between child cancer cases and pollutant emission in the atmosphere [2]. In criminology, co-location patterns based on criminal events in the study region along with socioeconomic and socio-demographic factors are found to be helpful to develop and implement effective control strategies to reduce crime rate etc.

These co-located objects found are useful to identify a facility located near to a person or to a location. It has an application in recommendation of neighbourhood like hospitals, restaurants, banks, malls, educational institutions etc. [1].

Spatial data is usually obtained from any location based services like maps obtained from google earth, google maps, weibo, Bhuvan, bing, yahoo etc. The spatial data could be represented as points, a set of points form "line", a collection of points in an area having boundaries to form "polygons" etc. Figure.1 shows a sample dataset of spatial features with objects '+' are mostly co-located along objects 'o'. and objects '*' are co-located with '∇' where {+, o, *, ∇,♦} is the feature set . A spatial dataset is a collection of entities called as spatial features $F = \{f_1, f_2, \ldots f_n\}$ and its instances called as objects $O = \{o_1, o_2, o_k\}$, where $o_j \in fi$ along with latitude and longitude as its attributes [2].

A basic concept in traditional spatial pattern mining is the notion of prevalence index which determines the significance of a pattern. For a co-location rule to be prevalent or of interest, where co-location rule takes the form as:

$$C1 \rightarrow C2(PI, cp) \qquad (1)$$

here C1, C2 are co-location patterns, PI is the prevalence or the participation index [6,10] and cp is the conditional probability. The prevalence index PI(C) of a co-location pattern C is defined as:

$$PI(C) = min_{fi \in C}\{pr(C, f_i)\} \quad (2)$$

where pr {C, f_i} is the participation ratio of a feature f_i in a colocation C and is given by

$$pr(C, f_i) = \frac{the\ number\ of\ distinct\ instances\ of\ f_i\ in\ instances\ of\ C}{total\ instances\ of\ f_i} \quad (3)$$

A co-location pattern is prevalent or interesting pattern, if for a feature, minimum PI% of instances are in a clique with other features instances. Which relies on neighbourhood or close proximity relationship and is a measure of the distance between these pair of objects. A co-location pattern or rule is considered as significant if it has a high level of prevalence given that spatial features are independent of each other.

The figure 2. depicts a general approach in spatial colocation pattern mining. On the spatial data collected pertaining to a study region, neighbourhood is enumerated. For every feature instance, as per the close proximity or distance measure, neighbourhood instances of other features are generated. The neighbourhood instances are evaluated for interestingness based on the prevalent thresholds with algorithm to discovery prevalent co-location patterns.

The aim of the review is to analyse and present the current scientific literature relating to the co-location pattern mining based on various approaches, applications and open challenges in the field.

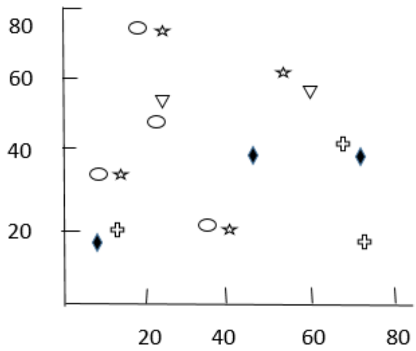

Figure 1: An illustration of different spatial features co-locations.

The paper is organized as section 2 highlights the related work, section 3 is discussion and section 4 is conclusion.

II. RELATED WORK

2.1 Spatial Statistical Approaches

The statistical approaches that are used for Co-location mining are based on Ripley's Cross-k functions Monte-Carlo simulations [3], pair–wise correlation function, spatial autocorrelation and regression models [4]. Monte-Carlo simulations are used for theoretical distribution estimation on spatial point process. The distribution estimation methods based on distance measure applied are pair wise distance, mean nearest neighbour distance, empty space distance for clustering tendency [5].

2.2 Spatial data mining approaches

Spatial data mining approach is a process to uncover interesting and useful patterns computationally embedded in spatial dataset. Colocation pattern mining in spatial data is one such study that is explored by many researchers. Colocation mining general framework was first proposed by Shekhar and Huang in 2001[6]. Spatial data collected as maps is location data and is not present in the form of transaction. So to analyse such data by applying association rule mining, transformation spatial data into transaction is necessary. The pattern mining problem is analysed as association rule problem with prevalence and conditional probability measures. Their work adopts Apriori algorithm [7-9] with enumeration of neighbourhood candidates to explore collocation patterns. According to Shekhar and Huang [6], different ways to identify the neighbourhood are categorized as

- Local,
- Reference feature centric
- Window centric
- Event centric.

In the local model, each arbitrary partition is transformed to a transaction for mining patterns without emphasizing on partition boundaries. As given in Figure 2. in the reference centric model, transactions are formed based on neighbouring instances close-by to the reference features instances. It is suitable for focusing on a domain specific boolean spatial feature example cancer. The window centric model generates all possible windows to form transactions. It is suitable for geology, surveying applications. The event centric model captures subset of spatial features occurring in the neighbourhood of given event types. It is suitable for ecological application where man spatial features take boolean values [6].

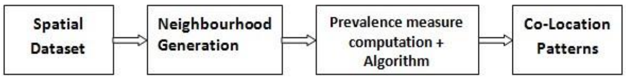

Figure 2: General approach to spatial co-location pattern mining.

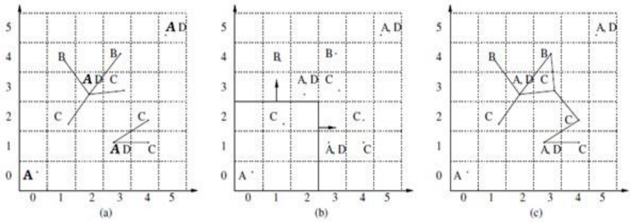

Figure 3: Neighbourhood generation models: (a) Reference feature centric, (b) Window centric, (c) Event centric. Ref: S.Shekhar. et.al (2003).

The various spatial data mining approaches majorly explored by researchers are

1. Join-based.
2. Partial join-based.
3. Join-less based.
4. Graph based.
5. Distributed – map and reduce based

Huaang and Shekhar [10] propose a join-based algorithm on geometric, combinatorial and hybrid methods for neighbourhood enumeration with pruning. Apriori algorithm is used to discover the collocation patterns and later these candidates are pruned based on prevalence measure.

- Initially, size-1 colocation instances are prevalent due to its participation index being 1.
- Size-2 Colocation candidates are generated by geometrical approach.
- Candidate participation index is computed, if PI <min_prev threshold than it is prunes.
- Compute the Colocation rules.

Yoo [11] has proposed a partial join approach which is equally computationally expensive. Spatial data is transformed into transactions. Neighbourhood is enumerated as clique instance and where relationship cuts on the cliques, for those instances each clique is converted into a transaction. Then the algorithm follows apriori property to find intra-X and inter-X instances and uses bitmap. It saves computational time as total cutrelations formed will be less.

Yoo [12] proposed join-less approach of star instance look-up scheme in order to eliminate the need of join operation. The neighbourhood relations are examined through star neighbourhood partition model. “The star neighbourhood of an object is a set of the canter object and objects in its neighbourhood whose feature types are greater than the feature type of the center object in a lexical order”. Star neighbourhood instances are generated for each candidate instance to form the superset based on neighbourhood relationship. Generate candidate co-location for size 1 and 2. Those candidates above size as 2, apply generalized apriori algorithm. Now using the participation index, pruning of candidates is performed and required collocation enumeration is complete. Check is performed to see all instances in the given star instance forms a clique. Compute participation index and generate co-location rules.

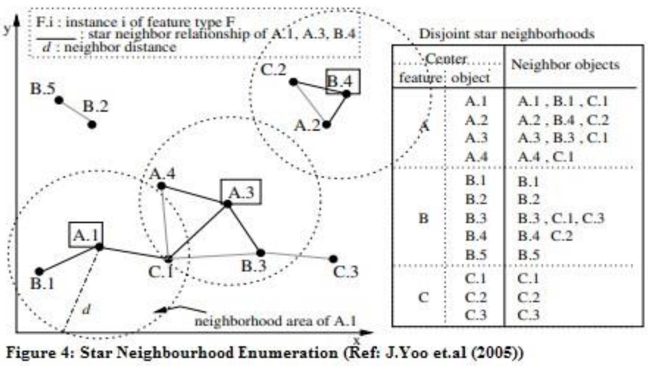

Figure 4: Star Neighbourhood Enumeration (Ref: J.Yoo et.al (2005))

Co-location Pattern Instance (CPI) tree [30] is another join-less based approach which does not perform expensive join computations. The method enumerates neighbourhood relationships in optimized way with the recursive and hierarchical properties of tree data structure. Size-(k + 1) colocation instance is generated from respective size-k colocation instances without waiting for all size-k co-location instances are enumerated.

The graph based approach is presented in [13] where the spatial data is studied in graph domain. The scope of their work is to discover co-location patterns based on clique enumeration on neighbourhood graph. Three algorithms CliqueEnumG, CliqueEnumK and CliqueExtend have been proposed. The outline of these algorithms is which works on first to enumerate the candidate clique and further validate it for neighbourhood relations using graph traversal.

In [1], a distributed approach using map and reduce is proposed to discover co-location patterns in spatial data. The method: first starts with counting total objects of a feature in the entire dataset. The R-proximity neighbourhood which provides the similarity measure between objects and is a distance measure. The mapper transforms the unstructured data into structured key-value pair for those objects which satisfy R proximity. Based on CP the conditional probability of a rule and user defined minimum support δ, frequent colocation patterns are found. Reducer does the summarization of all frequent patterns. The candidate patterns are generation is based on clique property.

Types of Co-Location Patterns

Based on recent years' papers published on various co-location pattern, we classify the co-location patterns types as:

- Contrast and common sets.
- Multi-level - Global and local patterns.
- Symbiotic patterns.
- Interaction Patterns – co-location and segregation.
- Clique Pattern.

Contrast and Common Sets: Traditional colocation pattern mining methods are based on frequency thresholds which leads to discard of rare patterns if the threshold is high and detects noisy patterns when the threshold is low. To overcome these limitations a type of co-location pattern which is contrast set and common sets is proposed in [14].

"Contrast sets are another class of associative patterns which are used to characterize a particular class and contrast it from the others. Contrast sets are defined as conjunctions of attribute value pairs, $X \subset A$, defined on mutually exclusive classes from C such that no $A_i \in X$ occurs more than once." [14-15].

Using class association rule $X \rightarrow C_i$ contrast sets are detected where X is a contrast set for class C_i to distinguishes C_i from other classes.

$$\exists_{i,j}\ P(X\,|c_i) \neq P\big(X\,|c_j\big) \tag{4}$$

$$\max_{i\ j} |support(X, c_i) - support(X, c_j)| \geq \min - dev \tag{5}$$

where support(X, c_i) is the support of X in class c_i. Equation 4 states that a contrast set is significantly different from various groups. Equation 5 states that difference of support of a contrast among various groups is large [14].

$$pf(X \rightarrow G_{x,y}) \leq pf\,(X \rightarrow G_{p,q}) \tag{6}$$

$$\max_{i\ j} |support\,(X, G_{x,y}) - support(X, G_{p,q})| \geq \min - dev \tag{7}$$

where X is spatial contrast set for the group $G_{x,y}$ over the other groups of interest $G_{p,q}$ and pf ($X \rightarrow G_{x,y}$) is the Fisher's p-value for the co-location pattern and support (X, $G_{x,y}$) is the support of X in the data of $G_{x,y}$ [14]. The Fisher's exact test is to measure the statistical significance of rules rather than using prevalent threshold. Fisher's exact test is most commonly used for categorical data to classify objects in two different ways.

Contrast set has been applied in health care domain to identify a relationship between neonatal birth abnormalities and air pollution in various cities of Canada. A research question posed for mining in [14] as "Are there any combinations of industrial air pollutants present which leads to lower birth weight in Toronto area than any other city in Canada?" A contrast set or discriminative collocation pattern has helped to successfully contrast Toronto city from many other cities of Canada [31]. Spatial common sets can be used to identify patterns which are commonly seen in many spatial regions/groups.

$$pf(X \rightarrow G_{x,y}) - pf(X \rightarrow G_{p,q}) \leq \max - pf - diff \qquad (8)$$

$$|support(X, G_{x,y}) - support(X, G_{p,q})| \leq \max - dev \qquad (9)$$

where max-pf-diff is a user defined threshold. It has been used to control the changes in significance of a common set within the given set of spatial groups. max-dev is the allowed maximum support difference between any two different groups. These two constraints make sure that the statistical significance and the support of the common set does not vary significantly across spatial groups. These two equations will not allow large differences between the groups.

Multi-level - Global and local patterns

According to Liu [16], due to the spatial heterogeneity of unevenly distributed features, co-location patterns are present at multiple levels – some co-location patterns are present in the entire study area are global co-location patterns. Some exist only in the local regions of the study area can be termed as local co-location patterns [17], [18].

The uncovering of such multilevel co-location patterns provides insights on the interaction among the different spatial phenomena. Multi-level co-location patterns can be found using the participation index(PI), if the participation index of the candidate pattern is greater than the threshold then it is global pattern. Otherwise the candidate pattern is identified as local pattern. The natural neighbourhood relationship along with PI classifies global and local patterns. An example in Figure 5 depicts global and local patterns. In the paper [32], for finding multi-level co-location patterns spatial adaptive clustering is used along with overlapping method.

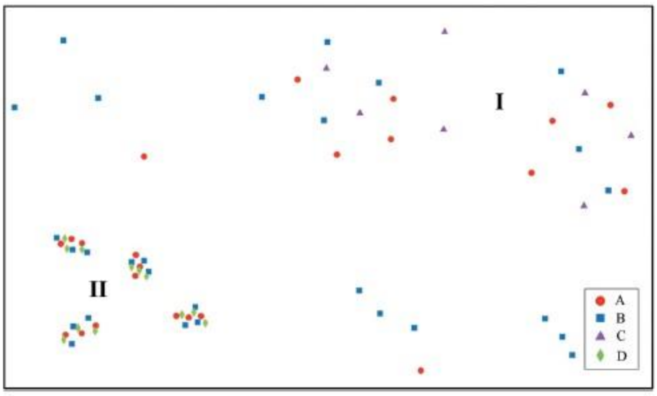

Figure 5: A sample dataset depicting global and local multi-level patterns. Ref: Liu et.al (2020)

Consider, if threshold $T = 0.5$; then, $\{A, B\}$ ($PI = 0.68$) is identified as a global co-location pattern, whereas $\{A, B, C\}$ ($PI = 0.25$) and $\{A, B, D\}$($PI = 0.42$) will be identified as candidate local co-location patterns.

Symbiotic patterns:

As in [19], symbiotic pattern is a concept borrowed from ecological science or biology. According to Wikipedia[20], Symbiotic patterns have three main categories. Mutualism is when every feature in the pattern receives benefit from the relationship and it is the strongest symbiotic pattern. Commensalism is when few features in the pattern receive the benefit and others are not affected which is called general symbiotic pattern. Whereas Parasitism is a pattern, here some features get the benefit while others are harmed and is not worked upon in their study. The "benefit "in symbiotic pattern signify that the features coexist and their existence or survival entirely depends on other features. An algorithm based on changed star neighbourhood and changed star instances is used to identify the co-changed colocation instances and co-location pattern in the neighbourhood which share life or death for existence.

Interaction Patterns – co-location and segregation: According to Barua [5] "An interaction pattern is a subset of k different features f1, f2, . . . , f_k having a spatial interaction within a given distance Rd. Rd is called as the interaction distance. A group of features are said to have a spatial interaction if features of each possible pairs are neighbours of each other. Two feature instances are neighbours of each other if their Euclidean distance is not more than the interaction distance Rd. Let C = {f1, f2, . . ., f_k} be an interaction pattern. In an instance of C, one instance from each of the k features will be present and all these feature instances are neighbours of each other."

According to Barua [5], spatial data interaction patterns are of two types: **co-location** which possess positive interaction and **segregation** patterns possess negative or inhibition interaction. "Segregation patterns, representing negative interactions, can be defined as subsets of boolean spatial features whose instances are infrequently seen to be located at close spatial proximity (i.e., whose co-locations are "unusually" rare)."

Clique Pattern:
A clique pattern is discovered in neighbourhood graph where the neighbourhood relationship is represented as a clique [13]. A clique is a subset of vertices in a graph that are adjacent to each other. Features of a pattern are considered to be co-located if they satisfy the neighbourhood relation using great-circle distance or orthodomic distance measured on the surface of a sphere and is the shorter distance. Enumerating candidate cliques and validating the clique instances is performed with the graph traversal algorithms.

Applications of Co-Location Mining
***Recommendation* System:**
The tabular patterns data in table.1 [5], demonstrates the statistically significant collocation patterns. This signifies low density residential locality highly co-exist with universities. Universities coexist with a fire station. And in some location, low density residential locality, a university, a fire station coexists together. The conclusions drawn are: 1. The key feature is University 2. a health care facility (i.e. nursing home) is missing from classification result and does not coexist with other facilities. This knowledge may serve as an insight to develop the urban infrastructure. If a person lands in a new city and wants information on a nearest landmark, a hotel, a hospital, a cinema hall, a market place etc. within a close proximity to him, these facilities could then be recommended with the help of generated co-location mining patterns.

Low density resident	University	Fire Station	Police Station	College	emergency service	Nursing home	School	Prevalent Index
√	√							0.969
√		√						0.079
√			√					0.157
	√			√				0.263
	√	√						0.95
	√		√					0.12
	√							0.022
		√	√		√			0.095
√			√					0.9
√	√	√				√		0.47
√	√		√					0.087
√	√				√			0.22
	√	√		√				0.0158
	√	√	√					0.317
√	√	√	√					0.0158
	√	√					√	0.012

Table 1: Statistically significant co-locations. A feature present in a co-location is shown by √

These kind of applications of co-location may be served in a recommendation system application.

Environment study for healthcare:
The co-location patterns are identified to establish a relationship between air pollutants in the surrounding area at the emission points and child cancers cases [2]. If a colocation patterns are found highly prevalent in the area surrounding the emission points, so the control measures may be initiated to protect human lives.

Urban Planning:
A proper understanding of urban function or facility (i.e. point of interest) in terms of types of facilities and how these urban facilities are spatially organized [21]–[23]. This knowledge may serve to promote enhanced built environment from the perspective of spatial co-located facilities. By aggregating the resulting CPs, and its representation in graphs for every city to reflect the inter-connections of different function types.

Ecology: In ecological application co-location patterns are found to establish the symbiotic relationships among plants of different species, ages, and sizes to understand the dynamics and specific structures of an ecosystem [24].

Criminology:
In criminology, co-location patterns that are formed by criminal events, socio-economic and socio-demographic factors may derive crucial insights for developing effective policing strategies to reduce crime in the study regions [25].

III. DISCUSSION

Spatial Statistical approaches
The statistical methods are found to be computationally expensive as exponential candidate subset are to be tested and are found inefficient when explored with many spatial features. So spatial data mining approaches has received much attention.

Table 2: Types of Co-location Patterns with comparison of various techniques and features.

Reference Paper	Type of pattern	Neighbourhood relationship/Interesting measure	Features	Remarks	Dataset
[19]	Symbiotic	Star neighbourhood	Changed star neighbourhood, changed star instances, symbiotic index	SSP Average Symbiotic Index (ASI)=0.71 , Average Participation Index(API)=0.84, PCP ASI=0.45,API=0.81	Real-1
[16]	Multi-level -global and regional or local	Natural neighbourhood	Delaunay triangulation, MLminer method	Result Precision 1, Recall 1	Wetland
[32]		Clique-proximity neighbourhood	Adaptive spatial clustering, overlap method, Delaunay triangulation	Result Pattern size=2, Precision 1, Recall 1	Lansing Woods
[5]	Significant co-location and segregation	neighbourhood-Euclidean distance	Spatial interaction, participation ratio, participation index, neighbourhood sampling-grid based space partitioning.	Result Precision 0.64, Recall 1, F-measure 0.78	Ants, Bramble Cranes, Lansing Woods,
[1]	Co-location and Segregation	R-proximity neighbourhood	Map and Reduce framework	Average execution time is 15minutes for data over 100K points.	Bramble Cranes, Lansing Woods,
[14]	Contrast sets and Common sets	Statistically significant association rules	Transactionization, Fishers test, DiSConS, DiSComS	on average 495 co-location rules per CMA discovered.	CNN
[13]	Clique	Neighbourhood graph relation	Neighbourhood graph traversal, Apriori property, Candidate Clique enumeration and instance validation	PI=0.01,Size of SCP=7, Time taken by CliqueExtend=80.8 sec,CliqueEnum$_G$=162.975 sec, CliqueEnumK=343.379sec	Titan

Spatial Data mining approaches

The join-based approach uses bitmap technique to compute participation ratio and does not hold intermediate results. The algorithm performs neighborhood check for all instances in the join step and makes it computationally expensive. To improve the efficiency of join based approach, algorithms are presented in [26][27].

Yoo[11] has proposed a partial join approach which is equally computationally expensive when the number of cut-relations are very high. In the worst case when each instance forms as separate transaction in which case the performance of this approach will be same like join-based approach.

Yoo [12] proposed join-less approach of star instance look-up scheme, to reduce the computational cost as compared to the join based approach for generating co-location instances. It saves computational time as join operation is not performed.

In [13], the graph the vertices represent the instances of the features and edges represent the neighbourhood relationship.co-location patterns discovery problem is mapped to enumeration of cliques on all vertices of the graph which is a NP-Hard problem.

In distributed approach using map and reduce [1] large number of resources are required as data is kept or distributed at many nodes and all nodes have to participate in every process of computation.

.

In [5][14] authors argues that most of the collocation pattern mining algorithms use a single user defined prevalence threshold or distance threshold for the detection of interesting co-location patterns. If the distance of the instance is smaller than the distance threshold, a neighbour is identified. However, it is really difficult to give a perfect distance threshold. Due to which certain rare co-location patterns are either missed or discarded. Instead, statistical significance test applied on colocation patterns will not discard such interesting but rare events. Some methods like voronoi diagrams for cliques, Delaunay triangulation etc. would be more suitable for neighbour relationship which works without threshold. This will not lose some rare yet crucial patterns.

The graph based co-location pattern mining authors claim their algorithm supports interactive-user analysis and the neighborhood constraint parameters can be varied over a range.

In [14] the authors argue that that the prevalence Index threshold should not be global and pre-defined, but should be determined depending on the distribution of the data and the total number of instances of each individual feature involved in an interaction so that we do not lose patterns which are rare and are interesting. To identify statistically significant patterns statistical significance tests-Fishers p value is used to evaluate a rule. This approach does not scale well with large data sets.

The multi level co-location patterns[16] method employs heuristic without time consuming enumeration of all possible localities. However the existing significance tests cannot be used to identify local co-location patterns.

The authors in [19] argue that mining Prevalent Co-location Patterns (PCPs) is more expensive as compared to mining Strong Symbiotic Patterns(SSPs), for following reasons 1) the candidates PCPs being all co-location patterns are found on the entire database, and (2) the candidates of SSPs being PCPs searched on changed dataset which is much smaller than the entire database. Furthermore, we find that most of the methods have not considered autocorrelation.

IV. CONCLUSION AND FUTUREWORK

Co-location pattern mining is a research area with variety of applications which comes as an aid to tackle human problems or to serve for betterment of human lives. The paper refreshes, co-location pattern mining conception, various types of patterns, approaches and limitations of methods, applications of colocation mining and comparison of colocation mining methods by review of recent research work is presented. The comparison of the recent research work is presented in terms of types of patterns, neighbourhood relationship, features of the methods, dataset and result.

Moreover, following issues may be considered for further exploration and research. A contrast and common sets co-location mining problem in spatial datasets is an underexplored area of research. No significant work has been done to find co-location patterns which are in contrast or common to many spatial groups as well. To statistically quantify significance of Fishers p-value test is used to evaluate a rule, this method does not scale well with large data sets. Work can be carried out to improve its performance. Further research also has to be carried out in symbiotic pattern mining as it is not explored in greater detail. Along with mining of competitive patterns and causal rules patterns on dynamic spatial databases is another future work. Existing significance test are not effective to detect global and local patterns so efficient significance test has to be devised for multilevel co-location patterns based on natural neighbourhood in future work. The temporal dimension of the spatial features also needs to be consider for adaptively construct neighbouring relationship. In the graph based approach, a key-value store has been employed to generate the candidate cliques instances or to validate them, at a time but not both. So we see the possibility of using the key-value store for both of the processes together.

REFERENCES

[1] S. Maiti and R. B. V. Subramanyam, "Mining co-location patterns from distributed spatial data," *J. King Saud Univ. - Comput. Inf. Sci.*, Aug. 2018, doi: 10.1016/j.jksuci.2018.08.010.

[2] J. Li, A. Adilmagambetov, M. S. Mohomed Jabbar, O. R. Zaïane, A. Osornio-Vargas, and O. Wine, "On discovering colocation patterns in datasets: a case study of pollutants and child cancers," GeoInformatica, vol. 20, no. 4, pp. 651–692, Oct. 2016, doi: 10.1007/s10707-016-0254-1.

[3] C. Tilke, "Statistics for spatial data : N.A.C. Cressie (1991) New York: John Wiley & Sons, 920 pp., ISBN 0-471-84336-9, 71 [pound sign] sterling," *Comput. Stat. Data Anal.*, vol. 14, no. 4, pp. 547–544, 1992.

[4] "Exploring spatial analysis in geographic information systems / Yue-Hong Chou. - rove."https://trove.nla.gov.au/work/7455149 (accessed Aug. 03, 2020).

[5] S. Barua, "Mining Statistically Significant Co-location and Segregation Patterns.," *IEEE Trans Knowl Data Eng*, vol. 26, no. 5, pp. 1185–1199, 2014, doi: 10.1109/TKDE.2013.88.

[6] S. Shekhar and Y. Huang, "Discovering Spatial Co-location Patterns: A Summary of Results," in *Advances in Spatial and Temporal Databases*, Berlin, Heidelberg, 2001, pp. 236–256, doi: 10.1007/3-540-47724-1_13.

[7] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," in *Proceedings of the 20th International Conference on Very Large Data Bases*, San Francisco, CA, USA, Sep. 1994, pp. 487–499, Accessed: Aug. 02, 2020. [Online].

[8] K. Koperski and J. Han, "Discovery of spatial association rules in geographic information databases," in *Advances in Spatial Databases - 4th International Symposium, SSD 1995, Proceedings*, Jan. 1995, pp. 47–66, doi: 10.1007/3-540-601597_4.

[9] R. Munro, S. Chawla, and P. Sun, "Complex spatial relationships," *Third IEEE Int. Conf. Data Min.*, 2003, doi:10.1109/ICDM.2003.1250924.

[10] Y. Huang, S. Shekhar, and H. Xiong, "Discovering colocation patterns from spatial data sets: a general approach," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 12, pp. 1472–1485, Dec. 2004, doi: 10.1109/TKDE.2004.90.

[11] J. S. Yoo, S. Shekhar, J. Smith, and J. P. Kumquat, "A partial join approach for mining co-location patterns," in *Proceedings of the 12th annual ACM international workshop on Geographic information systems*, Washington DC, USA, Nov. 2004, pp. 241–249, doi: 10.1145/1032222.1032258.

[12] J. S. Yoo, S. Shekhar, and M. Celik, "A join-less approach for co-location pattern mining: A summary of results," in *Proceedings - Fifth IEEE International Conference on Data Mining, ICDM 2005*, Dec. 2005, pp. 813–816, doi: 10.1109/ICDM.2005.8.

[13] S. V. Mehta, S. Sodhani, and D. Patel, "Spatial Co-location Pattern Mining - A new perspective using Graph Database," *ArXiv181009007 Cs*, Oct. 2018, Accessed: Aug. 05, 2020. [Online]. Available: http://arxiv.org/abs/1810.09007.

[14] M. Shazan, M. Jabbar, O. R. Zaïane, and A. Osornio-Vargas, "Discovering spatial contrast and common sets with statistically significant co-location patterns," in *Proceedings of the Symposium on Applied Computing*, Marrakech, Morocco, Apr. 2017, pp. 796–803, doi: 10.1145/3019612.3019665.

[15] S. D. Bay and M. J. Pazzani. Detecting group differences: Mining contrast sets. Data Mining and Knowledge Discovery, 5(3)

[16] Q. Liu, W. Liu, M. Deng, J. Cai, and Y. Liu, "An adaptive detection of multilevel co-location patterns based on natural neighborhoods," *Int. J. Geogr. Inf. Sci.*, pp. 1–26, Jun. 2020, doi: 10.1080/13658816.2020.1775235.

[17] M. Celik, J. M. Kang, and S. Shekhar, "Zonal co-location pattern discovery with dynamic parameters," in *Proceedings of the 7th IEEE International Conference on Data Mining, ICDM 2007*, Dec. 2007, pp. 433–438, doi: 10.1109/ICDM.2007.102.

[18] W. Ding, C. F. Eick, X. Yuan, J. Wang, and J.-P. Nicot, "A framework for regional association rule mining and scoping in spatial datasets," *GeoInformatica*, vol. 15, no. 1, pp. 1–28, Jan. 2011, doi: 10.1007/s10707-010-0111-6.

[19] J. Lu, L. Wang, Y. Fang, and J. Zhao, "Mining strong symbiotic patterns hidden in spatial prevalent co-location patterns," *Knowl.-Based Syst.*, vol. 146, pp. 190–202, 2018.

[20] "Symbiosis," *Wikipedia*. Jun. 14, 2020, Accessed: Aug. 05,2020.[Online].Available:https://en.wikipedia.org/w/index.php?title=Symbiosis&oldid=962566681.

[21] Y. Chen, X. Chen, Z. Liu, and X. Li, "Understanding the spatial organization of urban functions based on co-location patterns
mining: A comparative analysis for 25 Chinese cities," Cities, vol. 97, p. 102563, Feb. 2020, doi:10.1016/j.cities.2019.102563.

[22] M. Manca, L. Boratto, V. M. Roman, O. M. i Gallissà, and A. Kaltenbrunner, "Using social media to characterize urban mobility patterns: State-of-the-art survey and case-study," Online Soc. Netw. Media, vol. 1, pp. 56–69, 2017.

[23] M. Sahana, H. Hong, and H. Sajjad, "Analyzing urban spatial patterns and trend of urban growth using urban sprawl matrix: A study on Kolkata urban agglomeration, India," *Sci. Total Environ.*, vol. 628, pp. 1557–1566, 2018.

[24] F. Goreaud and R. Pélissier, "Avoiding misinterpretation of biotic interactions with the intertype K12-function: population independence vs. random labelling hypotheses," *J. Veg. Sci.*, vol. 14, no. 5, pp. 681–692, 2003, doi: 10.1111/j.16541103.2003.tb02200.x.

[25] P. Phillips and I. Lee, "Mining co-distribution patterns for large crime datasets," *Expert Syst. Appl.*, vol. 39, no. 14, pp. 11556– 11563, Oct. 2012, doi: 10.1016/j.eswa.2012.03.071.

[26] X. Xiao, X. Xie, Q. Luo, and W.-Y. Ma, "Density based colocation pattern discovery," in *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, Irvine, California, Nov. 2008, pp. 1–10, doi: 10.1145/1463434.1463471.

[27] J. S. Yoo, "Spatial query processing and data mining methods for location based services," phd, University of Minnesota, USA, 2007.

[28] S. Shekhar, M. R. Evans, J. M. Kang, and P. Mohan, "Identifying patterns in spatial information: A survey of methods," *WIREs Data Min. Knowl. Discov.*, vol. 1, no. 3, pp. 193–214, 2011, doi: 10.1002/widm.25.

[29] Gorijala Mohana Krishna," Approaches towards Co-location Rule Mining "https://pdfs.semanticscholar.org/3ecd/d2cb0b804441b195b1a63725afc5983f782e.pdf

[30] Wang, L., Bao, Y., Lu, J., Yip, J. (2008). A new join-less approach for co-location pattern mining. In Q. Wu, X. He, Q.V. Nguyen, W. Jia, M.L. Huang (Eds.), CIT (pp. 197–202). Sydney: IEEE.

[31] https://www.toronto.ca/city-government/council/2018-council-issue-notes/

[32] Min Deng, Jiannan Cai, Qiliang Liu, Zhanjun He & Jianbo Tang (2017) Multi-level method for discovery of regional co-location patterns, International Journal of Geographical InformationScience, 31:9, 18461870, DOI: 10.1080/13658816.2017.1334890

About the authors:

Swati Meshram, is a research scholar at Government College of Engineering, Sant Gadge Baba Amravati University. She has received Master's degree in Computer Engineering from University of Mumbai. She is currently working as Assistant Professor in P.G. Department of Computer Science, SNDT Women's University, Mumbai. Her current research interest is Data Mining, Machine Learning and Databases.

Kishor P.Wagh received his Diploma in Computer Engineering from Government Polytechnic, Jalgaon, BE (Computer Science and Engineering) degree from Government College of Engineering, Aurangabad, ME (Computer Science and Engineering) degree from Walchand College of Engineering, Sangali and Ph.D. in Computer Engineering from North Maharashtra University, Jalgaon in 2018. Presently He is working as Assistant Professor in Information Technology Department at Government College of Engineering, Amravati where he has been involved in teaching and research in the field of Data Mining, Machine Learning, Artificial Intelligence and Neural Networks for over a decade. His current research interests are in the areas of Data Mining, Machine Learning, Artificial Intelligence, Neural Networks and Web Mining. He has authored and c o-authored over 30 papers in referred academic journals and national/international conference proceedings.