ARTICLE IN PRESS

Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Adaptive detection of statistically significant regional spatial co-location patterns

Jiannan Cai, Qiliang Liu*, Min Deng, Jianbo Tang, Zhanjun He

Department of Geo-informatics, Central South University, Changsha 410083, China

A R T I C L E I N F O

Keywords:
Spatial heterogeneity
Regional spatial co-location patterns
Significance test
Non-parametric
Adaptive spatial clustering

A B S T R A C T

Regional spatial co-location patterns refer to subsets of spatial features that often co-occur in close geographical proximity in certain localities of space. Discovering regional spatial co-location patterns is still very challenging because it is difficult to specify appropriate thresholds for prevalence measures without prior knowledge and to detect natural localities of regional spatial co-location patterns automatically. On that account, an adaptive method is proposed in this study. First, a non-parametric significance test is constructed to evaluate the prevalence of spatial co-location patterns. Then, an adaptive pattern clustering approach is developed to detect hotspots of each candidate regional spatial co-location pattern. Finally, all statistically significant regional spatial co-location patterns and their localities are detected by iteratively expanding these hotspots. Comparisons between this adaptive method and two state-of-the-art methods are carried out with both simulated and ecological datasets (i.e. a wetland species dataset in northeast China). Experiments show that the proposed adaptive method allows detecting regional spatial co-location patterns effectively and with less prior knowledge than the state-of-the-art methods.

1. Introduction

In spatial datasets, positive interactions among different geographic phenomena usually indicate spatial co-location patterns formed by subsets of spatial features that are frequently located together, such as symbiotic species in ecology (Keddy, 2010). Discovery of spatial co-location patterns is vital in many application domains, such as ecology, epidemiology, business, criminology, climatology, environmental science, etc. (Hu, 2008). Because of spatial heterogeneity, spatial co-location patterns are usually geographically regional (Ding, Eick, Yuan, Wang, & Nicot, 2011). Regional spatial co-location patterns can be represented as a collection of spatial features that are frequently located together in certain localities (i.e. sub-regions) in the study area (Mohan et al., 2011). For example, in ecology, symbiotic relationships among species usually vary in different places due to various environmental factors (e.g. illuminance, water content, soil property, etc.). Discovery of regional spatial co-location patterns will provide more insight into the exploration of spatial interactions among geographic phenomena.

Existing regional spatial co-location pattern discovery methods usually need to first identify localities in the study area where regional spatial co-location patterns may exist, and then the prevalence of regional co-location patterns should be evaluated in each locality. In existing methods, a series of user-specified parameters are required to evaluate the prevalence of regional co-location patterns and to detect localities of regional co-location patterns. In practice, it is difficult to determine these parameters, and inappropriate parameters may result in omission or misstatement of regional co-location patterns. To reduce the difficulty in determination of these parameters, an adaptive method is proposed in this study. The prevalence of regional co-location patterns is statistically evaluated by using a non-parametric significance test. The significance level specified in significance test has clear meanings in statistics and can be easily set in practice. An adaptive pattern clustering approach without user-specified parameters is developed to detect localities of regional spatial co-location patterns in an automatic way.

The rest of this paper is organized as follows. In Section 2, related work on mining regional spatial co-location patterns is reviewed and analyzed, followed by our new strategy. The adaptive method is fully performed in Section 3 and 4. In Section 5, the proposed method is compared with existing methods using both simulated and real datasets. In Section 6, conclusions of this paper together with directions of future work are given.

* Corresponding author at: Department of Geo-informatics, Central South University, Changsha, Hunan, China.
E-mail addresses: jiannan.cai@csu.edu.cn (J. Cai), qiliang.liu@csu.edu.cn (Q. Liu), dengmin028@yahoo.com (M. Deng), jianbo.tang@csu.edu.cn (J. Tang), zhanjun.he@csu.edu.cn (Z. He).

http://dx.doi.org/10.1016/j.compenvurbsys.2017.10.003
Received 20 October 2016; Received in revised form 24 October 2017; Accepted 26 October 2017
0198-9715/ © 2017 Elsevier Ltd. All rights reserved.

Please cite this article as: Cai, J., Computers, Environment and Urban Systems (2017), http://dx.doi.org/10.1016/j.compenvurbsys.2017.10.003

2. Discovering regional spatial co-location patterns: existing work and a new strategy

Two methods for detecting regional spatial co-location patterns are currently available, namely partitioning- and clustering-based methods.

2.1. Partitioning-based methods

Partitioning-based methods first partition the whole study area into several local regions and then extract prevalent spatial co-location patterns in each region. The methods identify each prevalent spatial co-location pattern discovered in a local region as a regional spatial co-location pattern and use the local regions containing that regional spatial co-location pattern to represent its localities.

The existing partitioning-based methods mainly differ in the space partitioning schemes used to identify local regions. For example, Celik, Kang, and Shekhar (2007) applied a user-specified quad-tree-based zoning scheme to partition the study area. To reduce the subjectivity in determining a partitioning scheme, Qian, Chiew, He, and Huang (2014) developed a *k*-nearest-neighbor-based partitioning method that partitions the study area into several homogenous sub-regions. In each sub-region, the weights of the edges in the *k*-nearest-neighbor graph vary slightly.

Although partitioning-based methods can discover some regional spatial co-location patterns that cannot be detected at a global level, the user-specified space partitioning schemes cannot precisely determine the actual localities of the regional spatial co-location patterns (Mohan et al., 2011). Therefore, it is possible to miss or incorrectly segment the localities of the regional patterns. To overcome this limitation, clustering-based methods were developed.

2.2. Clustering-based methods

Clustering-based methods consider localities of a regional co-location pattern as clusters formed by concentrated and co-located instances of spatial features that participate in that pattern.

When a spatial dataset is organized similar to the market basket dataset, in which a transaction consists of a collection of item types purchased together by a customer, a spatial transaction of the spatial dataset can be defined as a set of adjacent instances of different features (Shekhar & Huang, 2001). Then, localities of regional co-location patterns can be detected by clustering these spatial transactions. For example, Eick et al., (2008) proposed a prototype-based clustering method in which the objective function is designed based on the products of the *z*-scores of continuous variables. Prior knowledge is required to set the number of clusters, and the method may be computationally expensive due to the brute-force testing of all possible patterns. To overcome these limitations, Ding et al. (2011) first developed a grid-based clustering method to identify hot spots of an interesting feature, in which a fitness function designed based on probability distribution of that feature was maximized. Then, regional spatial co-location patterns associated with that interesting feature were identified in each hot spot. Finally, the grid-based clustering method was employed to detect the localities of each regional co-location pattern by using another fitness function designed based on the prevalence of co-location patterns.

To detect regional co-location patterns without the definition of spatial transactions, Mohan et al. (2011) further developed a neighborhood graph based method. The method used the neighborhood graph to connect the adjacent instances of a co-location pattern and identified the convex hulls of the connected instances as candidate localities. It then evaluated the prevalence of the regional patterns in each candidate locality using a regional participation index. Some clustering methods developed based on spatial scan statistics also can be used to detect regional co-location patterns. For example, Wang, Huang, and Wang (2013) developed a heuristic region expansion method based on frequentist and Bayesian statistics. This method only can be used for discovery of 2-size regional co-location patterns. Jung, Kulldorff, and Richard (2010) proposed a multinomial spatial scan statistic. A cluster detected by this statistic indicates that the intensity of at least one kind of spatial features in a sub-region (or a scanning window) is higher than expected. It has been found that the clusters discovered by multinomial spatial scan statistic cannot always guarantee the co-located relationship among different spatial features (Páez, López-Hernández, Ortega-García, & Ruiz, 2015). In addition, these methods only report the most likely cluster (locality with maximized statistic), and some localities may be omitted due to the shadowing effect of the most likely cluster (Li, Wang, Yang, Li, & Lai, 2011).

2.3. A critical analysis of existing work

Although some efforts have been made to discover regional spatial co-location patterns, the mining results are seriously affected by user-specified parameters for evaluating the prevalence of regional co-location patterns and detecting localities of regional co-location patterns:

(a) In almost all the existing methods, a prevalent regional co-location pattern is usually determined by comparing the value of certain prevalence measure (e.g. participation index) with a user-specified prevalence threshold. In practice, determining this threshold without any domain knowledge is difficult. Although the significance test developed by Barua and Sander (2014a, 2014b) does not require user-specified prevalence threshold, this significance test only can detect co-location patterns at global level. In addition, user-specified parameters are also required to define the spatial distribution of the data sets.

(b) Although existing clustering-based methods are more flexible than the partitioning-based methods in discovering the localities of regional co-location patterns (user-specified space partitioning schemes are not required), user-specified clustering parameters, such as the number of clusters (Eick, Parmar, Ding, Stepinski, & Nicot, 2008) or the size of the grids (Ding et al., 2011; Wang et al., 2013), are still difficult to determine.

To reduce the difficulty in determination of user-specified parameters in regional co-location pattern discovery, a new strategy should be developed.

2.4. An adaptive strategy for mining statistically significant regional spatial co-location patterns

Based on the analysis of existing work, we found that the participant spatial features in each locality of a regional spatial co-location pattern should maintain the following two properties:

(1) Co-located: the instances of different spatial features are frequently located in close geographic proximity.
(2) Clustered: the density of instances of that regional co-location pattern is higher than expected.

Fig. 1 shows an example regional spatial co-location pattern {A, B}. One can see that in the localities I and II, the instances of features A and B are co-located, and the instances of pattern {A, B} are clustered.

Motivated by these two properties, an adaptive strategy is proposed in this study. The procedure of the adaptive strategy can be described as follows:

(1) Generation of candidate regional co-location patterns: although regional co-location patterns and their localities are prior unknown, the candidate regional co-location patterns and their instances can be generated by identifying non-prevalent co-location patterns in the whole study area. To evaluate the prevalence of a spatial co-location pattern without user-specified parameters, a non-parametric significance test based on the pattern reconstruction will be constructed in this study.

Download English Version:

https://daneshyari.com/en/article/6921869

Download Persian Version:

https://daneshyari.com/article/6921869

Daneshyari.com