JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 35, 1343-1363 (2019)
DOI: 10.6688/JISE.201911_35(6).0011

Spatial Co-location Pattern Mining Based on Fuzzy Neighbor Relationship

MEI-JIAO WANG[1,2], LI-ZHEN WANG[1,+] AND LI-HONG ZHAO[1]
[1]*Department of Computer Science and Engineering*
Yunnan University
Kunming, 650504 P.R. China
[2]*School of Information Science and Technology*
Chuxiong Normal University
Chuxiong, 675000 P.R. China
E-mail: meijiaow@cxtc.edu.cn; lz hwang@ynu.edu.cn; 1657611732@ynu.edu.cn

A co-location pattern is a subset of spatial objects whose instances are frequently located together in geography space. The traditional co-location mining algorithms treated the spatial proximity relationship between the instances as unanimous by binary logic, which weakened the accuracy and effectiveness of the results. In this paper, the co-location pattern mining based on fuzzy neighbor relationship is studied. Firstly, fuzzy neighbor relationship (FNR) is defined to measure the proximity level between instances, and then the fuzzy participation ratio and the fuzzy participation index are defined. Secondly, the algorithm for spatial co-location pattern mining based on FNR (CPFNR) is proposed by the basic idea of the Join-less algorithm. Moreover, optimizing strategy is adopted for the CPFNR algorithm. Finally, the effectiveness of the CPFNR algorithm is verified by experiments on the real datasets, and the performance of our algorithm is evaluated on the synthetic datasets.

Keywords: data mining, fuzzy set, spatial co-location pattern, fuzzy neighbor relationship, fuzzy participation index

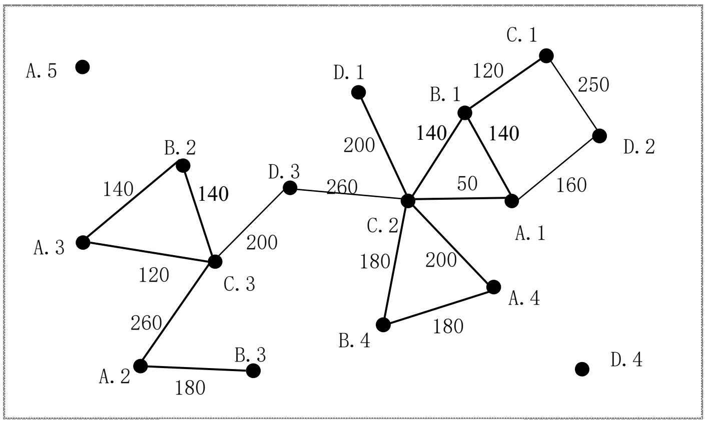
1. INTRODUCTION

Spatial co-location pattern mining, which is an important branch of spatial data mining, has attracted more and more researchers. A co-location pattern is a subset of spatial objects whose instances are frequently located together in geography space. Spatial co-location pattern mining discovers the association relation of spatial objects from the spatial datasets, helping for human decision-making. For example, epidemiologists have found that the Nile crocodile and Egyptian plovers are frequently co-located. Botanists have found that eighty percent of the semi-humid evergreen broad-leaved forests grow in places where there is orchid according to the distribution of the symbiotic vegetation. Spatial co-location pattern mining has been applied to many applications, including species distribution [1], location services [2], public security [3], environmental management [4] and so on.

The traditional co-location mining algorithm employed the participation index (PI) to measure the prevalence of a co-location pattern [5]. Given a co-location pattern $c = \{o_1, o_2, \ldots, o_k\}$ where $o_1, o_2, \ldots, o_k$ are object types, the participation index of c is defined as $PI(c) = \min_{o_i \in c} PR(c, o_i)$, where $PR(c, o_i)$ is the participation ratio of the object o_i in c,

Received July 31, 2018; revised November 8, 2018; accepted December 31, 2018.
Communicated by Tzung-Pei Hong.
+ Corresponding author.

calculated by $PR(c, o_i) = \dfrac{\text{number of distinct instances of } o_i \text{ in row instances of } c}{\text{number of instances of } o_i}$, where a row instance of c is an instance cliques of c, which is composed of neighbor instances whose object types match the types in c. But this PI calculation defined the neighbor relationship (R-neighbor) between instances by setting a distance threshold. Namely, when the distance between two instances is no less than the distance threshold, the two instances satisfy R-neighbor and they are neighbors, otherwise they don't. This binary judgment didn't take into accounts the proximity level, thus weakening the accuracy and validity of the mining results.



(a) An example spatial dataset.

PI({B,C})=min{3/4,1}
=3/4

table instance of {B,C}

(PI=3/4) B	C	distance
B.1	C.1	120
B.1	C.2	140
B.2	C.3	140
B.4	C.2	180
PR: 3/4	1	

(PI=3/4) C	D	distance
C.1	D.2	250
C.2	D.1	200
C.2	D.3	260
C.3	D.3	200
PR: 1	3/4	

distance between C.1 and D.2

(b) Table instances and PI&PR values of the co-locations {B, C} and {C, D}.
Fig. 1. A motivating example.

Fig. 1 (a) shows an example spatial dataset. There are four different objects O = {A, B, C, D}. Each instance is marked by its object type and a numeric ID value. For example, the instance A.1 is the first instance of object A. Let the distance threshold is 300. The solid line between two instances indicates the R-neighbor, and the value on the line is the Euclidean distance between the two instances. Object A has five instances, B has four instances, C has three instances, and D has four instances. Fig. 1 (b) illustrates the table instances (the set of row instance) and the PI values of co-location patterns {B, C} and {C, D}. We can get that PI({B, C}) = PI({C, D}). Namely, the co-location {B, C} has the same prevalence as the co-location {C, D}. But, obviously, the distances between the instances of B and the instances of C are much smaller than those between the instances of C and the instances of D. However, the PI calculation treats them equally without considering the proximity level between instances.

To address the above problem, this paper defines fuzzy neighbor relationship (FNR) to measure the proximity level by the fuzzy set theory. But it is a big challenge that mining co-location pattern based on FNR (CPFNR) instead of R-neighbor because the participation index (PI) which is defined based on R-neighbor as the metric of the prevalence of a co-location pattern will not fit for the co-location pattern based on FNR. And the traditional mining methods are not suitable for CPFNR either.

Our contributions are as follows: (1) We define FNR to measure the proximity level between instances; (2) The fuzzy participation ratio and fuzzy participation index are defined based on FNR, which are very different from the participation ratio and index in the conventional algorithms; (3) The CPFNR algorithm as well as the optimization strategy is proposed; (4) Extensive experiments are conducted to verify that the proposed algorithm is effective and the proposed algorithm as well as the optimization techniques achieves satisfactory performance.

The remainder of this paper is organized as follows: Section 2 describes related work; Section 3 gives the concepts and properties of co-location patterns based FNR; Section 4 proposes the algorithm for co-location pattern mining based on FNR; Section 5 verifies the effectiveness and the performance of the proposed algorithms; Section 6 presents the conclusion and future work.

2. RELATED WORK

Co-location was originally proposed by Shekhar and Huang for discovering the distribution of ecological species [5], and the Join-based co-location mining algorithm which is an Apriori-like strategy was put forward. But when the datasets is dense or the length of the co-location pattern increases, the connections between the table instances become huge. The partial join approach [6] and the Join-less algorithm [7] were presented for reducing the connections between table instances. Wang *et al.* proposed 3 Join-less co-location pattern mining approaches [8-10]. The CPI-tree (Co-location Pattern Instances tree) algorithm [8] materialized the neighbor relation of the instances by a tree structure. The iCPI-tree algorithm [9] integrated the Apriori pruning and the tree structure of CPI-tree. The work in [10-13] studied the problem of maximal co-location pattern mining. The incremental mining and competitive pairs mining of co-location patterns were studied in [14, 15]. Prevalent co-location redundancy reduction problem was discussed in [16]. The paper [17] presented a new lossless condensed representation of prevalent co-location collections. High utility co-location pattern mining methods were presented in [18-20]. The research on co-location pattern mining in big data was conducted in [21, 22].

Researchers have not only studied the co-location pattern mining methods, but also extended the research objects from the classical data to the special data. The co-location patterns were mined from spatial uncertain data [23], from the uncertain data with probability intervals [24] and from interval data [25]. The paper [26] proposed the co-location pattern mining algorithm for rare objects.

In 1965, the concept of fuzzy set was proposed by Professor Zadeh, which is used for studying the problem with vagueness or uncertainty, then it developed rapidly. In previous work, the idea of fuzzy set theory has been introduced into the study of spatial co-location pattern mining. Our paper [27] proposed a spatial co-location pattern mining method for fuzzy objects first. Co-location pattern mining was studied for spatial datasets with fuzzy attributes in [28]. However, so far, no research work has been done on applying the fuzzy set to the neighbor relationship in co-location pattern mining, which will be addressed in this paper. The most similar work to ours is [29], in which a kernel-density-estimation (KDE) model was adopt to measure the proximity level between instances. And a KDE-based prevalence index (PI-K) was defined in the SGCT-K algorithm for mining the maximal co-location patterns. However, the SGCT-K algorithm has the following limitations: (1) It needs a distance threshold, which is difficult to set, to identify neighboring instances; (2) The PI-K is rather small, leading too hard to give a prevalence threshold to select the prevalent co-locations; (3) Because only the KDE-based prevalence index of a co-location pattern was deduced, the SGCT-K algorithm can't be applied to the applications that concern on the prevalence ratio of an object, such as co-location pattern mining with rare objects. Based on the membership function, the FNR defined in this paper doesn't require a distance threshold. And compared to the KDE-based prevalence index, the fuzzy participation index (ratio) is much more closer to participation index (ratio) in classic mining methods.

3. RELATED CONCEPTS AND PROPERTIES

This section first defines some concepts, and then the related properties are followed.

3.1 Basic Definitions

Spatial objects represent different kinds of events in space. The spatial object set is denoted as $O = \{o_1, o_2, ..., o_n\}$. The object in each specific location is called spatial instance, expressed in i. The collection of the spatial instances are called instance set, denoted as $S = S_1 \cup S_2 \cup ... \cup S_n$, where $S_j(1 \leq j \leq n)$ is the instances set of the spatial object o_j. In this paper, two representations of a spatial instance are used: (1) Any spatial instance in S is denoted as $i_s(1\leq s \leq |S|)$; (2) Given an object $o_j\,(1 \leq j \leq n)$, an instance of o_j is denoted as $i_j^s\,(1\leq s \leq |S_j|)$.

Definition 1 (fuzzy neighbor relationship (FNR)): Taking the Euclidean distances D between the spatial instances as the domain, where $D\rightarrow[0, \infty)$, the fuzzy neighbor relationship FNR is defined as a set of proximity relationship based on the distances D. The distance between two specific instances is denoted as $dist(.)$, then we can give the following mapping: μ: $D\rightarrow[0, \infty)$, $dist(.)\rightarrow\mu(dist(.))$. We say that μ determines a fuzzy subset FNR on D, and μ is the membership function of FNR, $\mu(dist(.))$ is the membership value of $dist(.)$ which indicates the probability of $dist(.)$ belonging to FNR.

The FNR maps the distance between two instances to a value in the interval [0,1]. Different from the R-neighbor in traditional co-location pattern mining framework, the FNR can measure the proximity level between two spatial instances. Namely, the larger the membership value, the higher the proximity level. In practical applications, the membership function is often determined by expert experience. In the example datasets in Fig. 2, we give the membership function μ as follows:

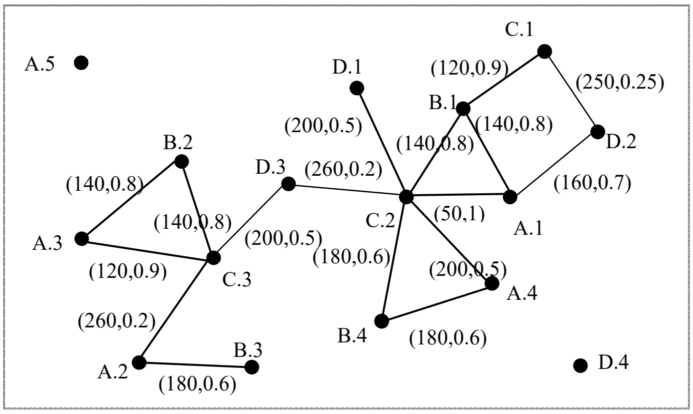


Fig. 2. The example spatial datasets based on FNR with $\alpha = 0.1$.

$$\mu(dist(.)) = \begin{cases} 1 & dist(.) \le 100 \\ -\frac{1}{200}(dist(.) - 100) + 1 & 100 < dist(.) \le 300 \\ 0 & dist(.) > 300 \end{cases} \tag{1}$$

In Eq. (1), the value 300 is called the boundary distance. According to Eq. (1), in Fig. 2, we can get the FNR: $\mu(dist(\text{B.1, C.1})) = 0.9$, $\mu(dist(\text{C.1, D.2})) = 0.25$, $\mu(dist(\text{A.1, B.1})) = 0.8$, $\mu(dist(\text{A.1, C.2}) = 1$, $\mu(dist(\text{A.2, C.3})) = 0.2$, *etc.*

Definition 2 (α-cut set): Given a user-defined membership threshold $\alpha \in [0,1]$, the **α-cut set** of FNR is defined as: $\text{FNR}_\alpha = \{\mu(dist(.)) \mid (dist(.) \in D, \mu(dist(.)) \ge \alpha\}$, in which the membership values between instances satisfy α.

Definition 3 (fuzzy neighbors): Given two spatial instances $i_s \in S$ and $i_t \in S$, if $\mu(dist(i_s, i_t)) \in \text{FNR}_\alpha$, then we call i_s and i_t fuzzy neighbors, denoted as $FNeib(i_s, i_t)$.

If two spatial instances are fuzzy neighbors, then the solid line is connected between them in the graph. As shown in Fig. 2, they are *FNeib*(B.1, C.1), *FNeib*(A.1, C.2), *etc.*

A **co-location pattern** c is a set of spatial objects, *i.e.*, $c \subseteq O$. The number of objects in c is called the size of c. For example, in Fig. 2, the co-location {A, B, C} is a size-3 co-location pattern. A **row instance** is an instance set in which the instances form a clique under the fuzzy neighbors. A row instance of c is denoted as ***RI*(*c*)**. The collection of all row instances of c is called the **table instance** of c, denoted as ***TI*(*c*)**. For example, in Fig. 2, the instance set {A.1, B.1, C.2} is a row instance of the co-location pattern {A, B, C}, and the row instances {A.1, B.1, C.2}, {A.3, B.2, C.3} and {A.4, B.4, C.2} constitute the table instance of {A, B, C}.

3.2 Fuzzy Participation Ratio and Fuzzy Participation Index

In this section, we define the fuzzy participation ratio and the fuzzy participation index based on FNR. Firstly, we give the concept of the contribution of an instance.

Definition 4 (the contribution of an instance): The contribution of an instance refers to the contribution to the fuzzy participation index of its object, and is defined as the smallest value in the membership values between the instance and all its fuzzy neighbors in the row instance. The contribution of the instance $i_s(1 \leq s \leq t)$ in the row instance $RI(c) = \{i_1, i_2, \ldots, i_t\}$ is expressed as:

$$Contri_{RI(c)}(i_s) = \min_{j=1}^{t}(\mu(dist(i_s, i_j))) \quad j \neq s. \tag{2}$$

For example, in Fig. 2, the contribution of B.1 in row instance {B.1, C.1} is $Contri_{\{B.1,C.1\}}(B.1) = 0.9$. And the contribution of A.1 in row instance {A.1, B.1, C.2} is $Contri_{\{A.1,B.1,C.2\}}(A.1) = min(\mu(dist(A.1, B.1)), \mu(dist(A.1, C.2))) = 0.8$.

Next, we will introduce the definitions of fuzzy participation ratio and fuzzy participation index, the names of which are the same as definitions in [27] but their meanings are quite different. Fuzzy participation ratio and fuzzy participation index are defined based on FNR in this paper while they are defined for fuzzy objects in [27].

Definition 5 (fuzzy participation ratio and fuzzy participation index): Given a co-location pattern $c = \{o_1, o_2, \ldots, o_k\}$, the fuzzy participation ratio $FPR(c, o_u)$ of the object $o_u \in c(1 \leq u \leq k)$ in c is a fraction, of which the molecule is the sum of the contribution of the instances of o_u in the table instance of c, and the denominator is the number of the instances of o_u. *i.e.*,

$$FPR(c, o_u) = \frac{\sum_{i_u^s \in RI(c), RI(c) \in TI(c)} Contri_{RI(c)}(i_u^s)}{|TI(\{o_u\})|}. \tag{3}$$

In Eq. (3), when an instance of o_u appears in more than one row instances, only the largest contribution value is summed in the summation expression.

The **fuzzy participation index** of the co-location pattern c is the smallest fuzzy participation ratio of the object in c. *i.e.*,

$$FPI(c) = \min_{u=1}^{k}\{FPR(c, o_u)\}. \tag{4}$$

Given a user-defined **fuzzy prevalence threshold** *min_fprev*, a co-location pattern c is prevalent if $FPI(c) \geq min\_fprev$.

For example, in Fig. 2, let $min\_fprev = 0.4$. The table instance of the co-location pattern $c = \{A, B, C\}$ is {{A.1, B.1, C.2}, {A.3, B.2, C.3}, {A.4, B.4, C.2}}. The contribution of A.1 in {A.1, B.1, C.2} is $Contri_{\{A.1,B.1,C.2\}}(A.1) = 0.8$, the contribution of A.3 in {A.3, B.2, C.3} is $Contri_{\{A.3,B.2,C.3\}}(A.3) = 0.8$, and the contribution of A.4 in {A.4, B.4, C.2} is $Contri_{\{A.4,B.4,C.2\}}(A.4) = 0.5$. So the fuzzy participation ratio of the object A in c is $FPR(c, A) = \frac{0.8+0.8+0.5}{5} = 0.425$. Similarly, we can get $FPR(c, B) = 0.55$ and $FPR(c, C) = 0.53$ respectively. Therefore the fuzzy participation index of c is $FPI(c) = min(FPR(c, A),$

$FPR(c, \text{B})$, $FPR(c, \text{C})) = 0.425$, then we can get that c is prevalent. We can also obtain that $FPI(\{\text{B, C}\}) = 0.575$ and $FPI(\{\text{C, D}\})=0.3125$. Compared to the result $PI(\{\text{B, C}\}) = PI(\{\text{C, D}\}) = 3/4$ in the motivating example in Fig.1, the FPI could give a different result, $FPI(\{\text{B, C}\}) > FPI(\{\text{C, D}\})$. It can be seen that, because the calculation of FPI takes into accounts the proximity levels between instances, the prevalence of the co-location patterns can be better differentiated.

In the definition of the fuzzy participation ratio, when all of the membership values between instances are 1, the contribution of each instances is "1", in this case, the fuzzy participation ratio is the participation ratio in the classical algorithm. When the membership value between instances is in the interval (0,1], the higher the membership value, the much the contribution, the larger the fuzzy participation ratio. Therefore, the definition of fuzzy participation ratio is reasonable.

3.3 Properties

The fuzzy participation ratio and the fuzzy participation index possess the anti-monotonicity property.

Lemma 1: Fuzzy participation ratio and fuzzy participation index are monotonically non-increasing with increases in the size of the co-location.

***Proof*:** Given a size-k co-location pattern $c = \{o_1, o_2, \ldots, o_k\}$ and a size-k+1 co-location $c' = c \cup o_{k+1}$, for any object $o_u (1 \leq u \leq k)$, if an instance of o_u contributes to the fuzzy participation ratio of o_u in c', it certainly does so in c, but not vice versa. So the fuzzy participation ratio is monotonically non-increasing with increases in the size of the co-location. As the fuzzy participation index is the minimum participation ratio of all the objects in the co-location, it is also monotonically non-increasing with increases in the size of the co-location.

Theorem 1: If a co-location pattern c is prevalent, all of its subsets $c' \subseteq c$ are prevalent; conversely, if c is not prevalent, all of its supersets $c' \supseteq c$ are also not prevalent.

***Proof*:** According to Lemma 1, the fuzzy participation index of the co-location pattern c is no larger than the fuzzy participation index of any subset of c, and it is no less than the fuzzy participation index of the superset of c. Therefore, if c is prevalent, all of its subsets are prevalent; if c is not prevalent, all of its supersets are also not prevalent.

4. CO-LOCATION PATTERNS MINING BASED ON FNR

In this section, an algorithm for the co-location pattern mining based on FNR (CPFNR) is designed by improving the classic Join-less algorithm [7]. And we give a strategy to optimize the algorithm.

4.1 The Algorithm for Co-location Pattern Mining Based on FNR (CPFNR)

The CPFNR algorithm first constructs the fuzzy star neighbor set, and then iteratively executes the following steps: Generate the candidate co-location patterns; Collect star instances of the candidate co-location patterns; Prune the candidate co-locations; Check the clique relationships of the star instances; Select the co-location patterns whose fuzzy participation index is no less than the fuzzy prevalence threshold.

(A) The Star Model

Similar to the star model in the Join-less algorithm, in the CPFNR algorithm, the star model is used to store the fuzzy star neighbour set.

Definition 6 (fuzzy star neighbor): Given an $o_u \in O(1 \leq u \leq n)$, for any instance i_u^s $(1 \leq s \leq |S_u|)$, the fuzzy star neighbor of i_u^s is defined as:

$$SNeib(i_u^s) = \{i_v^t \mid i_v^t = i_u^s \vee (v > u \wedge FNeib(i_u^s, i_v^t))\} \tag{5}$$

where o_u is referred as the central object and i_u^s is the central instance.

According to Definition 6, the fuzzy star neighbor of an instance is a collection of the instance and its fuzzy neighbors, requiring that the type of its fuzzy neighbor is larger than the type of the instance itself in the dictionary order. Unlike the star model in the Join-less algorithm, the star model in the CPFNR algorithm stores the star neighbor of the instance as well as the membership values between the instance and its star neighbor. For example, in Fig. 2, the star neighbor of the instance A.1 is {A.1, (B.1 0.8), (C.2 1), (D.2 0.7)}, where the membership value of A.1 and B.1 is 0.8, of A.1 and C.2 is 1 and of A.1 and D.2 is 0.7.

Given a collection of instances $I = \{i_1, i_2, ..., i_m\}$, if all instances excluding i_1 are fuzzy neighbors of the instance i_1 in I, I is called a **star instance** of the co-location pattern which is composed of the object types of the instances in I. For example, in Fig.2, the star neighbor {A.2, (B.3 0.6), (C.3 0.2)} of A.2 is a star instance of the co-location {A, B, C}. The star instance of the co-location pattern c is denoted as ***SI*(*c*)**, and the collection of ***SI*(*c*)** is called the **star table instance**, recorded as ***STI* (*c*)**.

(B) Coarse Pruning

In order to reduce the complexity of checking the clique relationships of the star instance, we prune the candidate co-locations by **the upper bound of fuzzy participation index** on the star instance level. In the star instance, it is easy to get the contribution of the central instance; for the non-central instances, we can only get the membership value between it and the central instance, so we can't compute the contribution of it. But we treat the membership value between it and the central instance as the contribution of it instead. Then we can compute a fuzzy participation ratio of its object, which is called **the upper bound of fuzzy participation ratio** because the star instance may be not true row instance. So we can give the following definitions.

Definition 7 (the upper bound of fuzzy participation ratio and the upper bound of fuzzy participation index): Given a size-k candidate co-location pattern $c = \{o_1, o_2, \ldots, o_k\}$, on the star instance level, for any object $o_u \in c(1 \leq u \leq k)$, the upper bound fuzzy participation ratio of o_u is denoted as $U\_FPR(c, o_u)$, which is defined as follows:

$$U\_FPR(c,o_u)=\begin{cases}\dfrac{\sum_{i_u^s\in SI(c),SI(c)\in STI(c)} Contri_{SI(c)}(i_u^s)}{|TI(\{o_u\})|} & u=1\\ \dfrac{\sum_{i_1^s,i_u^t\in SI(c),SI(c)\in STI(c)} \mu(dist(i_1^s,i_u^t))}{|TI(\{o_u\})|} & u\neq 1\end{cases}. \tag{6}$$

In Eq. (6), when an instance of o_u appears in more than one star instances, only the largest contribution value is summed in the summation expression.

The upper bound of the fuzzy participation index of the co-location pattern c, denoted as $U\_FPI(c)$, and is defined as the minimum upper bound of the fuzzy participation ratio of all objects in c. Namely,

$$U\_FPR(c,o_u)=\min_{u=1}^{k}\{U\_FPR(c,o_u)\}. \tag{7}$$

Theorem 2: For any co-location pattern c, if $U\_FPI(c)\leq min\_fprev$, then c could be pruned.

***Proof*:** Given a size-k co-location $c=\{o_1, o_2, \ldots, o_k\}$, for any object $o_u\in c(1\leq u\leq k)$, $FPR(c, o_u)\leq U\_FPR(c, o_u)$, and $FPI(c)=\min_{u=1}^{k}\{FPR(c,o_u)\}\leq\min_{u=1}^{k}\{U\_FPR(c,o_u)\}=U\_FPI(c)$. So if $U\_FPI(c) < min\_fprev$, we can get $FPI(c) < min\_fprev$, then c could be pruned.

For example, in Fig. 2, let $min\_fprev=0.5$. The star instance set of the co-location pattern $c=\{A, B, C\}$ is {{A.1, B.1, C.2}, {A.2, B.3, C.3}, {A.3, B.2, C.3}, {A.4, B.4, C.2}}. The upper bound of the fuzzy participation ratio of the object A, B and C in c are $U\_FPR(c,A)=\frac{0.8+0.5+0.8+0.2}{5}=0.46$, $U\_FPR(c,B)=\frac{0.8+0.6+0.8+0.6}{4}=0.7$, $U\_FPR(c,C)=\frac{1+0.9}{3}=0.63$ respectively, then the upper bound of the fuzzy participation index of c is $U\_FPI(c)=0.46 < min\_fprev=0.5$, and c could be pruned.

(C) Description for the CPFNR Algorithm

The description for the CPFNR algorithm is as following:

Algorithm 1: the CPFNR algorithm

Input: $O=\{o_1, o_2, ..., o_n\}$: spatial object set, S: spatial instance set, μ: membership function, min_fprev: fuzzy prevalence threshold, α: membership threshold.

Output: co-location patterns with $FPI\geq min\_fprev$

Variables: k: size of a co-location pattern, FSN: fuzzy star neighbor set, C_k: a set of size-k candidate co-locations, SI_k: star instance set of size-k candidate co-locations, P_k: a set of size-k prevalent co-location patterns, TI_k: table instance of size-k candidate co-locations.

Steps:

(1) FNR = get_membership_value(S, μ);

(2) FSN = gen_star_neighbor (O, S, FNR_α);

(3) $P_1=O$;

(4) while(not empty P_{k-1}) do
    (4.1) C_k = gen_candidate_co-locations(P_{k-1});
    (4.2) SI_k = get_star_instances(C_k, FSN);
    (4.3) if k = 2 then TI_k = SI_k;
        else
        {
            C_k = prune_candidate_co-locations(C_k, SI_k, *min_fprev*);
            TI_k = check_clique_instance(C_k, SI_k);
        }
    (4.4) P_k= select_prevalent_co-locations(C_k, TI_k, *min_fprev*);
    (4.5) $k = k+1$;
    end do

(D) The Computational Complexity Analysis

The detail analysis of the computational complexity of the classic Join-less algorithm is described in [7], in which the computational complexity of the Join-less algorithm $T_{jl} = T_{star\_neighborhood}(S) + T_{jl}(2) + \sum_{k>2} T_{jl}(k)$, where $T_{star\_neighborhood}(S)$ represents the cost in constructing the star neighbor set, and $T_{jl}(k)(k \geq 2)$ represents the cost for finding size-k co-location patterns. Compared to the Join-less algorithm, the CPFNR algorithm cost more time to compute the membership values between instances and more memory to storage them in constructing the fuzzy star neighbor set. For finding size-$k(k \geq 2)$ co-location patterns, the CPFNR algorithm spend more time on scanning the fuzzy star neighbor set to compute the upper bound of fuzzy participation index of the candidate co-locations for coarse pruning and the fuzzy participation index of the candidate co-locations for selecting the prevalent co-locations.

4.2 Optimizing Strategy

Although the CPFNR algorithm has shown good performance which is shown by the experiments in the later section, we can still provide a strategy to further improve the efficiency. In the steps (4.3) and (4.4) of the CPFNR algorithm, we need to scan the star neighbor set to get the contribution of the instance for computing the fuzzy participation index of its object, which will be time-consuming. We will give a tactics to improve the efficiency of these two steps.

Theorem 3: Given a co-location pattern $c = \{o_1, o_2, \ldots, o_k\}$, for any object $o_u(1 \leq u \leq k)$ in c, if $FPR(c, o_u) < min\_fprev$, then c is not prevalent and could be pruned.

***Proof*:** Because $FPI(c) = \min_{u=1}^{k}\{FPR(c, o_u)\} \leq FPR(c, o_u)$, if $FPR(c, o_u) < min\_fprev$, then $FPI(c) < min\_fprev$, we can get c is not prevalent, and it could be pruned.

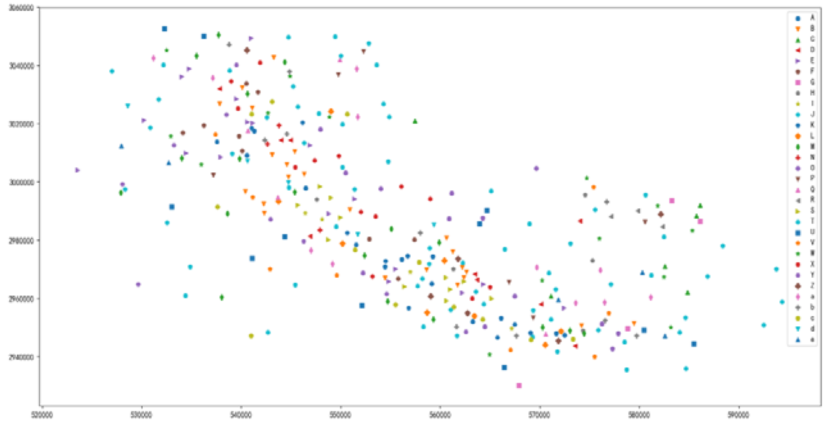
According to Theorem 3, once we acquire that the fuzzy participation ratio of some object of the co-location pattern is less than *min_fprev*, we can prune the co-location immediately without having to calculate the fuzzy participation ratio of all the other objects for the fuzzy participation index of the co-location. It will reduce the time spent on scanning the star neighbor set and improve the efficiency of the algorithm.

5. EXPERIMENTS

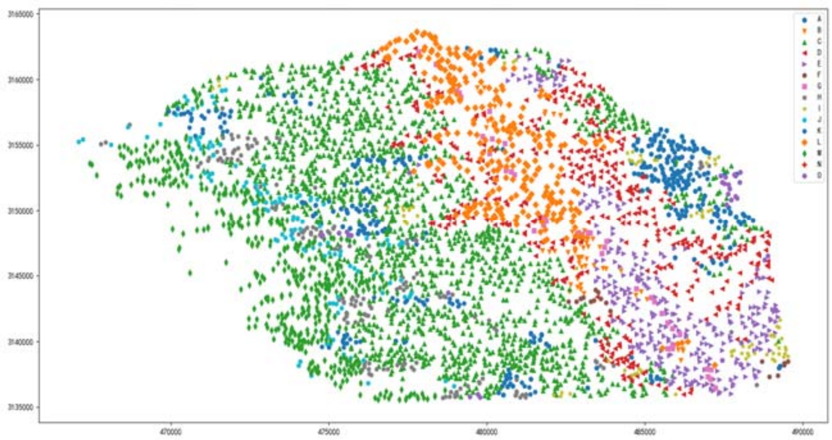
In this section, we perform experiments to verify the effectiveness and efficiency of the proposed algorithm on both real and synthetic datasets. All the algorithms are implemented in Java and run on a normal PC with core i7 3.40 GHz CPU and 16G memory.

5.1 Datasets

Fig. 3 shows the two real datasets used in our experiments. The summary of the two real datasets is demonstrated in Table 1. The datasets Real-1 is the vegetation distribution datasets of the Three Parallel Rivers of Yunnan Protected Area, which has 31 species of plants with 336 instances as shown in Fig. 3 (a). The datasets Real-2 is also the vegetation distribution datasets of the Three Parallel Rivers of Yunnan Protected Area. It has less objects but more instances than Real-1. The distribution of the datasets Real-2 is shown in Fig. 3 (b). Both of the two real datasets are normalized in the range of 2000*2000 in the experiments.



(a) Real-1



(b) Real-2

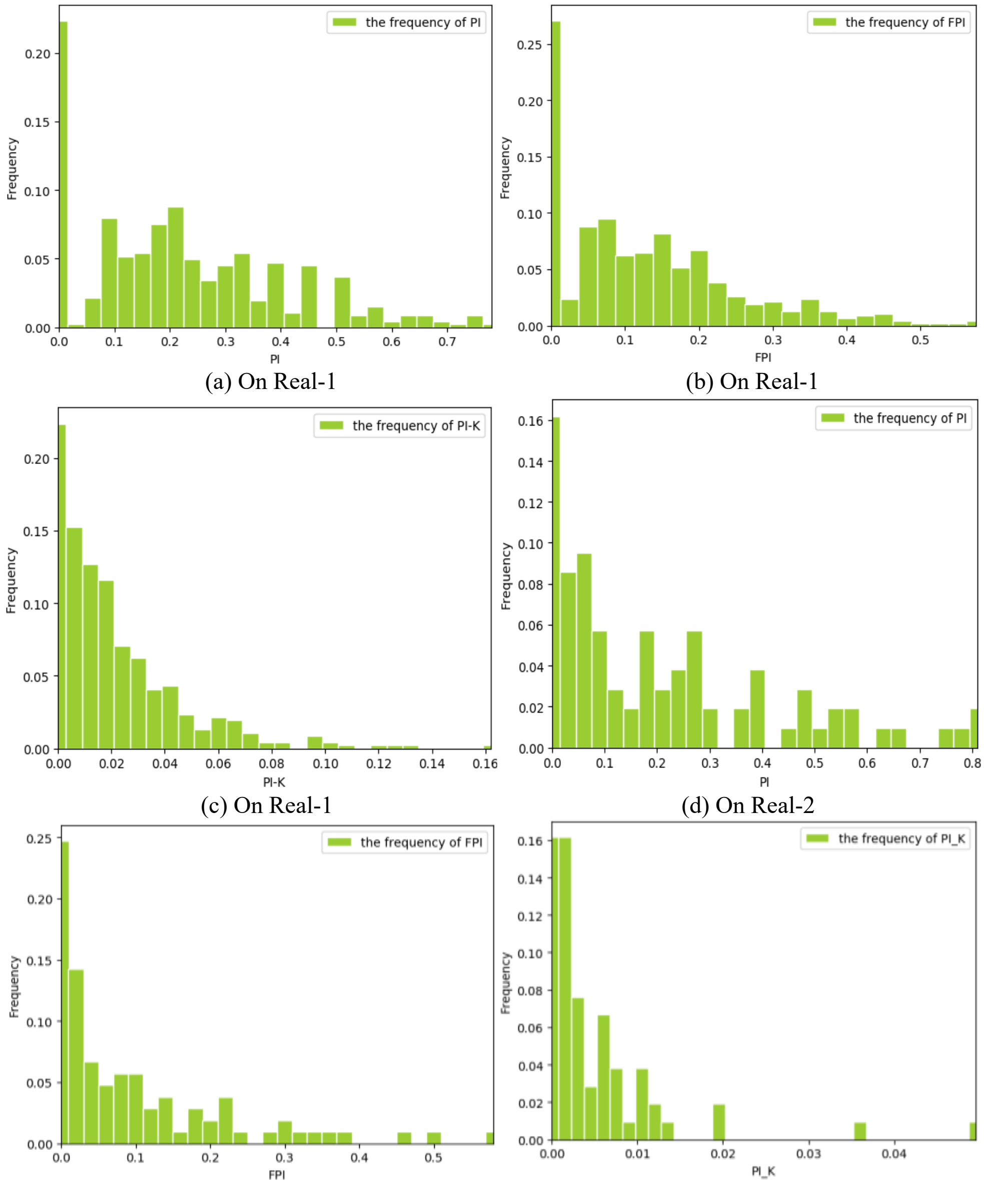
Fig. 3. The real datasets.

Table 1. Real datasets summary.

Dataset	Number of objects	Number of instances	(Max, Min)
Real-1	31	336	(62,3)
Real-2	15	3913	(1536,6)

(Max, Min): the maximum and minimum number of the object's instances in the real datasets.

The synthetic datasets used in the experiments are produced by a synthetic data generator similar to that used in the paper [5], distributed in a range of 2000*2000.



(a) On Real-1 (b) On Real-1
(c) On Real-1 (d) On Real-2
(e) On Real-2 (f) On Real-2

Fig. 4. The frequency histograms of the three prevalence metrics of the size-2 co-locations on the two real datasets.

5.2 The Effectiveness of the CPFNR Algorithm

(A) Comparisons of the prevalence metrics at the macro-level

We conduct experiments on the real datasets to compare the prevalence metrics of the CPFNR algorithm with that of the other two algorithms, namely, the classic Join-less algorithm and the SGCT-K algorithm. The prevalence metrics of the CPFNR algorithm, the classic Join-less algorithm [7] and the SGCT-K algorithm [29] are the fuzzy participation index (FPI), the participation index (PI), and the participation index based on kernel density evaluation (PI-K), respectively. We set the boundary distance in the membership function of FNR to be equal to the distance threshold of the Join-less algorithm and the SGCT algorithm for verifying the effectiveness of the CPFNR algorithm.

As shown in Table 2 and Fig. 4, the three kinds of prevalence metrics have different ranges when the boundary distance of FNR and the distance threshold of Join-less or SGCT-K are the same. Table 2 lists the minimum, the maximum and the average values of the PI, the FPI and the PI-K of the size-2 co-locations on the two real datasets. It can be seen that both the maximum and the average values of the FPI are smaller than those of the PI. The reason is that the participation ratio (PR) of an object in the Join-less algorithm is computed by counting the number of times its instances occur in the row instances without taking into account the proximity level between instances. However, the fuzzy participation ratio of an object is calculated based on the contribution of its instances, which is obtained from the membership values between it and its fuzzy neighbors. Accordingly, for an instance that appears in a row instance, in the former case, the contribution to the participation ratio is assigned the value of 1, while in the latter case, the contribution to the fuzzy participation ratio is assigned to the value in the interval [0,1]. Therefore, for the same co-location pattern, the FPI is always no larger than the PI. Table 2 and Fig.4 also show that the FPI is close to the PI while the PI-K is so small that there is great difference between the PI-K and the PI.

Table 2. The extreme values of the prevalence metrics of the size-2 co-locations on the real datasets.

Datasets	PI			FPI			PI-K		
	Min	Max	AVG	Min	Max	AVG	Min	Max	AVG
Real-1	0	0.7692	0.2174	0	0.5871	0.1365	0	0.1649	0.0221
Real-2	0	0.8176	0.1841	0	0.6941	0.1168	0	0.0501	0.0034

Nevertheless, from Fig. 4, It can be observed that, the three prevalence metrics accord the exponential-like distribution on the whole. The distribution of the FPI is much more similar to that of the PI than the PI-K. In table 2, the same characteristics are shown in the maximum and minimum values of the three metrics. The maximum values of the three in Real-1 are both smaller than those in Real-2, while the average values of the three in Real-1 are both larger than those in Real-2. From the macro point of view, we can draw a conclusion that the FPI has the similar effects in mining co-location patterns as the PI and the PI-K do.

(B) Comparisons of the prevalence metric values at the micro-level

We will analyze the differences among the three prevalence metrics in detail by the mining results on the datasets Real-1. Table 3 lists the plants in the top 10 size-2 co-location patterns mined by the CPFNR algorithm, which are listed in Table 4, and the symbols for each plant used in the experiments. Table 4 lists the FPI, the PI and the PI-K as well as the ranks of the 10 prevalent co-location patterns in the results of the three mining methods. As shown in Table 4, the ranks of the FPI and the PI-K of the 10 co-location patterns are different from that in the Join-less algorithm. Compared to the ranks of PI-K, the rank of FPI is more similar to that of PI. The co-location patterns {A, L} and {A, I}(marked by gray background) share the same prevalence level in the results of the Join-less algorithm but have different prevalence levels in the results of the CPFNR algorithm or the SGCT-K algorithm. The same thing happens to the co-locations {B, X} and {B, E} (marked by gray background). Obviously, this is because the influence of distances between instances on participation ratio is not considered in the Join-less algorithm, while the CPFNR algorithm uses the fuzzy set theory to measure the proximity level or the SGCT-K algorithm constructed the KDE model among the instances. Next we will describe how these happen in the CPFNR algorithm and the Join-less algorithm in detail (the explanations for the SGCT-K algorithm was discussed in [29]).

Table 5 gives the (fuzzy) participation ratio((F)PR) of the objects contained in the four co-location patterns mentioned above(marked by gray background in Table 4) for the CPFNR algorithm and the Join-less algorithm. We select the co-locations {B, X} and {B, E} as the representative to describe the problem. As we know, the FPR of an object is determined by the contribution of its instances in the row instances and the total number of instances. In Table 5, for the same object B, it has the same prevalence level in the

Table 3. The plants in top 10 size-2 co-locations in result of the CPFNR algorithm.

Tabs	Name of plants		Tabs	Name of plants
A	Abies georgei		L	Fritillaria delavayi franch
B	Taxus yunnanensis		S	Pseudotsuga forrestii craib
E	Yunnan Torreyn		X	Cephalotaxus lancceolata
I	Wake robin		Z	Megacarpaea delavayi Franch
J	Saussurea gossypiphora		c	Cordyceps sinensis
K	Magnolia sieboldii			

Table 4. The ranks of the three metrics of the top 10 size-2 co-locations.

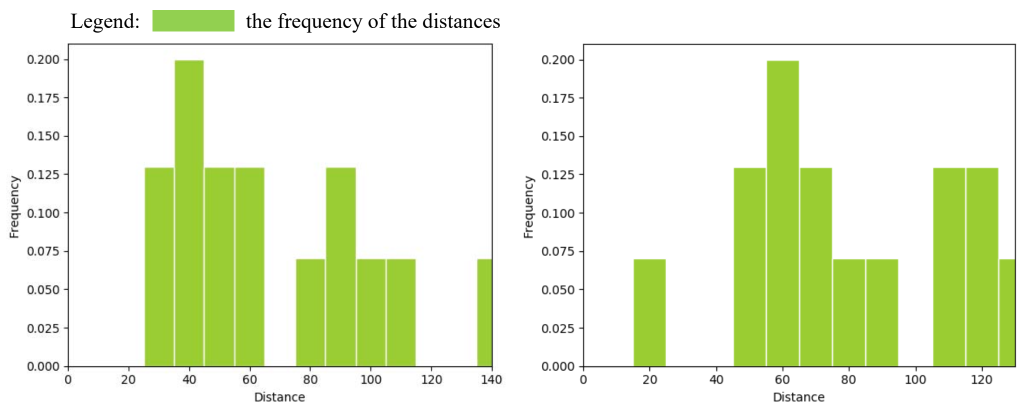
Patterns	CPFNR		Join-less		SGCT-K	
	FPI	Rank	PI	Rank	PI-K	Rank
{L, c}	**0.5871**	**1**	**0.6667**	**10**	**0.0684**	**21**
{L, Z}	0.5702	2	0.625	15	0.0846	13
{A, L}	**0.5605**	**3**	**0.75**	**2**	**0.0703**	**19**
{B, X}	**0.5335**	**4**	**0.6522**	**11**	**0.0451**	**55**
{A, J }	0.5078	5	0.6923	8	0.0627	32
{E, X}	0.4874	6	0.7222	7	0.0503	46
{A, I}	**0.4749**	**7**	**0.75**	**2**	**0.048**	**51**
{B, E}	**0.4727**	**8**	**0.6522**	**11**	**0.0417**	**67**
{A, Z}	0.4611	9	0.6154	16	0.0410	70
{A, K}	0.4601	10	0.7692	1	0.0757	14

Table 5. The (F)PR for the objects in the representative patterns.

Patterns	the FPR for CPFNR	the PR for Join-less
{A, L}	*FPR*{{A, L}, A}=0.5604	*PR*{{A, L}, A}=0.7692
	FPR{{A, L}, L}=0.5623	*PR*{{A, L}, L}=0.75
{A, I}	*FPR*{{A, L}, A}=0.4832	*PR*{{A, I}, A}=0.7692
	FPR{{A, L}, I}=0.4749	*PR*{{A, I}, I}=0.75
{B, X}	*FPR*{{B, X}, B}=0.5335	*PR*{{B, X}, B}=0.6522
	FPR{{B, X}, X}=0.7069	*PR*{{B, X}, X}=0.84615
{B, E}	*FPR*{{B, E}, B}=0.4726	*PR*{{B, E}, B}=0.6522
	FPR{{B, E}, E}=0.5416	*PR*{{B, E}, E}=0.7222

Join-less algorithm but has different prevalence levels in CPFNR algorithm. The only difference is that the distances between the instances of the object B and their fuzzy neighbors with object type X are different from that between the instance of B and their fuzzy neighbors with object type E, thus the membership values of different distance, which will be transformed to the contribution to the participation ratio of B, will be diverse in this two case. Fig. 5 shows the frequency of the distances between B's instances and their fuzzy neighbors with object type X or E for the co-locations {B, X} and {B, E}. The average distances are 66.13 for {B, X} and 80.53 for {B, E}, respectively. The smaller the distance, the larger the membership value. Therefore, there is *FPR*({B, X}, B) > *FPR*{{B, E}, B}. Further, it can be obtained that *FPI*({B, X}) > *FPI*({B, E}), which is a more accurate result than *PI*({B,X})=*PI*({B, E}).

The above analysis reveals that the CPFPR algorithm is more accurate than the Join-less algorithm. The co-location pattern {L, c} (marked by underlined text) in Table 4 ranks first in CPFNR while ranks tenth in Join-less. Namely, the plants L and c are frequently observed together than other pair of plants actually. This result is more valuable for the Botanists to make decisions.



Average = 66.13
(a) {B, X}

Average = 80.53
(b) {B, E}

Fig. 5. The frequency histograms of the distances for co-locations {B, X} and {B, E}.

Table 6. the NRR of the three metrics on the two real datasets.

Patterns	The NRR of FPI	The NRR of PI	The NRR of PI-K
Real-1	89.47	7.5	99.445
Real-2	99	87.878	98

From Table 5 and Fig. 5, we have come to conclusion that, it is more possible to share the same prevalence level using the PI than the FPI or the PI-K. To a certain extent, the *FPI* can distinguish the effects of different distances between instances that contribute to the fuzzy participation ratio. We formulate the non-repeat rate (NRR) of the prevalence metrics to evaluate the tree metrics as follows:

$$NRR = \frac{n_{non}}{n_T} * 100\%$$

where, n_T is the total number of the size-2 co-locations and n_{non} is the number of co-locations with non-repeat prevalence levels in the size-2 co-locations. A higher *NRR* indicates a more refined result.

Table 6 lists the NRRs of the three metrics on the two real datasets. We can see that the NRR of PI is the lowest of the three. The NRRs of FPI and the PI-K are much more higher than that of the PI. With more objects but less instances in Real-1, the NRR of PI on Real-1 is pretty small while both the FPI and PI-K earn a high score, which indicates that our algorithm is more stable than the Join-less algorithm.

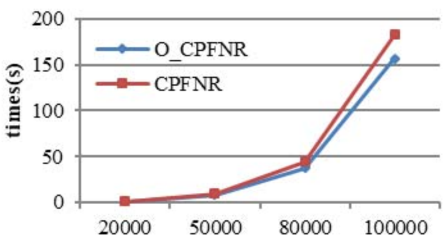
5.3 Performance Evaluation of the CPFNR Algorithm

Although the SCGT-K algorithm is the most similar to the CPFNR algorithm, it aimed to mine the maximal co-locations which is a lossy compression form of the mined co-location patterns, and the prevalence metrics PI-K of SGCT-K is extremely small (the FPI is several times larger on average than the PI-K), such that we couldn't compare the CPFNR algorithm with it for verifying the efficiency of CPFNR. We evaluate the performance of the CPFNR algorithm and its optimizing strategy (O_CPFNR) on synthetic datasets by changing the number of instances, the fuzzy prevalence threshold and the membership threshold. We also examine the scale of the results of the CPFNR algorithm. We also compare the number of the mined prevalent co-location patterns of the CPFNR algorithm and the classic Join-less algorithm by changing the number of instances, the (fuzzy) prevalence threshold and the membership threshold.

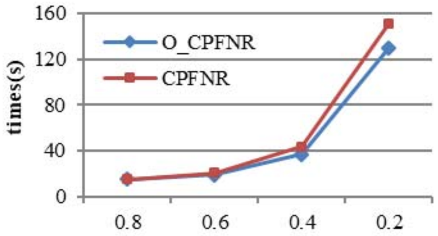
The default values of the parameters in the experiments are as follows: the number of spatial objects is 20, the number of instances is 80000, the fuzzy prevalence threshold *min_fprev* is 0.4, and the membership threshold α is 0.001, the distance threshold in the Join-less algorithm is set to be equal to the boundary distance of FNR.

(A) The efficiency of the CPFNR algorithm

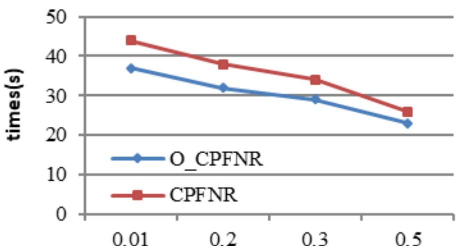
Effect of the number of instances. We first consider the influence of the different number of instances on the performance of the two algorithms. As you can see from Fig. 6 (a), both of the two algorithms show good performance. The running time of the two algorithms increases with the number of instances increasing. The CPFNR algorithm



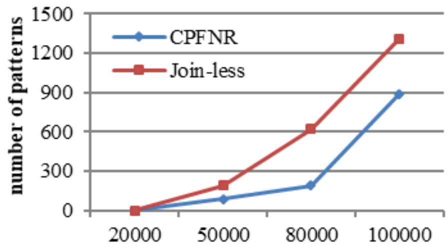
(a) Number of instances



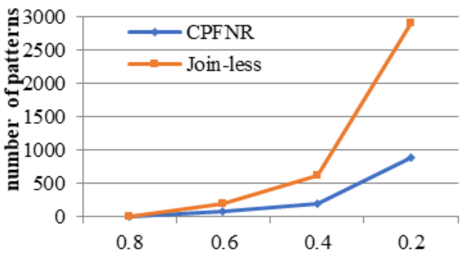
(b) *min_fprev*



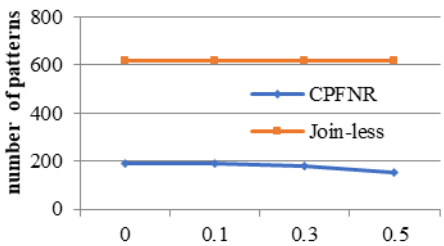
(c) Membership threshold α



(d) Number of instances



(e) *min_fprev*



(f) Membership threshold α

Fig. 6. The execution time and number of co-location patterns on different parameters.

costs more running time than the O_CPFNR algorithm, and the running time of the former grow faster than that of the latter, which indicates that the improving method perform better. We can get that the O_CPFNR algorithm reduces the time of CPFNR algorithm by up to 14.2%.

Effect of min_fprev. Then we study the effect of the different fuzzy prevalence threshold *min_fprev* on the performance of the two algorithms. In Fig. 6 (b), as *min_fprev* varies from 0.8 to 0.2, the running time of the two algorithms increases. And the running time of the O_CPFNR algorithm is less than that of the CPFNR algorithm all the time. When *min_fprev* = 0.2, the O_CPFNR algorithm reduces the time of CPFNR algorithm by up to 14%.

Effect of α. Finally we study the impact of the membership threshold α by varying α from 0.01 to 0.5. As shown in Fig. 6 (c), with α increasing, the running time of the two algorithms decreases. The reason is that FNR_α becomes small with α increasing, which means that the fuzzy neighbors involved in the mining process decrease. So the running time of the two algorithms decreases naturally. And the O_CPFNR algorithm reduces the time of CPFNR algorithm by up to 16%.

(B) Comparisons of the scale of the results

As fuzzy participation index in the CPFNR algorithm is not greater than participation index in the Join-less algorithm, the prevalent co-location set generated by CPFNR should be a subset of the result of Join-less, that is, if a co-location pattern appears in the result of CPFNR, it must also be in the result of Join-less, otherwise it is not. By the experimental results of the CPFNR and Join-less algorithms under different parameters, we evaluate the number of the prevalent co-location patterns generated of the two algorithms.

Effect of the number of instances. The effect of the number of instances on the number of prevalent co-location patterns generated by two algorithms is considered first. The number of instances increased from 20000 to 100000. As can be seen from Fig. 6 (d), the number of patterns generated by the two algorithms increases with the increase of the number of instances. Because the fuzzy participation index in the CPFNR algorithm is no larger than the participation index in the Join-less algorithm, resulting in the number of prevalent co-location patterns mined by the former will be less than by the latter when the other parameters are the same.

Effect of min_fprev. Next, the number of prevalent co-location patterns generated on different *min_fprev* by two algorithms is studied. In Fig. 6 (e), we can observe that, as *min_fprev* varies from 0.8 to 0.2, the number of patterns generated by the two algorithms grow faster. Obviously, the growth rate of the CPFNR algorithm is much faster than the Join-less algorithm.

Effect of α. In Fig. 6 (f), as the membership threshold α varies from 0 to 0.5, the number of prevalent co-location patterns mined by the Join-less algorithm remains the same, while the number of prevalent patterns obtained by the CPFNR algorithm decreases. The reason is that the Join-less algorithm does not take the proximity level into account, then the change of α will not affect its results; as the increase of α, FNR_α become small, resulting in the number of prevalent co-location patterns decreases. In addition, the number of the prevalent patterns produced by the CPFNR algorithm decreases slowly, this is because the membership values with low value are cut down in FNR_α, of which the impact on the fuzzy participation index is little, so the number of the prevalent patterns generated decreases slowly.

6. CONCLUSIONS

Although lots of the co-location pattern mining approaches were proposed, most of them neglected the proximity level between instances. This paper defines the fuzzy neighbor relationship (FNR) to measure the proximity level and proposes an efficient algorithm based on FNR (CPFNR). The experimental results demonstrate that our algorithm can distinguish the effects of different distances between instances, which will contribute to the fuzzy participation ratio (index). And due to the anti-monotonicity of the fuzzy participation index of the co-location pattern, our algorithm achieves good performance. This work is more valuable for human decision in practical application. The future work is to study the co-location pattern mining with maximum membership threshold.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (61966036, 61662086), the Natural Science Foundation of Yunnan Province (2016FA026), and the Project of Innovative Research Team of Yunnan Province (2018HC019).

REFERENCES

1. R. Sierra and C. R. Stephens, "Exploratory analysis of the interrelations between co-located boolean spatial features using network graphs," *Journal International Journal of Geographical Information Science*, Vol. 26, 2012, pp. 441-468.
2. X. Wang, Y. Zhao, L. Nie, *et al.*, "Semantic-based location recommendation with multimodal venue semantics," *IEEE Transactions on Multimedia*, Vol. 17, 2015, pp. 409-419.
3. D. G. Leibovici, C. Claramunt, D. Guyader, and D. Brosset, "Local and global spatio-temporal entropy indices based on distance-ratios and co-occurrences distributions," *International Journal of Geographical Information Science*, Vol. 28, 2014, pp. 1061-1084.
4. M. Akbari, F. Samadzadegan, and R. Weibel, "A generic regional spatio-temporal co-occurrence pattern mining model: a case study for air pollution," *Journal of Geographical Systems*, Vol. 17, 2015, pp. 249-274.
5. S. Shekhar and Y. Huang, "Discovering spatial co-location patterns: a summary of results," in *Proceedings of the 7th International Symposium on Advances in Spatial and Temporal Databases*, 2001, pp. 236-256.
6. J. S. Yoo and S. Shekhar, "A partial join approach for mining colocation patterns," in *Proceedings of the 12th Annual ACM International Workshop on Geographic Information Systems*, 2004, pp. 241-249.
7. J. S. Yoo, S. Shekhar, and M. Celik, "A join-less approach for co-location pattern mining: a summary of results," in *Proceedings of IEEE International Conference on Data Mining*, 2005, pp. 813-816.
8. L. Wang, Y. Bao, J. Lu, and J. Yip, "A new join-less approach for co-location pattern mining," in *Proceedings of IEEE 8th International Conference on Computer and Information Technology*, Vol. 164, 2008, pp. 197-202.
9. L. Wang, Y. Bao, and Z. Lu, "Efficient discovery of spatial co-location patterns using the iCPI-tree," *Open Information Systems Journal*, Vol. 3, 2009, pp. 69-80.
10. L. Wang, L. Zhou, J. Lu, *et al.*, "An order-clique-based approach for mining maximal co-locations," *Information Sciences*, Vol. 179, 2009, pp. 3370-3382.
11. X. Bao, L. Wang, and J. Zhao, "Mining top-*k*-size maximal co-location patterns," in *Proceedings of International Conference on Computer*, *Information and Telecommunication Systems*, 2016, doi: 10.1109/ CITS.2016.7546421.
12. J. S. Yoo and M. Bow, "Mining maximal co-located event sets," in *Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2011, pp. 351-362.
13. X. Yao, L. Peng, L Yang, and T. Chi, "A fast space-saving algorithm for maximal co-location pattern mining," *Expert System with Applications*, Vol. 63, 2016, pp.

310-323.
14. J. Lu, L. Wang, Q. Xiao, *et al.*, "Incremental mining of co-locations from spatial database," in *Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery*, 2015, pp. 648-653.
15. J. Lu, L. Wang, Y. Fang, *et al.*, "Mining competitive pairs hidden in co-location patterns from dynamic spatial databases," in *Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2017, pp. 467-480.
16. L. Wang, X. Bao, and L. Zhou, "Redundancy reduction for prevalent co-location patterns," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 30, 2018, pp. 142-155.
17. L. Wang, X. Bao, and H. Chen, "Effective lossless condensed representation and discovery of spatial co-location patterns," *Information Sciences*, Vol. 436-437, 2018, pp. 197-213.
18. S. Yang, L. Wang, X. Bao, *et al.*, "A framework for mining spatial high utility co-location patterns," in *Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery*, 2015, pp. 631-637.
19. L. Wang, W. Jian, H. Chen, and Y. Fang, "Efficiently mining high utility co-location patterns from spatial data sets with instance-specific utilities," in *Proceedings of the 22nd International Conference on Database Systems for Advanced Applications*, 2017, pp. 458-474.
20. J. Zhao, L. Wang, *et al.*, "Mining high utility co-location patterns based on importance of spatial region," in *Proceedings of International Conference on Geo-Spatial Knowledge and Intelligence*, 2017, pp. 43-55.
21. A. Garaeva, F. Makhmutova, *et al.*, "A framework for co-location patterns mining in big spatial data," in *Proceedings of IEEE International Conference on Soft Computing and Measurements*, 2017, pp. 477-480.
22. P. Yang, L. Wang, and X. Wang, "A parallel spatial co-location pattern mining approach based on ordered clique growth," in *Proceedings of International Conference on Database Systems for Advanced Applications*, 2018, pp. 734-742.
23. L. Wang, P. Wu, and H. Chen, "Finding probabilistic prevalent co-locations in spatially uncertain data sets," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25, 2013, pp. 790-804.
24. L. Wang, P. Guan, H. Chen, and L. Zhao, "Mining co-locations from spatially uncertain data with probability intervals," in *Proceedings of the 14th International Conference on Web-Age Information Management*, 2013, pp. 301-314.
25. L. Wang, H. Chen, L. Zhou, *et al.*, "Efficiently mining co-location rules on interval data," in *Proceedings of the 6th International Conference on Advanced Data Mining and Applications*, 2010, pp. 477-488.
26. Y. Huang, J. Pei, and H. Xiong, "Mining co-location patterns with rare events from spatial data sets," *GeoInformatica*, Vol. 10, 2006, pp. 239-260.
27. Z. Ouyang, L. Wang, and P. Wu, "Spatial co-location pattern discovery from fuzzy objects," *International Journal of Artificial Intelligent Tools*, Vol. 26, 2017, doi: 10.1142/S0218213017500038.
28. P. Wu, L. Wang, and Y. Zhou, "Discovering co-location from spatial data sets with fuzzy attributes," *Journal of Frontiers of Computer and Technology*, Vol. 7, 2013, pp. 348-358.

29. X. Yao, L. Chen, L. Peng, *et al*., "A co-location pattern-mining algorithm with a density-weighted distance thresholding consideration," *Information Sciences*, Vol. 396, 2017, pp. 144-161.

Mei-Jiao Wang (王美姣) received her B.S. degree in Computer Science from Guizhou Normal University, China, in 2007, and her M.S. degree in Computer Application Technology from Yunnan University, China, in 2010. She is working toward the Ph.D. degree in Communication and Information Systems at Yunnan University, since 2015. Her main research interests include spatial data mining.

Li-Zhen Wang (王丽珍) received her M.S. degree in Computational Mathematics from Yunnan University, China, in 1988 and her Ph.D. in Computer Science from the University of Huddersfield, UK, in 2008. She is currently a Professor and Ph.D. supervisor at Yunnan University. Her current research interests include spatial data mining, big data analytics, interactive data mining, and their applications.

Li-Hong Zhao (趙丽红) received the M.S. degree in Computer Science from George Mason University, the United States of American. Her research interests include data mining and communication networks.