Filomat 32:5 (2018), 1491–1497
https://doi.org/10.2298/FIL1805491D

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Mining spatial dynamic co-location patterns

Jiangli Duan[a], Lizhen Wang[a], Xin Hu[b], Hongmei Chen[a]

[a] *School of Information Science and Engineering, Yunnan University, Kunming, China*
[b] *College of Information Science and Technology, Beijing Normal University, Beijing, China*

Abstract. Spatial co-location pattern mining is an important part of spatial data mining, and its purpose is to discover the coexistence spatial feature sets whose instances are frequently located together in a geographic space. So far, many algorithms of mining spatial co-location pattern and their corresponding expansions have been proposed. However, dynamic co-location patterns have not received attention such as the real meaningful pattern $\{$*Ganoderma lucidum*$_{new}$, *maple tree*$_{dead}\}$ means that "*Ganoderma lucidum*" grows on the "*maple tree*" which was already dead. Therefore, in this paper, we propose the concept of spatial dynamic co-location pattern that can reflect the dynamic relationships among spatial features and then propose an algorithm of mining these patterns from the dynamic dataset of spatial new/dead features. Finally, we conduct extensive experiments and the experimental results demonstrate that spatial dynamic co-location patterns are valuable and our algorithm is effective.

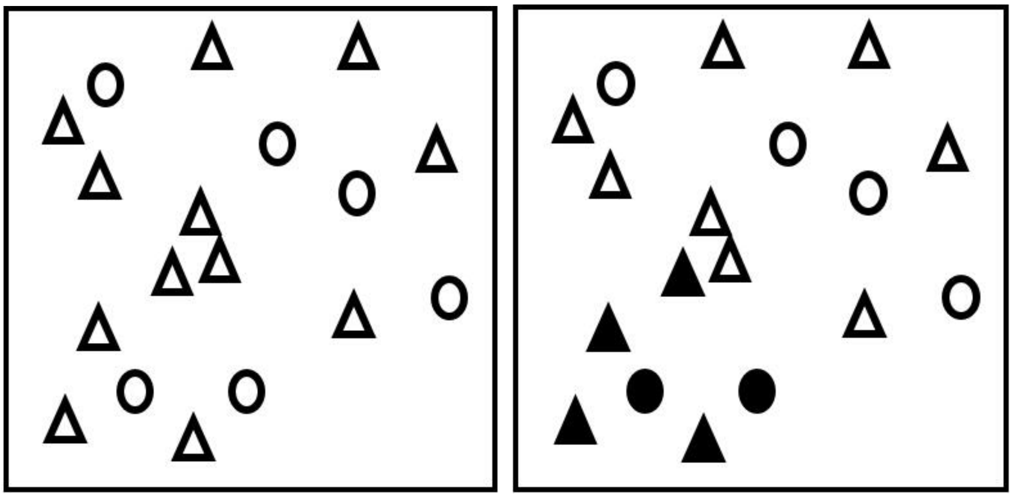
1. Introduction

Spatial co-location pattern mining is an important part of spatial data mining. A spatial co-location pattern represents a subset of spatial features whose instances are frequently located together in a geographic space. For example, if the city planner cannot find the prevalent pattern $\{school, supermarket, restaurant\}$ near the "*school*", it indicates that we need to build new "*supermarket*" or "*restaurant*" appropriately around the "*school*". Other application domains [1–3] include earth science, public health, biology, etc.

So far, various extensions for spatial co-location pattern mining have been proposed, such as spatial co-location pattern mining for probabilistic data [4–6], spatial co-location pattern mining with constraints [7–9, 15], and space-time co-location pattern mining [10, 11], etc. However, nobody has pay attention to the spatial dynamic relationships among the new/dead features that are produced as times goes on. Mining spatial dynamic co-location pattern can remedy the defects of existing frameworks of mining co-location pattern that are as follows:

Problem 1. Existing methods cannot find the relative change in the number of instances of multiple mutually inhibitory features as times goes on.

2010 *Mathematics Subject Classification*. Primary 68T10.
Keywords. Spatial data mining; Co-location pattern; Dynamic co-location pattern.
Received: 18 August 2017; Accepted: 27 October 2017
Communicated by Hari M. Srivastava
Research supported by grants (No.61472346, No.61662086) from the National Natural Science Foundation of China and in part by a grant (No.2015FB114, No. 2016FA026) from the Science Foundation of Yunnan Province.
Corresponding author: Xin Hu
Email addresses: jl_duan@126.com (Jiangli Duan), lzhwang2005@126.com (Lizhen Wang), 201431210011@mail.bnu.edu.cn (Xin Hu), hmchen@ynu.edu.cn (Hongmei Chen)



(a) time point t_0 (b) time point t_1

Figure 1: Sample of Problem 3

Example 1. The literature [12] deems that there is a competitive relationship between "*flaverias*" and "*abutilon theophrasti*", namely, the number of "*abutilon theophrasti*" will be decreased with the increase of the number of "*flaverias*" in same zone. However, because participation index is always unchanged for existing methods, they will get the prevalent pattern {*flaverias*, *abutilon theophrasti*} regardless of the decrease/increase in number of instances.

Problem 2. Existing methods cannot find whether the death/newborn of instances of a feature was caused by the newborn/death of instances of other features.

Example 2. "*Ganoderma lucidum*" grows on the "*maple tree*" which was already dead. However, existing methods mine patterns from the set of living plants. Therefore, the real meaningful pattern {$Ganoderma\ lucidum_{new}$, $maple\ tree_{dead}$} was missed.

Problem 3. Existing methods hold that the instances of multiple features were dead by (approximate) an equivalent percentage in one region that will not affect prevalent patterns.

Example 3. One application of spatial prevalent co-location patterns is that we can judge whether the environment in this region was polluted or not by comparing prevalent patterns at different time points. As shown in Figure 1, the instances of multiple features were dead (black shadow) by an (approximate) equivalent percentage because of environment disruption. However, existing methods believe that the environment has not been polluted because they get the same prevalent patterns with the same participate index at two time points (i.e., t_0 and t_1).

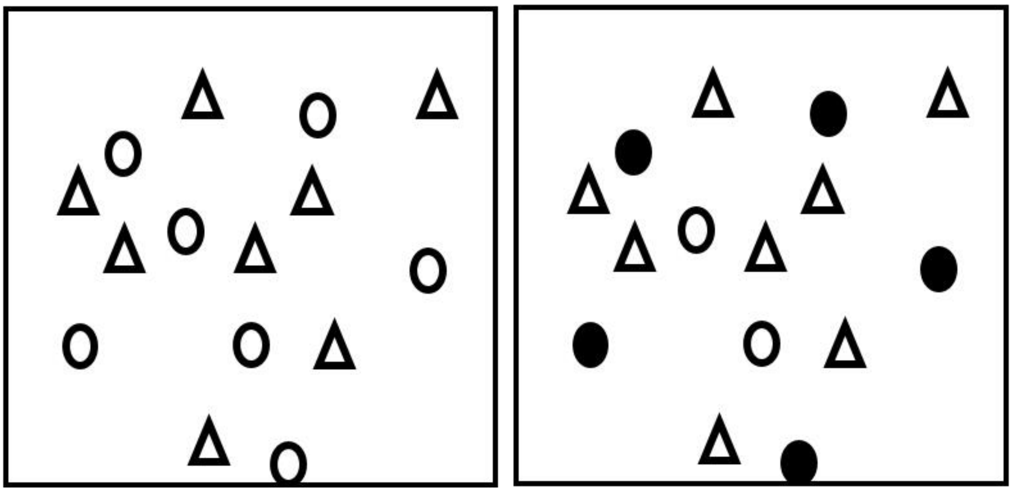
Problem 4. Existing methods hold that the instances of one feature were dead in one region that will not affect prevalent patterns.

Example 4. As shown in Figure 2, some instances of one feature were dead (black shadow) because of unknown factor. However, existing methods get the same prevalent patterns with the same participate index at two time points (i.e., t_0 and t_1).

In conclusion, the existing spatial co-location pattern mining frameworks have not considered the dynamic change of spatial data. Therefore, mining spatial dynamic co-location patterns can remedy the defects of existing frameworks of mining co-location pattern, and the contributions of this paper are as follows:

1. We propose the concept of spatial dynamic co-location pattern that can reflect the dynamic relationships among spatial features.

2. We propose the algorithm of mining spatial dynamic co-location pattern from the dynamic dataset of spatial new/dead features.



(a) time point t_0 (b) time point t_1

Figure 2: Sample of Problem 4

3. We conduct extensive experiments and the results demonstrate that spatial dynamic co-location patterns are valuable and our algorithm is effective.

2. Basic Concept

Definition 1. ***(Spatial Dynamic Feature/Instance):*** *Spatial dynamic feature represents the new/dead object in space, denoted as Df. Spatial dynamic instance is a spatial dynamic feature at a specific location, denoted as Ds.*

Example 5. In Figure 3, the spatial dynamic features A_{dead} and A_{new} belong to the same object, and $\{A_{dead}.1, A_{dead}.2, A_{dead}.3\}$ and $\{A_{new}.1, A_{new}.2, A_{new}.3, A_{new}.4\}$ are dynamic instances of A_{dead} and A_{new} , respectively.

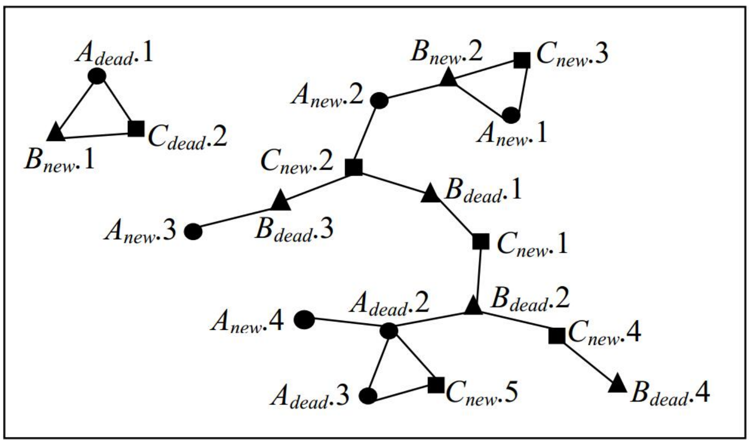


Figure 3: Dynamic features and its instances

Definition 2. ***(Spatial Dynamic Co-Location Pattern):*** *A spatial dynamic co-location pattern that can reflect the dynamic relationships among spatial features, and it contains multiple new/dead features, denoted as Dc. For example, Dc =* $\{A_{dead}, B_{new}, C_{dead}\}$ *is a spatial dynamic co-location pattern.*

Definition 3. ***(Time Span and Time Span Constraint):*** *Time span is the time difference between two adjacent dynamic data sets and represents the time interval in which a number of changes about instances have taken place. Time span constraint of a dynamic feature is the length of time that the feature has an effect on surrounding dynamic features, denoted as span (Df_i). Moreover, if the length of time is equal to k time spans, it can be represented as span (Df_i) = k (time spans).*

If a time span (designated by experts) is m years, we obtain the distribution of instances every m years for the same region. The effect of new dynamic feature on the surrounding dynamic features is the life cycle of this dynamic feature. For example, given a new dynamic feature "*human*", we assume that the time span is 3 years and the life cycle of "*human*" is 75 years, then the time difference between two data sets is 3 years and span ($human_{new}$) =25(time spans).

The time span constraint of dead dynamic feature is one time span because dead feature has little effect on the surrounding dynamic features and one time span can enough to cover the influence time of dead feature to surrounding features.

Definition 4. ***(Dynamic Distance Threshold D_d):*** *When the distance between two dynamic instances is less than D_d (designated by experts), it is considered that they have relationship otherwise they have no relationship.*

Definition 5. ***(Spatial Dynamic Neighbor Relationship D_R):*** *For two dynamic instance, when the distance between them satisfies D_d and the time difference between them satisfies max span, it is considered that they satisfy the spatial dynamic neighbor relationship D_R. The manner of judgement is as follows:*

DR (A_{new}.1, B_{dead}.1) $\Leftrightarrow$ (ΔT (A_{new}.1, B_{dead}.1) < max (span (A_{new}), span (B_{dead}))and distance (A_{new}.1, B_{dead}.1) $\leq$ D_d)

Definition 6. ***(Dynamic Participation Ratio (DPR)/Index (DPI)):*** *DPR(Dc, Df_i) of dynamic feature Df_i in a k-size dynamic co-location Dc = {Df_1, Df_2,. . . ,Df_k} is defined as :*

$$DPR(Dc, Df_i) = \frac{|\pi_{Df_i}(dynamic\_table\_instance(Dc))|}{|dynamic\_table\_instance(Df_i)|}$$

Where π is the relational projection operation with a duplication elimination.*DPI*(*Dc*) of *Dc* is defined as *DPI*(*Dc*)=$\min_1^k$ $DPR(Dc, Df_i)$, if *DPI*(*Dc*) is greater than a given minimum prevalence threshold *min_prev* that be designated by experts and be used to judge whether the pattern occurs frequently or not, we say *Dc* is a prevalent dynamic co-location pattern.

Example 6.In Figure 3,dynamic features B_{dead}/C_{new} have 4/5 dynamic instances, respectively, *min_prev* = 0.4 and the dynamic table instance of *Dc* is {{B_{dead}.1, C_{new}.1}, {B_{dead}.1, C_{new}.2}, {B_{dead}.2, C_{new}.1}, {B_{dead}.2, C_{new}.5}, {B_{dead}.3, C_{new}.2}}. For the dynamic co-location pattern *Dc* = {B_{dead}, C_{new}}, since only B_{dead}.1, B_{dead}.2 and B_{dead}.3 of all the 4 dynamic instances appear in the dynamic table instance so that *DPR* (*Dc*, B_{dead}) = 3/4. Similarly, *DPR* (*Dc*, C_{new}) = 3/5. Finally, *DPI* (*Dc*) = min {*DPR* (*Dc*, B_{dead}), *DPR* (*Dc*, C_{new})} = min {3/4, 3/5} = 3/5 > 0.4. Therefore, *Dc* is a prevalent dynamic co-location pattern.

3. Algorithm

Different from the traditional methods, on the one hand, we mine the spatial dynamic co-location patterns from the new/dead data set. On the other hand, during calculating D_R, we not only consider the distance but also the time span constraint. Based on the two points above, we propose a spatial dynamic co-location algorithm-DJ(Dynamic Join) algorithm that mainly includes four steps: 1) We produce the dynamic instance data set that contains new/dead instances by comparing the data at adjacent two time points (step1). 2) According to the definition of D_R, we produce the dynamic neighbor instance set from the spatial dynamic instance data set (step2). 3) We sort dynamic neighbor instance set (step3) and produce size-2 spatial dynamic co-location prevalent patterns with *min_prev* (step4). 4) We produce prevalent dynamic co-location patterns (join-based [1]) by iterative join process (step 6). The algorithm description is as follows:

- **Algorithm:DJ (Dynamic Join)**

- **Iutput:** $\{\delta_{T1}, \delta_{T2}, ..., \delta_{Tn}\}$: data sets of distributions of spatial features at different points in time
  δ_{life_time}:life cycle of each feature
  D_d:dynamic distance threshold
  min_prev:minimum participation index threshold

- **Output:** $\delta_{frequent} = \{\delta_{frequent\_size\_1}, \delta_{frequent\_size\_2}, ..., \delta_{frequent\_size\_n}\}$

1. δ_D=generate_D($\{\delta_{T1}, \delta_{T2}, ..., \delta_{Tn}\}$);
2. δ_R=generate_R($\delta_D = \{\delta_{\Delta T1}, \delta_{\Delta T2}, ..., \delta_{\Delta Tn-1}\}, \delta_{life\_time}, D_d$);
3. $\delta_{size2\_candidate}$=sort_and_generate_size2_candidate(δ_R);
4. $\delta_{frequent\_size2}$=generate_frequent_size2($\delta_{size2\_candidate}$,min_prev);
5. k=2;
6. while(not_empty($\delta_{frequent\_size\_k}$))
   6.1. $\delta_{candidate\_size\_k+1}$=generate_candidate_size_k+1($\delta_{frequent\_size\_k}$,k);
   6.2. δ_{size_k+1}=pruning_size_k+1($\delta_{candidate\_size\_k+1}$);
   6.3. $\delta_{frequent\_size\_k+1}$=generate_frequent_size_k+1(δ_{size_k+1},min_prev);
   6.4. k=k+1;

4. Experimental Analysis

There are two kinds of data sets: a real data set that consists of valuable infrastructures in a city, and two synthetic data sets (uniform/non-uniform distribution data).

The main purpose of this paper is to mine the spatial dynamic co-location patterns so that the efficiency of the algorithm is not our focus. In addition, there is no other researches about spatial dynamic co-location pattern mining, so there is no suitable algorithm to compare with our algorithm. Therefore, we only need to verify the effectiveness and scalability of our algorithm.

4.1. The Effectiveness of DJ over the Real Data Sets

We perform the traditional method join-based [1] and DJ (dynamic algorithm) over the real data set. By default, the D_d and min_prev are 1km and 0.4, respectively. Traditional method can find that "*school*", "*park*", "*hotel*", "*bank*", "*hospital*", "*supermarket*" and "*KTV*" are always coexistence for a long period, and the result is not meaningful.

In contrast, from the experimental results of DJ in Table 1, we can draw the following information:

1) The increase (decrease) in "*school*", "*hotel*", "*hospital*" and "*KTV*" will affect the increase (decrease) in "*bank outlet*".

2) Life-services like "*hospital*" and "*supermarket*" have a mutual exclusion relationship with entertainments such as "*KTV*". That is, the increase in "*hospital*", "*supermarket*" usually occurs with the decrease in "*KTV*" that represents the adjustment of urban regional structures. We can produce these patterns represents that we can solve Problem 1 and Problem 2.

3) "*Hotel*", "*KTV*" and "*bank*" always appear/disappear simultaneously and then we can conclude that they have strong symbiotic relationships. Simultaneous increase/decrease reflects the economic prosperity/recession in this region because they stand for the level of regional economic development. We can produce these patterns represents that we can solve the Problem 3.

4) From {$school_{new}, KTV_{new}$}/ {$school_{new}, KTV_{dead}$}, "*School*" has a positive and negative effect on "*KTV*" that looks like contradictory. In fact, after making a careful analysis, we find that "*college*" has a positive effect on "*KTV*"while "*primary school*" and "*secondary school*" have a negative effect on "*KTV*" as shown in Table 2. (In order to protect data privacy, we omit the specific name of the "*school*" and "*KTV*").

For Problems 1, 2, 3 in Introduction, it is obvious that the DJ algorithm can solve these problems. Although the problem 4 is not direct reflected from the results, but DJ algorithm can produce the pattern {B_{dead}} is prevalent and {A_{new}, B_{dead}}/ {A_{dead}, B_{dead}} are not prevalent (because there are only few instances of A_{new} and A_{dead}), so that we can consider that traditional pattern {A, B} is a non-coexistence relationship. Thus, this article can avoid the error in Problem 4.

In conclusion, the DJ algorithm can solve these problems in Introduction. In addition, two very interesting results were obtained, that is, the result 1) fully reflects the interaction among the features and result 4) brings us unexpectedly patterns.

Spatial dynamic prevalent co-location patterns			
	new, new	dead, dead	new, dead
Size-2	$school_{new}, supermarket_{new}$ $school_{new}, bank_{new}$ $school_{new}, KTV_{new}$ $KTV_{new}, bank_{new}$ $hospital_{new}, bank_{new}$ $hospital_{new}, supermarket_{new}$ $hotel_{new}, bank_{new}$ $hotel_{new}, KTV_{new}$ $part_{new}, hotel_{new}$	$school_{dead}, bank_{dead}$ $hospital_{dead}, supermarket_{dead}$ $hotel_{dead}, bank_{dead}$ $hotel_{dead}, KTV_{dead}$ $bank_{dead}, KTV_{dead}$	$hotel_{dead}, supermarket_{new}$ $supermarket_{new}, KTV_{dead}$ $hospital_{new}, KTV_{dead}$ $school_{new}, KTV_{dead}$
Size-3	$hotel_{new}, KTV_{new}, bank_{new}$	$hotel_{dead}, KTV_{dead}, bank_{dead}$	$hotel_{dead}, supermarket_{new}, KTV_{dead}$

Table 1: Prevalent patterns of urban unit data sets

$school_{new}, KTV_{new}$	$school_{new}, KTV_{dead}$
$**$ *junior college*$_{new}$,$**$ KTV_{new}	$**$ *kindergarten*$_{new}$, $**$ KTV_{dead}
$**$ *technical college*$_{new}$,$**$ KTV_{new}	$**$ *primary school*$_{new}$, $**$ KTV_{dead}
$**$ *college* $**$*institute*$_{new}$,$**$ KTV_{new}	$**$ *attached middle school*$_{new}$, $**$ KTV_{dead}
$**$ *college*$_{new}$, $**$ $KTV_{new} \ldots$	$**$ *attached primary school*$_{new}$, $**$ $KTV_{dead} \ldots$

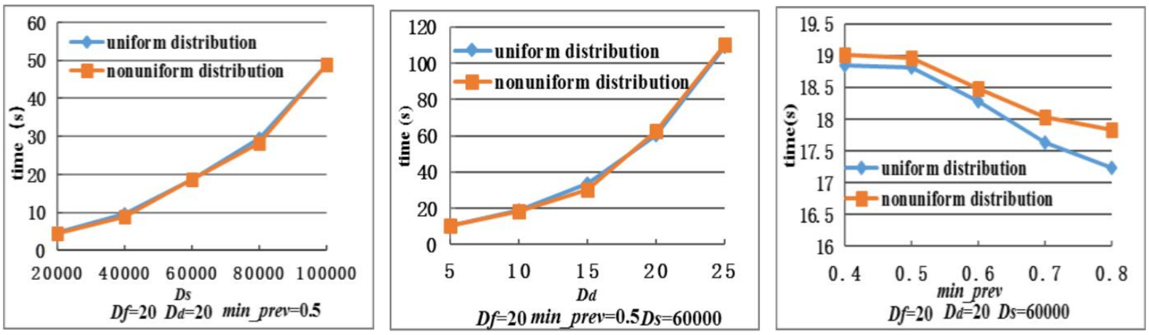
Table 2: Dynamic table instance of "*school*" and "*KTV*"

4.2. Scalability of DJ

In this part, in order to test the scalability of DJ algorithm, we compare the experimental results over two data sets (uniform/non-uniform distributed data sets) as shown in Figure 4. From the experimental results in Figure 4, we can draw the following conclusion: there is no significant difference about the efficiency of the algorithm over uniform/non-uniform distributed data sets, that is, the spatial dynamic co-location pattern mining algorithm DJ has perfect scalability.

5. Conclusion

On the one hand, different from current space-time data mining technology[13] that discovers co-occurrence patterns at the same time and space, dynamic spatial co-location patterns consist of the new/dead features which are co-occurrence at the same space and co-occurrence or sequence over time. Therefore, we cannot use the existing space-time data mining technology. On the other hand, different from some traditional spatial co-location pattern mining algorithms of mining positive and negative pattern[14] that



(a) (b) (c)

Figure 4: Scalability of DJ algorithm

represents coexistence and competitive relationship among features respectively, we can extract the interaction relationships among features by analyzing the dynamic changes of features.

In order to solve the problems that are defect of existing co-location pattern mining frameworks like Problem 1,2,3,4 in Introduction and show the interaction relationships among spatial features better, we propose the concept of spatial dynamic co-location patterns and the algorithm of mining these patterns. In the future, we will consider designing more efficient algorithm and doing a more in-depth analysis for spatial dynamic co-location patterns.

References

[1] Y. Huang, S. Shekhar, H. Xiong, Discovering colocation patterns from spatial data sets: a general approach, IEEE Trans. Knowl. Data Eng. 16 (2004), 1472-1485.

[2] F. Flouvat , J.F.N.V. Soc, E. Desmier, N. Selmaoui-Folcher, Domain-driven co-location mining Extraction, visualization and integration in a GIS, Geoinformatica.19(2015), 147-183.

[3] S.K. Kim , J.H. Lee, K.H. Ryu, U. Kim, A framework of spatial co-location pattern mining for ubiquitous GIS, Multimed. Tools Appl, 71(2014), 99-218.

[4] Z. Ouyang, L. Wang, Mining Spatial Co-Location Patterns for Fuzzy Object, Chinese journal of computers 34 (2011), 1947-1955.

[5] L. Wang, J. Han, H. Chen, J. Lu, Top-k probabilistic prevalent co-location mining in spatially uncertain data sets, Front.. Comput. Sci, 10 (2016), 488-503.

[6] L. Wang, P. Wu, H. Chen. Finding probabilistic prevalent co-locations in spatially uncertain data sets, IEEE Trans. Knowl. Data Eng, 25 (2013), 790-804.

[7] S.Y. Jin, M. Bow, Mining maximal co-located event sets, Proceedings of the Advances in knowledge discovery and data mining, 2011, 351-362.

[8] M. Shazan, M. Jabbar, O.R .Za?ane, A. Osornior-Vargas, Discovering spatial contrast and common sets with statistically significant co-location patterns, Proceedings of the Symposium on Applied Computing, 2017, 796-803.

[9] J. Li, O. R. Za?ane, A. Osornio-Vargas, Discovering statistically significant co-location rules in datasets with extended spatial objects, Proceedings of the International Conference on Data Warehousing and Knowledge Discovery, 2014,124-135.

[10] B. Aydin, D. Kempton ,V. Akkineni, R.Angryk, K.G. Pillai, Mining spatiotemporal co-occurrence patterns in solar datasets , Astron. Comput, 13 (2015), 136-144.

[11] M. Celik. Partial spatio-temporal co-occurrence pattern minin,Knowl. Inf. Syst, 44 (2015), 27-49.

[12] L. Yuan, G.Q. Wang, L. Zheng, N.I Han-Wen. Competitiveness of invasive plant Flaveria bidentis with native weed plants, Chinese Journal of Ecology, 30 (2011), 677-681.

[13] D. LiuH. ChenH. Qi, B. Yang, Advances in Spatiotemporal Data Mining, Journal of Computer Research and Development, 50 (2013), 225-239.

[14] Y. Jiang, L. Wang, Y. Lu, H. Chen, Discovering both positive and negative co-location rules from spatial data sets, Proceedings of the Software Engineering and Data Mining, 2010,398-403.

[15] F. Qian, Q. He, J. He, Mining Spatial Co-location Patterns with Dynamic Neighborhood Constrain, Proceedings of the Machine Learning and Knowledge Discovery in Databases, 2009, 238-253.