Knowledge-Based Systems 000 (2018) 1–13

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Mining strong symbiotic patterns hidden in spatial prevalent co-location patterns

Junli Lu, Lizhen Wang*, Yuan Fang, Jiasong Zhao

Department of Computer Science and Engineering, School of Information Science and Engineering, Yunnan University, Kunming, China

ARTICLE INFO

Article history:
Received 29 April 2016
Revised 23 January 2018
Accepted 2 February 2018
Available online xxx

Keywords:
Spatial data mining
Spatial co-location patterns
Strong symbiotic patterns
Dynamic spatial databases

ABSTRACT

Spatial co-location patterns represent the subsets of spatial features which are frequently located together in a geographic space. Spatial co-location pattern mining has been a research hot in recent years. However, maybe the features in a prevalent co-location pattern further have more interesting relationships such as symbiotic relationships, competitive relationships or causal relationships. This paper mines symbiotic relationships implied in prevalent co-location patterns from dynamic spatial databases. Firstly, after analyzing the existed definition of symbiotic patterns, a criterion of judging strong symbiotic patterns is proposed. Secondly, a novel algorithm to mine strong symbiotic patterns from prevalent co-location patterns is presented, named basic algorithm. Third, for improving the efficiency of the basic algorithm, an improved algorithm which integrates two expensive operations of the basic algorithm into together, and a pruning strategy with two pruning lemmas are presented. The experiments evaluate the effectiveness and efficiency of the proposed algorithms with “real + synthetic” data sets and the results show that strong symbiotic patterns are more concise and actionable compared to traditional prevalent co-location patterns.

1. Introduction

The explosive growth of spatial data and widespread use of spatial database emphasize the need of automated discovery of spatial knowledge [1]. Spatial data mining aims to discover interesting, previous unknown, but potentially useful spatial knowledge. Spatial co-location pattern mining is one of main research directions in spatial data mining. Due to the importance and difficulty of spatial co-location pattern mining, spatial co-location pattern mining has been a research hot. Many works [6–19] have been presented.

Spatial co-location pattern mining aims at finding subsets of spatial features frequently located together in spatial neighborhoods. Symbiotic patterns are divided into three main categories in Wikipedia [2]. When each feature of the pattern benefit from the relationship, the pattern is called *mutualism*. When some features benefit and others aren’t affected, the pattern is called *commensalism*. When some features benefit, and others are harmed, the pattern is called *parasitism*. The “benefit” [2] in symbiotic pattern means each one or some of the features entirely depend on each other for survival. The objective of this paper is to find the mutualism patterns from spatial databases, which show the strongest symbiotic relationship in the three categories and are called *strong symbiotic patterns*. Commensalism patterns are called *general symbiotic patterns* and parasitism is not referred to. The relationship of the patterns is showed in Fig. 1. Strong symbiotic patterns and general symbiotic patterns are contained in prevalent co-location patterns (see the details on the existed definition of symbiotic patterns as below).

Now, we give three examples to explain prevalent co-location patterns, general symbiotic patterns and strong symbiotic patterns. Example 1: restaurants and ice cream shops both appear in the same community, university town, or downtown street. Example 2: T. matsutake and pine trees are frequently located together. Example 3: sea anemones and hermit crabs are often seen attaching together. In prevalent co-location pattern mining, all the three examples can be discovered and be treated in the same way. However, pattern {restaurants, ice cream shops} is a prevalent co-location pattern but not a symbiotic pattern. Since neither restaurants nor ice cream shops can get benefit from the co-located relationship, each of them can survive without the other one. {T. matsutake, pine trees} is a prevalent co-location pattern and a general symbiotic pattern. Since T. matsutake relies on the appropriate environment by the shadow of pine trees, although pine trees do not get anything from their co-located relationship. {Sea anemones, hermit crabs} is a prevalent co-location pattern and a strong symbiotic pattern. Sea anemones attach themselves to the

* Corresponding author.
E-mail addresses: ljl1982_3_6@126.com (J. Lu), lzhwang@ynu.edu.cn (L. Wang), fy1990825@163.com (Y. Fang), zhaojs75@163.com (J. Zhao).

https://doi.org/10.1016/j.knosys.2018.02.006
0950-7051/

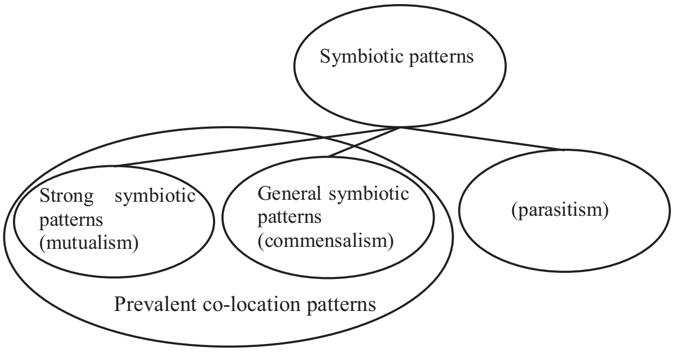

Fig. 1. The relationship of the patterns.

shell of hermit crabs, so sea anemones can eat the food particles that are left floating when a hermit crab eats. Hermit crabs can benefit from the sea anemones that can sting predators that put the hermit crabs in danger. Therefore, general symbiotic patterns and strong symbiotic patterns have more information than prevalent co-location patterns.

Another typical example of strong symbiotic patterns is termites and trichonympha, termites live on wood but they cannot digest the wood fibers. Trichonympha, a genus of parabasalid protists that live in the intestines of termites, can help break down the cellulose in the wood and plant fibers termites eat. And they themselves are fed by engulfing wood and plant fibers through phagocytosis. Symbiotic interactions of microorganisms are widespread in nature, and support fundamentally important processes in diverse areas of biology that range from health and disease to ecology and the environment [3]. Strong symbiotic patterns in this paper refer to not only microorganisms, but also other events such as plants, organizations, services. Strong symbiotic patterns can be used in prediction and reasoning. With identified strong symbiotic patterns, we can predict potential danger before actually carrying out some actions.

The term symbiotic pattern was first used by Heinrich Anton de Bary, a German botanist, in 1879 to describe the relationships between fungi and algae in the formation of lichens. His speech was translated into English and published in the journal "Symbiosis" in 2016 [4]. He referred to "the living together of two dissimilar organisms, usually in intimate association, and usually to the benefit of at least one of them." It can be considered as the definition of general symbiotic pattern which meets two conditions, (1) they live together, and in intimate association, and (2) they get benefit of at least one of them. The first condition refers to they frequently located together (i.e., they forms a prevalent co-location pattern) and the second condition refers to "they get benefit of at least one of them". Therefore, this paper identifies strong symbiotic patterns from prevalent co-location patterns. And the second condition of strong symbiotic patterns is adjusted as "each feature in it gets benefit from the others". However, "each feature in it gets benefit from the others" is hard to identify. In ecological science or biology, it needs a lot of controlled experiments, observations and domain knowledge. It unfortunately is often expensive to alter artificially each feature and observe the change of the other features, and it is also infeasible with large number of prevalent co-location patterns and features. We discover "each feature in it gets benefit from the others" from observational data in this paper, and it needs time to examine, so the methods are conducted on dynamic spatial databases. We have known the "benefit" in symbiotic patterns is not ordinary benefit, and the feature cannot live without the "benefit" from the other features. So, "each feature in it gets benefit from the others" means the features share life or death.

We assume one spatial database at two selected time points as old database S_{old} and updated database $S_{updated}$. Comparing with S_{old}, some data disappeared and some data added in $S_{updated}$. For a prevalent co-location pattern c, if c's features in their neighborhoods (i.e., they located together) disappeared or added consistently in $S_{updated}$, it means they share life or death. Many application domains [5] include location-based services, rare animals and plants protection, transportation and environmental monitoring which collecting their data periodically or continuously.

Discovering strong symbiotic patterns from dynamic spatial databases, however, is also a challenging task. Firstly, this is the first time to solve the problem of mining strong symbiotic patterns hidden in spatial prevalent co-location patterns, so a generally acceptable definition about this problem is challenging. Secondly, the computational cost for the discovery is high due to large number of the prevalent co-location patterns and amount of features in a co-location induced by a small minimum prevalence threshold and a large distance threshold. Thirdly, the existent algorithms cannot be reused to discover strong symbiotic patterns.

In this paper, we propose an approach of strong symbiotic pattern discovery. In comparison with previous works, we have made three main contributions.

Firstly, we discover "each feature in it gets benefit from the others" from dynamic spatial databases and give the criterion of judging strong symbiotic patterns, which is strict and rational.

Secondly, we propose a basic algorithm to mine the strong symbiotic patterns, it is novel and reasonable.

Thirdly, an improved algorithm which integrate one expensive operation into another, and a pruning strategy with two pruning lemmas, are proposed to reduce the computation considerably and make it computational feasible.

Scope: This paper focuses on mining strong symbiotic patterns from prevalent co-location patterns having been given a database with two time points. The following issues are outside the scope of this paper: (i) selecting two time points of database for getting more accurate strong symbiotic pattern set, (ii) determining thresholds for prevalent co-location pattern mining, and (iii) indexing and query processing issues related to generate co-changed co-location instances.

The rest of paper is organized as follows. Related work is introduced in Section 2. In Section 3, we give the preliminary and definitions to measure the strong symbiotic patterns. Section 4 describes our algorithms for mining strong symbiotic patterns, which include a basic algorithm, an improved algorithm and a pruning strategy. Then in Section 5 we experimentally show the effectiveness and efficiency of the algorithms and compare the results with prevalent co-location patterns. In Section 6 we conclude the paper and suggest the future works.

2. Related work

Research on spatial co-location pattern mining has attracted much attention. Huang et al. [6] firstly proposed a general framework for co-location pattern mining and defined the minimum participation ratio to measure the prevalence of a co-location. It is like Apriori [7] method, a classical method in association rule mining proposed by Agrawal and Srikant in 1993. After that, many researchers did lots of works on improving the efficiency of mining process [1,8,9], fitting for different data types [10–11] and for special cases [12–19]. Partial join [8] and joinless [1] avoid the expensive join operation [6]. A compact format, prefix-tree, is used in [9] to store star neighborhoods and help to efficiently prune candidates. Spatial co-location patterns for fuzzy objects and uncertain objects are studied in [10] and [11,12] respectively. The maximum participation ratio is proposed in [13] for the co-location patterns with rare features, while it is improved by a weighted participation