



Continuous Sub-prevalent Co-location Pattern Mining

Qilong Wang¹, Hongmei Chen^{1,2}(✉), and Lizhen Wang^{1,2}

¹ School of Information Science and Engineering, Yunnan University,
Kunming, China
hmchen@ynu.edu.cn

² Yunnan Key Laboratory of Intelligent Systems and Computing, Yunnan University,
Kunming, China

Abstract. Spatial co-location pattern represents a subset of spatial features whose instances are frequently located together in space. Sub-prevalent co-location pattern mining discovers patterns with richer spatial relationships based on star instance model instead of clique instance model. Further, discovering spatiotemporal sub-prevalent co-location pattern is important to reveal the spatiotemporal interaction between spatial features and promote the application of patterns. However, the methods for mining spatiotemporal sub-prevalent co-location pattern measure the interestingness of patterns by the frequency of patterns in time slice set, and ignore the duration of patterns which is an important spatiotemporal information in patterns. Thus, this paper presents mining spatiotemporal sub-prevalent co-location pattern by considering the duration and the frequency of patterns. Specifically, a novel pattern, is proposed by defining the continuous sub-prevalent index. Then, an efficient algorithm is designed to mine the proposed patterns by utilizing the anti-monotonicity of continuous sub-prevalent index to prune unpromising patterns. Extensive experiments on synthetic and real datasets verify the practicability of the proposed patterns and the effectiveness of the proposed algorithm.

Keywords: Spatiotemporal data mining · Spatial sub-prevalent co-location pattern · Spatiotemporal sub-prevalent co-location pattern

1 Introduction

With the rapid development of spatial information technology such as the global positioning system, spatial data has shown an explosive growth. Spatial co-location pattern mining is an important branch of spatial data mining, which has draw the attention of researchers due to practicality of co-location patterns in environmental protection [1], public security [2] and public health [3]. A prevalent co-location pattern is a subset of spatial features whose instances are frequently collocated within a neighbourhood. For example, that Egyptian Plover occur frequently in near Nile Crocodile can be expressed as a prevalent co-location pattern {Egyptian Plover, Nile Crocodile}.

Prevalent co-location pattern mining determines the neighbor relationship of spatial instances through the distance threshold, generates all row instances of patterns based on clique instance model (i.e., a row instance forms a clique), calculates the participation indices of patterns on the basis of row instances, and finally discovers all prevalent patterns through the participation index threshold [4]. In order to capture richer spatial relationships, sub-prevalent co-location pattern and star participation index based on star instance model are proposed, which loosen the clique constraint of spatial instances in a row instance.

Both prevalent co-location pattern and sub-prevalent co-location pattern ignore the time factor of patterns, i.e., patterns vary with time. For example, seagulls migrate to lakes in Yunnan to spend the cold winter, and leave from lakes when spring comes. This shows the pattern {Seagull, Lake} change with the season.

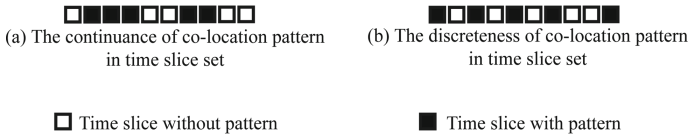


Fig. 1. The distribution of co-location patterns in time slice set

To mine time-varying patterns, spatiotemporal co-location pattern mining was introduced. Celik et al. [5] mine co-location patterns on each time slice and finds patterns which appear on many time slices. Li et al. [6,7] mine sub-prevalent co-location patterns existing on many time slices. These spatiotemporal co-location patterns consider the frequency of patterns on time slices. However, besides the frequency of patterns, we argue that the duration of patterns is also important. Let us see Fig. 1. The frequencies of pattern in Fig. 1(a) and Fig. 1(b) are the same (i.e., 50%), but the duration of pattern in Fig. 1(a) is longer than that in Fig. 1(b). This implies the pattern lasts for a period of time besides repeatedly appearing. This kind of patterns is also meaningful. For instance, instead of passing by, seagulls spending the cold winter on lakes may be more meaningful for environmental protection.

Motivated by the above, we propose continuous sub-prevalent co-location pattern by taking into account the duration and the frequency of sub-prevalent co-location pattern. The main contributions of the paper are as follows:

- We define the continuous sub-prevalent index to measure pattern, and propose the novel continuous sub-prevalent co-location pattern.
- We prove the anti-monotonicity of the proposed measure, and design an efficient algorithm to mine the proposed patterns by utilizing the anti-monotonicity.
- We conduct extensive experiments on real and synthetic data sets. The experimental results show that the proposed pattern is practical and the proposed algorithm is efficient.

The rest of the paper is organized as follows. We review the related work in Sect. 2. Section 3 introduces preliminaries and defines the proposed continuous sub-prevalent co-location pattern. Section 4 describes the algorithm to mine the proposed patterns. Section 5 presents the experimental evaluation and Sect. 6 concludes the paper.

2 Related Work

2.1 Prevalent Co-location Pattern

Huang et al. [4] defined the prevalent co-location pattern based on clique instance model and proposed the Join-based mining algorithm, studies have proposed optimization algorithms [8–12] to solve the low efficiency issues of the Join-based algorithm.

Due to the impact of time factor on the co-location pattern, Celik et al. [5] analyzes the time-varying co-location pattern, and proposed a mixed spatiotemporal MDCOPs algorithm. Andrzejewski and Boinski [13] proposed the MAXMDCOP-Miner algorithm to solve the low efficiency issue of the MDCOPs algorithm. Qian et al. [14] proposed a sliding window model, which introduces the impact of event time intervals into the index to measure the spatiotemporal co-location pattern. Ma et al. [15] proposed a two-step framework to mine evolving pattern over time. Yang and Wang [16] proposed a spatiotemporal co-location congestion pattern mining method to discover the orderly set of roads with congestion propagation in urban traffic.

2.2 Sub-prevalent Co-location Pattern

To mine co-location pattern with richer spatial relationship, Wang et al. [17, 18] proposed the sub-prevalent co-location pattern based on star instance model which loosens the clique constraint of spatial instances in a row instance, and designed the PTBA and PBA algorithms to mine sub-prevalent co-location patterns. Ma et al. [19] proposed the sub-prevalent co-location pattern with dominant feature. Xiong et al. [20] presented mining fuzzy sub-prevalent co-location pattern with dominant feature.

Taking into account the importance of time factor, Li et al. [6, 7] proposed the spatiotemporal sub-prevalent co-location pattern by considering the frequency of pattern.

Distinct from the spatiotemporal sub-prevalent co-location pattern in [6, 7], our proposed continuous sub-prevalent co-location pattern not only considers the frequency of pattern but also the duration of pattern.

3 Preliminaries and Problem Definition

3.1 Spatial Sub-prevalent Co-location Pattern

Let $F = \{f_1, f_2, \dots, f_n\}$ be the set of spatial features in a spatial dataset, $S = S_1 \cup S_2 \cup \dots \cup S_n$ be the set of spatial instances where S_i is the instance set of f_i ,

and d be a user-specified distance threshold. For two instances $i_i, i_j \in S$, if the Euclidean distance between them satisfies $\text{distance}(i_i, i_j) \leq d$, they satisfy the neighbor relationship $R(i_i, i_j)$. Figure 2(a) shows a spatial dataset, where $F = \{A, B, C, D\}$, $S = \{A.1, A.2, A.3, B.1, B.2, B.3, C.1, C.2, C.3, D.1, D.2\}$, and the neighbor relationships between instances are expressed by lines. A sub-prevalent co-location pattern c is a subset of the feature set, i.e., $c \subseteq F$, and the number of features in c is called the size k of c , i.e., $k = |c|$. The related definitions of sub-prevalent co-location pattern [17, 18] is as follows.

Definition 1. (*Star Neighborhoods Instance, SNsI*). The set of star neighborhoods instances of an instance $i_j \in S$ is defined as $SNsI(i_j) = \{i_k | \text{distance}(i_j, i_k) \leq d, i_k \in S\}$.

Definition 2. (*Star Participation Instance, SPIns*). The star participation instance of a feature $f_i \in c$ is defined as $SPIns(f_i, c) = \{i_j | i_j \in f_i \text{ and the feature set of } SNsI(i_j) \text{ contains all features in } c\}$.

In Fig. 2(a), $SNsI(A.1) = \{A.1, B.2, C.3\}$, $SPIns(A, \{A, B, C\}) = \{A.1, A.2\}$

Definition 3. (*Star Participation Ratio, SPR*). The star participation ratio of a feature $f_i \in c$ is defined as the ratio of the number of star participation instances of f_i to the number of instances of f_i :

$$SPR(f_i, c) = |SPIns(f_i, c)| / |S_i| \quad (1)$$

Definition 4. (*Star Participation Index, SPI*). The star participation index of a pattern c is defined as the minimum of the star participation rates of all features in c :

$$SPI(c) = \min_{f_i \in c} \{SPR(f_i, c)\} \quad (2)$$

Definition 5. (*Sub-prevalent Co-location Pattern, SCP*). Given a user-specified star participation index threshold θ , if $SPI(c) \geq \theta$, the pattern c is called a sub-prevalent co-location pattern.

In Fig. 2(a), for a pattern $\{A, B, C\}$, $SPR(A, \{A, B, C\}) = 0.67$, $SPR(B, \{A, B, C\}) = 0.67$, $SPR(C, \{A, B, C\}) = 0.67$ and $SPI(\{A, B, C\}) = \min\{0.67, 0.67, 0.67\} = 0.67$, if $\theta = 0.5$, the pattern $\{A, B, C\}$ is a SCP.

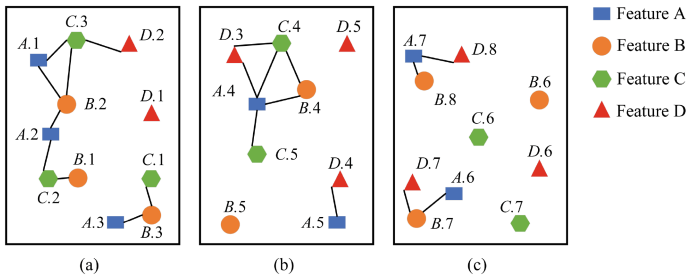


Fig. 2. A spatiotemporal dataset with three time slices

3.2 Continuous Sub-prevalent Co-location Pattern

Distinct from mining sub-prevalent co-location patterns from a spatial dataset, we propose mining continuous sub-prevalent co-location patterns from a spatiotemporal dataset by taking into account the duration and the frequency of patterns.

Let $T = \{t_1, t_2, \dots, t_m\}$ be the set of time slices in a spatiotemporal dataset, $F = F^1 \cup F^2 \cup \dots \cup F^m$ be the set of spatial features where $F^i = \{f_1^i, f_2^i, \dots, f_n^i\}$ is the feature set on time slice t_i , $S = S^1 \cup S^2 \cup \dots \cup S^m$ be the set of spatial instances where $S^i = S_1^i \cup S_2^i \cup \dots \cup S_n^i$ be the instance set on t_i and S_j^i is the instance set of f_j on t_i . Figure 2 shows a spatiotemporal dataset with 3 time slices each of which is Fig. 2(a), Fig. 2(b) and Fig. 2(c) respectively.

We define the related concepts of continuous sub-prevalent co-location pattern as follows.

Definition 6. (*Duration time slices, DTS*). Given a sub-prevalent co-location pattern c , if it appears on time slice t_a , disappears on time slice t_d , and keeps on time slices $t_i (t_a \leq t_i < t_d)$, then a set of duration time slices $DTS(c)$ of c is defined as:

$$DTS(c) = \{t_a, t_{a+1}, \dots, t_{d-1}\} \quad (3)$$

Intuitively, a pattern may have multiple sets of duration time slices which can be noted by $DTSs(c) = \{DTS^1(c), DTS^2(c), \dots, DTS^l(c)\}$.

Definition 7. (*Continuous sub-prevalent index, CSI*). Given a sub-prevalent co-location pattern c , the continuous sub-prevalent index $CSI(c)$ of c is defined as:

$$CSI(c) = \frac{\sum_{i=1}^{|DTSs(c)|} |DTS^i(c)| * \frac{|DTS^i(c)|}{|T|}}{Non(c) * \frac{1}{|T|} + \sum_{i=1}^{|DTSs(c)|} |DTS^i(c)| * \frac{|DTS^i(c)|}{|T|}} \quad (4)$$

Where $\frac{|DTS^i(c)|}{|T|}$ is the weight of duration time slice set $DTS^i(c)$ of c , $Non(c)$ is the number of time slices without c , i.e., $Non(c) = |T - \bigcup_{i=1}^{|DTSs(c)|} DTS^i(c)|$, and $\frac{1}{|T|}$ is the weight of time slices without c .

Definition 8. (*Continuous sub-prevalent co-location pattern, CSCP*). Given a sub-prevalent co-location pattern c and a user-specified continuous sub-prevalent index threshold ρ , if $CSI(c) \geq \rho$, c is called a continuous sub-prevalent co-location pattern.

In Fig. 2, $T = \{t_1, t_2, t_3\}$, if $\{A, B, C\}$ is a sub-prevalent co-location pattern on time slice t_1 and t_2 , then $DTSs(\{A, B, C\}) = \{\{t_1, t_2\}\}$, $CSI(\{A, B, C\}) = \frac{2 * \frac{2}{3}}{1 * \frac{1}{3} + 2 * \frac{2}{3}} = 0.8$. if $\rho = 0.5$, then $\{A, B, C\}$ is a CSCP.

Problem Statement. Given a spatiotemporal dataset with the time slice set T , the spatial feature set F , the spatial instance set S , a distance threshold d , a star participation index threshold θ , and a continuous sub-prevalent index threshold ρ , continuous sub-prevalent co-location pattern mining is to find all CSCPs that satisfy all thresholds.

4 Mining Algorithm

Starting from size-2 patterns, all CSCPs can be mined level-by-level. To improve the efficiency of the algorithm, the anti-monotonicity of continuous sub-prevalent index will be proven, and be used for pruning unpromising patterns.

Lemma 1. (*Anti-monotonicity of continuous sub-prevalent index*). Let c, c' be two co-location patterns such that $c \subset c'$, then $CSI(c) \geq CSI(c')$.

Proof. (1) According the anti-monotonicity of star participation index [17, 18], For $\forall t \in T$, $SPI^t(c) \geq SPI^t(c')$ holds. If c' is a SCP ($SPI^t(c') \geq \theta$), c is also a SCP ($SPI^t(c) \geq \theta$).

(2) Suppose $DTSs(c) = \{DTS^1(c), \dots, DTS^l(c)\}$, $DTSs(c') = \{DTS^1(c'), \dots, DTS^l(c')\}$. According to (1), for $\forall DTS^{i'}(c') \in DTSs(c')$, $\exists DTS^i(c) \in DTSs(c)$, $DTS^{i'}(c') \subseteq DTS^i(c)$ holds.

Suppose $DTS^{i_1'}(c'), \dots, DTS^{i_{l'}'}(c') \subseteq DTS^i(c)$. For $\forall i_1' \leq i', j' \leq i_{l'}', i' \neq j'$, $DTS^{i_1'}(c') \cap DTS^{j'}(c') = \emptyset$ and $\bigcup_{i_1'=i_1'}^{i_{l'}'} DTS^{i_1'}(c') \subseteq DTS^i(c)$ hold.

Then, we have $\sum_{i_1'=i_1'}^{i_{l'}'} |DTS^{i_1'}(c')|^2 \leq |DTS^i(c)|^2$. Further, we have $\sum_{i=1}^{|DTSs(c)|} |DTS^i(c')|^2 \leq \sum_{i=1}^{|DTSs(c)|} |DTS^i(c)|^2$. On the other hand, we have $Non(c') \geq Non(c)$.

So, we have $\frac{Non(c)}{\sum_{i=1}^{|DTSs(c)|} |DTS^i(c)|^2} + 1 \leq \frac{Non(c')}{\sum_{i=1}^{|DTSs(c')|} |DTS^i(c')|^2} + 1$, i.e.,

$\frac{\sum_{i=1}^{|DTSs(c)|} |DTS^i(c)|^2}{Non(c) + \sum_{i=1}^{|DTSs(c)|} |DTS^i(c)|^2} \geq \frac{\sum_{i=1}^{|DTSs(c')|} |DTS^i(c')|^2}{Non(c') + \sum_{i=1}^{|DTSs(c')|} |DTS^i(c')|^2}$. Further, we have

$\frac{\sum_{i=1}^{|DTSs(c)|} |DTS^i(c)| \cdot \frac{|DTS^i(c)|}{|T|}}{Non(c) * \frac{1}{|T|} + \sum_{i=1}^{|DTSs(c)|} |DTS^i(c)| * \frac{|DTS^i(c)|}{|T|}} \geq \frac{\sum_{i=1}^{|DTSs(c')|} |DTS^i(c')| * \frac{|DTS^i(c')|}{|T|}}{Non(c') * \frac{1}{|T|} + \sum_{i=1}^{|DTSs(c')|} |DTS^i(c')| * \frac{|DTS^i(c')|}{|T|}}$,
i.e., $CSI(c) \geq CSI(c')$. \square

Pruning Strategy 1. If $c \subset c'$ and $CSI(c) < \rho$, then $CSI(c') < \rho$, i.e., c and its all supersets are not CSCPs, and can be pruned.

Based on lemma 1 and pruning strategy 1, we propose an efficient Continuous Sub-prevalent Co-location Pattern Mining Algorithm (CSCPMA) for mining all CSCPs. Generally, the CSCPMA adopts the level-by-level and generation-and-test method. Starting from size-2 patterns, it generates size- k candidate patterns based on size- $(k-1)$ patterns, and tests size- k candidate patterns to get size- k patterns.

Initialization (Step 1–4): The star neighbors instance set of each spatial instance in each time slice $t \in T$ is generated by Gen_SNsI(). All size-1 patterns, i.e., a subset of F , are the beginning the of the iterations(Step 5–13).

Generating Candidate Patterns and Sub-prevalent Patterns (Step 6–9): Size- k candidate patterns are generated by calling Gen_CCSCP(). In Gen_CCSCP(), size- k candidate patterns are generated based on the set

Algorithm 1. CSCPMA

Input :

T : a set of time slices
 F : a set of spatial features
 S : a set of spatial instances
 d : a distance threshold
 θ : a star participation index threshold
 ρ : a continuous sub-prevalent index threshold

Output :

$CSCP$: a set of CSCPs

Variables :

k : a co-location size
 SCP_k : a set of size- k SCPs
 $CCSCP_k$: a set of size- k candidate patterns
 $CSCP_k$: a set of size- k CSCPs

Algorithm :

```

1: for each  $t \in T$  do
2:    $SNsI^t = \text{Gen\_SNsI}(F, S, d)$ 
3: end for
4:  $k = 2; SCP_1 = F; CSCP_1 = \text{Gen\_CSCP}(\rho, SCP_1)$ 
5: while  $CSCP_{k-1} \neq \emptyset$  do
6:    $CCSCP_k = \text{Gen\_CCSCP}(CSCP_{k-1})$ 
7:   for each  $t \in T$  do
8:      $SCP_k = SCP_k \cup \text{Gen\_SCP}(CCSCP_k, SNsI^t, \theta)$ 
9:   end for
10:   $CSCP_k = \text{Gen\_CSCP}(\rho, SCP_k)$ 
11:   $CSCP = CSCP \cup CSCP_k$ 
12:   $k = k + 1$ 
13: end while
14: return  $CSCP$ 

```

$CSCP_{k-1}$ of size- $(k-1)$ CSCPs according to Lemma 1 and Pruning Strategy 1. Then, In $\text{Gen_SCP}()$, these candidate patterns are tested for generating size- k SCPs.

Generating Continuous Sub-prevalent Co-location Patterns (Step 10):

For each size- k SCP, $\text{Gen_CSCP}()$ is called to calculate the continuous sub-prevalent index of the pattern according Definition 7. If the pattern satisfies the threshold ρ , it is a size- k CSCP.

The time cost of the CSCPMA mainly are Step 2, Step 8 and Step 10. The time complexity of Step 2 for generating star neighbor instance sets is $O(|T||S^i|^2)$, the time complexity of Step 8 for generating SCPs is $O(|T| \sum_{k=1} |C_k||S_j^i|^k)$, the time complexity of Step 10 for generating CSCP is $O(|T| \sum_{k=1} |SCP_k|)$, where C_k is the set of size- k candidate patterns and SCP_k is the set of size- k SCPs.

5 Experimental Results and Analysis

In this section, we conduct experiments to evaluate the efficiency of the proposed CSCPMA on synthetic datasets and compare the mining results with the STSCPMA [6, 7] on real datasets.

Datasets. We analyze the patterns by two algorithms on two real datasets and evaluate the efficiency of two algorithms by 11 synthetic datasets. The synthetic datasets are randomly generated according to Poisson distribution with different number of features, the number of instances, the number of time slices or the range. The real datasets include the Tokyo-Checkin containing 14 POI types (features) and 115202 checkin (instances), and the ChicagoCrime dataset containing 19 crime types (features) and 373933 crime events (instances). The datasets are described in Table 1. The default parameters of both algorithms are shown in Table 2.

Environment Setting. Both the CSCPMA and the STSCPMA are coded in Python and run on a PC with Intel Core i7 CPU, 32 GB RAM and Windows 10.

Table 1. Experimental datasets

Data Sets	Number of Features	Number of Instances	Number of Time slices	Range
Tokyo-Checkin	14	115202	11	40000x40000
ChicagoCrime	19	373933	24	40000x40000
Synthetic 1/2/3/4	10/15/20/25	8×10^4	10	10000x10000
Synthetic 5/6/7	20	$(7/9/10) \times 10^4$	10	10000x10000
Synthetic 8/9/10	20	10×10^4	10/15/20/25	10000x10000
Synthetic 11	25	8×10^4	10	12000x12000

Table 2. The default parameters of CSCPMA and STSCPMA

Name	d	θ	ρ	TSF [6, 7]
Tokyo-Checkin	1300	0.4	0.6	0.6
ChicagoCrime	800	0.4	0.6	0.6
Synthetic 1-11	350	0.4	0.6	0.6

5.1 Efficiency Analysis

Effect of Number of Features. We compare the running time of two algorithms on Synthetic 1–4 under different number of features. From Fig. 3 we can see that the running time of both algorithms increases as the number of features varies from 10 to 15, then the running time decreases when varying from 15 to 25. The reason may be that candidate patterns are generated and tested in plenty as the number of features varies from 10 to 15, then unpromising patterns are pruned effectively when varying from 15 to 25. The running time of CSCPMA is slightly higher than that of STSCPMA because CSCPMA spends more time discovering patterns that STSCPMA cannot find.

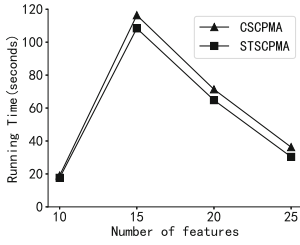


Fig. 3. Effect of number of features

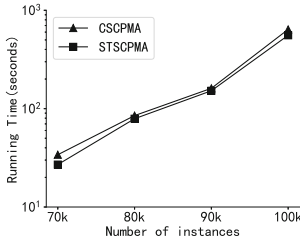


Fig. 4. Effect of number of instances

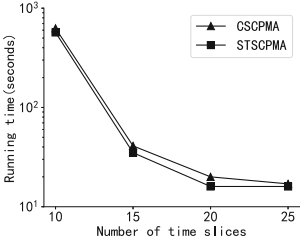
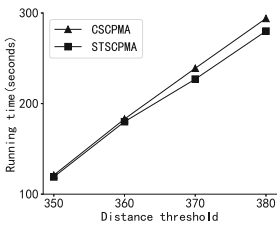


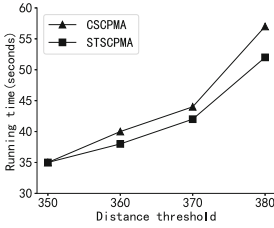
Fig. 5. Effect of number of time slices

Effect of Number of Instances. We compare the running time of two algorithms on Synthetic 3 and 5–7 under different number of instances. In Fig. 4, as the number of instances gradually increases, the running time of both algorithms also increases. This is because more and denser instances lead to more star neighbors and patterns generated.

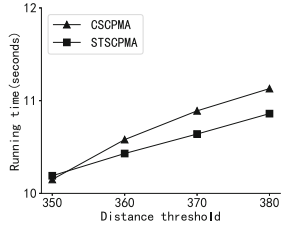
Effect of Number of Time Slices. We compare the running time of two algorithms on Synthetic 7–10 under different number of time slices. In Fig. 5, the running time of both algorithms decreases as the number of time slice increases, The reason is that sparser instances result in less patterns generated and tested.



(a) Synthetic 2



(b) Synthetic 4



(c) Synthetic 11

Fig. 6. Effect of distance threshold

Effect of Distance Threshold. We compare the running time of two algorithms on Synthetic 2,4,11 under different distance thresholds. In Fig. 6, as distance thresholds gradually increases, the running time of both algorithms also increases. The reason is that more instances satisfy the neighbor relationship when increasing distance threshold, which leads to more star neighbors and more patterns.

Effect of Star Participation Index Threshold. We compare the running time of two algorithms on Synthetic 2,4,11 under different star participation index thresholds. In Fig. 7, both algorithms show a gradual decrease in running time as the star participation index threshold increases. This is because the larger star participation index threshold will lead to the fewer patterns satisfying the threshold.

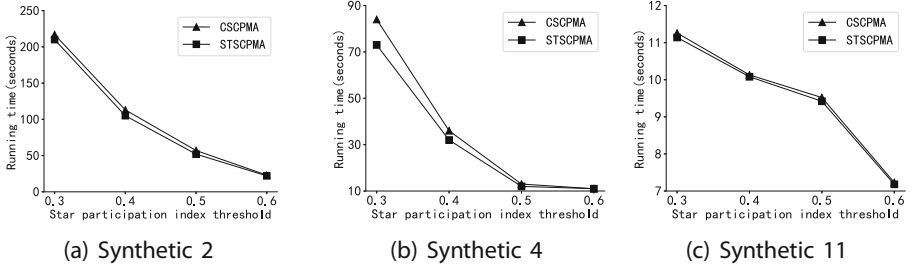


Fig. 7. Effect of star participation index threshold

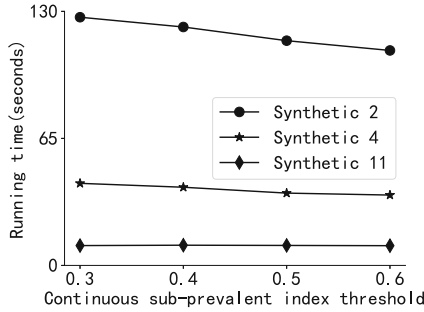


Fig. 8. Effect of continuous sub-prevalent index threshold

Effect of Continuous Sub-prevalent Index Threshold. We analyze the running time of the CSCPMA on datasets Synthetic 2,4,11 under different continuous sub-prevalent index thresholds as only the CSCPMA has this threshold. In Fig. 8, the running time of the CSCPMA on Synthetic 2 gradually decreases as the continuous sub-prevalent index threshold increases since the pruning strategy of CSCPs work effectively. The running time of the CSCPMA on Synthetic 4 and 11 is relatively flat, the reason is that generating candidate patterns costs a lot of time, but further generating CSCPs is relatively low-cost.

5.2 Case Analysis on Real Datasets

This section will analyze the cases on two real datasets to illustrate that the CSCPMA can mine more reasonable patterns by considering the duration and the frequency of patterns in time slice set.

On the Tokyo-Checkin Dataset. Table 3 list some size-2 and size-3 patterns mined by the CSCPMA and that mined by the STSCPMA on the Tokyo-Checkin dataset. We can see that the CSCPMA mines the size-2 pattern {Movie Theater,Parking} that cannot be mined by the STSCPMA, and the pattern is reasonable in real world. Usually, there are parkings near the movie theaters for convenience of audiences. The size-3 pattern {Bar,Parking,Hotel} indicates that

Table 3. Results comparison of CSCPMA and STSCPMA on Tokyo-Checkin

Size	CSCPMA	STSCPMA
2	{Hotel,Coffee Shop} {Convenience Store,Bar} {Movie Theater,Parking}	{Hotel,Coffee Shop} {Convenience Store,Bar}
3	{Restaurant,Mall,Coffee Shop} {Convenience Store,Restaurant,Hospital} {Bar,Parking,Hotel} {Mall,Coffee Shop,Barbershop}	{Restaurant,Mall,Coffee Shop} {Convenience Store,Restaurant,Hospital}

(Notes: the bold pattern is mined by the CSCPMA rather than STSCPMA)

consumers who have drunk alcohol should not drive, and should find hotels to rest. The size-3 pattern {Mall,Coffee Shop,Barbershop} shows that coffee shops and barbershops close to malls can have the benefit from malls. Therefore, the CSCPMA can mine more reasonable patterns than the STSCPMA on Tokyo-Checkin dataset.

On the ChicagoCrime Dataset. Let us take the pattern {Robbery, Sex Offense, Battery} as an example which is mined by the CSCPMA on ChicagoCrime dataset. The pattern indicates that robbery, sexual offense, and battery occur continuously and frequently in areas with poor public security.

6 Conclusion

This paper study the spatiotemporal sub-prevalent co-location pattern to better reveal the spatiotemporal relationship of patterns. First, we consider the duration and the frequency of patterns in time slice set, and propose the continuous sub-prevalent co-location pattern. Then, we propose an efficient mining algorithm to mine the proposed pattern. Extensive experiments on synthetic datasets and real datasets verify that the proposed algorithm can efficiently mine more reasonable patterns with richer spatiotemporal relationship.

Acknowledgements. This work is supported by the National Natural Science Foundation of China (62266050, 62276227), the Program for Young and Middle-aged Academic and Technical Reserve Leaders of Yunnan Province (202205AC160033), Yunnan Provincial Major Science and Technology Special Plan Projects (202202AD080003), the Open Project Program of Yunnan Key Laboratory of Intelligent Systems and Computing (ISC22Z02), Yunnan Fundamental Research Projects (202201AS070015).

References

1. Akbari, M., Samadzadegan, F., Weibel, R.: A generic regional spatio-temporal co-occurrence pattern mining model: a case study for air pollution. *J. Geogr. Syst.* **17**(3), 249–274 (2015)

2. Phillips, P., Lee, I.: Mining co-distribution patterns for large crime datasets. *Expert Syst. Appl.* **39**(14), 11556–11563 (2012)
3. Zeng, L., Wang, L., Zeng, Y., Li, X., Xiao, Q.: Discovering spatial co-location patterns with dominant influencing features in anomalous regions. In: Jensen, C.S., et al. (eds.) *DASFAA 2021. LNCS*, vol. 12680, pp. 267–282. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-73216-5_19
4. Huang, Y., Shekhar, S., Xiong, H.: Discovering co-location patterns from spatial data sets: a general approach. *IEEE Trans. Knowl. Data Eng.* **16**(12), 1472–1485 (2004)
5. Celik, M., Shekhar, S., Rogers, J.P., Shine, J.A.: Mixed-drove spatiotemporal co-occurrence pattern mining. *IEEE Trans. Knowl. Data Eng.* **20**(10), 1322–1335 (2008)
6. Li, X., Chen, H., Xiao, Q., Wang, L.: Spatiotemporal sub-prevalent co-location pattern mining. *J. Southwest Univ. Nat. Sci. Ed.* **42**(11), 68–76 (2020)
7. Li, X.: Mining spatiotemporal sub-prevalent co-location patterns based on star model. Master's thesis, Yunnan University (2021)
8. Yoo, J.S., Shekhar, S.: A joinless approach for mining spatial colocation patterns. *IEEE Trans. Knowl. Data Eng.* **18**(10), 1323–1337 (2006)
9. Wang, L., Bao, Y., Lu, J., Yip, J.: A new join-less approach for co-location pattern mining. In: 2008 8th IEEE International Conference on Computer and Information Technology, pp. 197–202. IEEE (2008)
10. Wang, L., Bao, Y., Lu, Z.: Efficient discovery of spatial co-location patterns using the iCPI-tree. *Open Inf. Syst. J.* **3**(1) (2009)
11. Wang, L., Zhou, L., Lu, J., Yip, J.: An order-clique-based approach for mining maximal co-locations. *Inf. Sci.* **179**(19), 3370–3382 (2009)
12. Yang, P., Wang, L., Wang, X., Zhou, L.: A spatial co-location pattern mining approach based on column calculation. *Sci. Sin. Inf.* **52**(6), 1053–1068 (2022)
13. Andrzejewski, W., Boinski, P.: Maximal mixed-drove co-occurrence patterns. *Inf. Syst. Front.* 1–24 (2022)
14. Qian, F., Yin, L., He, Q., He, J.: Mining spatio-temporal co-location patterns with weighted sliding window. In: 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems, vol. 3, pp. 181–185. IEEE (2009)
15. Ma, Y., Lu, J., Yang, D.: Mining evolving spatial co-location patterns from spatio-temporal databases. In: 2022 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 129–136. IEEE (2022)
16. Yang, L., Wang, L.: Mining traffic congestion propagation patterns based on spatio-temporal co-location patterns. *Evol. Intel.* **13**(2), 221–233 (2020)
17. Wang, L., Bao, X., Zhou, L., Chen, H.: Maximal sub-prevalent co-location patterns and efficient mining algorithms. In: Bouguettaya, A., et al. (eds.) *WISE 2017. LNCS*, vol. 10569, pp. 199–214. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68783-4_14
18. Wang, L., Bao, X., Zhou, L., Chen, H.: Mining maximal sub-prevalent co-location patterns. *World Wide Web* **22**(5), 1971–1997 (2019)
19. Ma, D., Chen, H., Wang, L., Xiao, Q.: Dominant feature mining of spatial sub-prevalent co-location patterns. *J. Comput. Appl.* **40**(2), 465–472 (2020)
20. Xiong, K., Chen, H., Wang, L., Xiao, Q.: Mining fuzzy sub-prevalent co-location pattern with dominant feature. In: *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, pp. 1–10 (2022)