Computer Science and Application 计算机科学与应用, 2018, 8(3), 328-338
Published Online March 2018 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2018.83038

Hans汉斯

Research on Spatial Co-Locations Mining: A Survey

Pingping Wu[1], Lizhen Wang[1,2], Shikun Deng[1], Yu'e Liu[1]

[1]Department of Computer Science and Engineering, Dianchi College, Yunnan University, Kunming Yunnan
[2]School of Information Science and Engineering, Yunnan University, Kunming Yunnan
Email: fjwpingping@126.com

Received: Mar. 10th, 2018; accepted: Mar. 21st, 2018; published: Mar. 28th, 2018

Abstract

Due to the widespread use of mobile phones, GPS, sensors and other wireless devices, spatial data set is rapidly growing. The unique complexity and widely application in the real world make spatial data mining a promising field. As one of the important researches in spatial data mining, the spatial co-location pattern mining attracts more and more attention. Spatial co-location pattern mining aims to find the spatial features whose instances frequently co-located in neighborhood. We briefly introduce the current research of spatial co-location pattern mining from three aspects: type of co-location pattern, method of mining and application. At last, we conclude some interesting challenges in the field of co-location pattern mining.

Keywords

Spatial Pattern Mining, Co-Location Pattern, Algorithm

空间并置模式挖掘研究

吴萍萍[1]，王丽珍[1,2]，邓世昆[1]，刘玉娥[1]

[1]云南大学滇池学院理工学院计算机科学与工程系，云南 昆明
[2]云南大学信息学院，云南 昆明
Email: fjwpingping@126.com

收稿日期：2018年3月10日；录用日期：2018年3月21日；发布日期：2018年3月28日

摘 要

由于移动电话、GPS、传感器和其他的无线设备的广泛使用，空间数据迅速增长。空间数据独特的复杂性以及其在现实中的广泛应用，使得空间数据挖掘成为一个很有前途的发展方向。而作为空间数据挖掘

文章引用: 吴萍萍, 王丽珍, 邓世昆, 刘玉娥. 空间并置模式挖掘研究[J]. 计算机科学与应用, 2018, 8(3): 328-338.
DOI: 10.12677/csa.2018.83038

领域的重要任务之一，空间并置模式挖掘也受到越来越多的关注。空间并置模式挖掘旨在寻找其实例在邻近域中频繁地并置出现的空间特征。本文从挖掘的并置模式类型、挖掘方法和应用三个方面简要介绍空间并置模式挖掘的研究现状，并总结了研究中一些有趣的挑战。

关键词

空间模式挖掘，并置模式，算法

1. 引言

一般情况下，空间数据是指存放在地理数据库中与地球空间相关的数据。人们相信收集到的空间数据肯定是有价值的，但如何从这些数据中分析提取出有用的信息已成为人们必须面对的挑战。空间数据挖掘中空间数据独特的复杂性以及其在现实中的广泛应用，使得空间数据挖掘成为一个很有前途的发展方向。作为空间数据挖掘领域的重要任务之一，空间并置模式挖掘也受到了越来越多的关注。空间并置模式即 co-location 模式，也称为空间同位模式，是指在同一区域内频繁关联的特征的集合。目前，国内外许多专家学者对空间并置模式进行了深入的研究。空间并置模式挖掘研究主要围绕三个问题展开：挖掘的模式是什么类型的？采用什么方法挖掘？挖掘算法的应用前景如何？本文先从一个引例出发介绍并置模式的应用动机，再从挖掘的模式类型、挖掘方法和应用三个方面来介绍空间并置模式挖掘的研究现状。简要概述如图 1 所示。

2. 引例

我们把空间数据集中的不同事物称之为特征，同一事物的若干对象称之为实例。形式化定义如下：

定义 1 空间特征和空间特征的实例。空间特征代表了空间中不同种类的事物。空间特征集代表了空

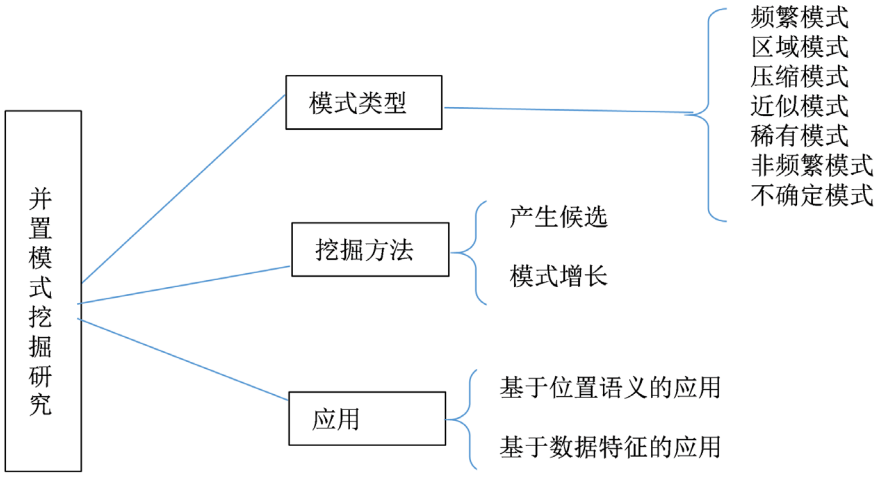

Figure 1. Spatial Co-locations mining
图 1. 并置模式挖掘研究

间中不同种类事物的集合，令 $\mathcal{F} = \{f_1, f_2, \cdots, f_n\}$是空间特征的集合。对于一个空间数据集合$\mathcal{L}$，$\mathcal{L} = L_1 \cup L_2 \cup \cdots \cup L_n$，其中 L_i $(1 \le i \le n)$是对应空间特征 f_i 的实例集合。

空间数据有很强的自相关性[1]，即空间中每一个事物都与其它事物相关，事物越邻近，其相关性越强。因此，感兴趣的特征很可能在邻近的空间上共存。空间并置模式在空间数据中发现用户感兴趣的、潜在有用的模式，在现实中有广泛的应用。

例 1 如图 2 是一个空间数据集的经典例子。图中有枯木、火灾、房屋和鸟共 4 个特征，每个特征各有若干个实例，总的有 16 个实例。观察图可以得出，枯木和火灾的实例、房屋和鸟的实例频繁地出现在一起。这种实例在空间邻近位置上频繁出现的特征集就称为并置模式，并置模式挖掘通过分析共生物种、具有燃烧源的火灾事件、自然植被间的相互影响等空间特征的关联关系可以帮助生物学家发现隐含的生态联系。

定义 2 空间并置模式。一个并置模式 F 是一个空间特征集合的子集，即 $F = \{f_1, f_2, \cdots, f_k\} \subseteq \mathcal{F}$。如例 1 中，{枯木，火灾}是一个并置模式。模式里空间特征的个数称为此模式的阶，记为$|F|$。例如，|{枯木，火灾}| = 2。

除了生物学上的应用，并置模式挖掘还广泛应用在其他的领域中，比如，在公共健康方面，用来寻找流行疾病的源头；在国防中，用来寻找不同寻常的事件；在商业服务方面，通过分析各类服务请求的位置依赖规律，帮助定制位置相关的服务广告。并置模式挖掘的应用现状将在第 4 节更详细的介绍。

3. 模式类型

由于应用需求的扩展，人们从不同类型的模式中得到有趣的、有助于决策的知识，本节将介绍空间并置模式挖掘研究中常见的一些模式类型。

3.1. 频繁模式

所谓的频繁模式指的是比较流行的模式。频繁模式的一般定义如定义 3 所示。

定义 3 如果并置模式 F 的流行程度的度量值 prevalence (F)大于等于给定阈值 *min_prev*，即 prevalence (F) $\ge$ *min_prev*，则称模式 F 为频繁模式。

Morimoto 2001 年在论文[2]中定义了基于距离的模式，称之为 k 邻近类集(k-neighboring class sets)，

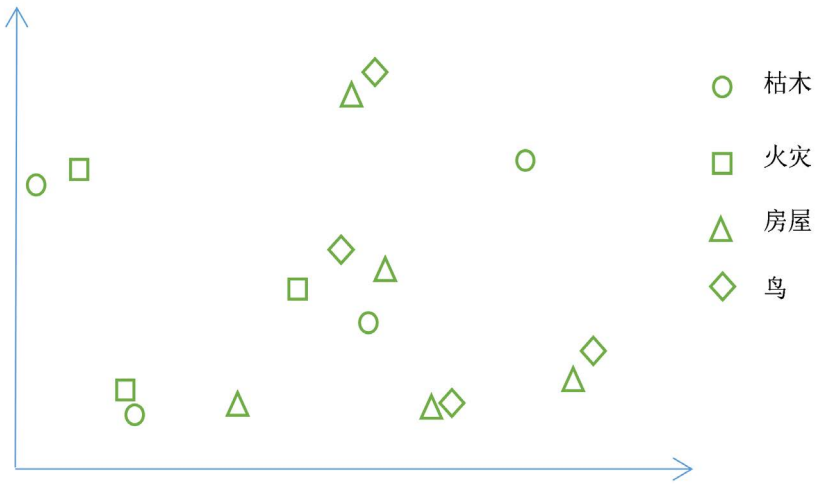

Figure 2. A classic example of spatial data sets
图 2. 空间数据集经典例子

第一次从数据挖掘的角度关注地理空间数据集上的紧密的相关关系。论文[2]采用了模式的实例的数目作为模式流行程度的度量，2004 年论文[3]中提出了参与度(participation index，PI)概念作为新的度量标准，并证明了参与度在空间统计上的可解释性，以及频繁并置模式和挖掘结果的完备性和正确性。在频繁模式挖掘上还有许多相关的工作，如文献[4] [5] [6] [7]。本节将采用文献[3]提出的框架来介绍频繁模式挖掘的基本定义。

定义 4 行实例和表实例。设 R 是一个给定的空间邻近关系，可以是空间拓扑关系、距离关系、混合关系等。当两个实例间满足关系 R 时，我们称这两个实例 R 邻近。给定模式 $F = \{f_1, f_2, \cdots, f_k\}$，一个实例集合 $L = \{l_1, l_2, \cdots, l_k\}$，若 $\forall i \in [1,k]$，有 l_i 是 f_i 的一个实例。$\forall i, j \in [1,k]$，有 l_i 和 l_j 之间 R 邻近，则称 L 是模式 F 的行实例。

一个并置模式 F 的**表实例**指该模式的所有行实例的集合，记为 $TI(F)$。

定义 5 参与率和参与度。参与率：给定 k 阶模式 $F = \{f_1, f_2, \cdots, f_k\}$，特征 $f_i \in F$ 的参与率 $PR(F, f_i)$指的是 f_i 在 $TI(F)$中不重复出现的实例个数与其总实例个数的比率，即

$$\mathrm{PR}(F, f_i) = \frac{\left|\pi_{f_i}(TI(F))\right|}{\left|TI(\{f_i\})\right|}$$

模式 F 的**参与度**取模式中的所有特征的参与率中的最小值，即

$$\mathrm{PI}(F) = \min_{i \in [1,k]} PR(F, f_i)$$

例 2 设空间数据库中，空间特征 A 有 3 个实例 $A.1$、$A.2$ 和 $A.3$，B 有 4 个实例 $B.1$、$B.2$、$B.3$ 和 $B.4$。对于并置模式 $F = \{A,B\}$，它的表实例为{{$A.2$,$B.1$}，{$A.3$,$B.3$}，{$A.3$,$B.4$}}。则根据定义，A 在模式 F 上的参与率 $PR(F, A) = 2/3$，类似地，$PR(F, B) = 3/4$。模式 F 的参与度：$PI(F) = \min(PR(F, A), PR(F, B)) = 2/3$。根据文献[3]，参与度为模式流行程度的度量，若给定阈值 $min\_prev = 0.6$，由于 $PI(F) = 2/3 > 0.6$，则本例子中模式 F 为频繁模式。

3.2. 区域模式

例 3 在商业服务方面，服务提供商尝试推出某个省、某个地区或者全国范围的服务套餐，不同区域内经济、文化等差异会导致不同范围内的服务请求很可能是不一样。

在各类应用中，经常会出现例子中对不同区域的挖掘要求。

根据不同的区域划分标准，区域模式可以分为基于位置信息的区域模式、基于位置语义信息的区域模式和基于特征语义信息的区域模式这三类。

基于位置信息的区域模式

例 3 中就是一个典型的根据地理位置来划分区域，把这类问题称为基于位置信息的区域模式挖掘。论文[8]关注空间矩形区域内的模式挖掘问题，挖掘矩形区域和全局的频繁模式，值得注意的是论文中还讨论了区域的动态变化引发的重新计算问题。频繁模式挖掘时会出现有些模式的实例集中出现在几个区域里而另一些模式的实例是分布在全局范围的，显然这是不一样的模式。论文[9]提出用离散度描述模式实例的空间分布特征，用户通过调整离散度来挖掘在空间中均匀出现的模式或者在空间中集中出现(在部分区域里出现)的模式。

基于位置语义信息的区域模式

所谓的位置语义指每个经纬度坐标表示的地理位置所代表的地方，比如商业区、生活区等。比如例 3 中，商业区域内的服务请求模式往往与游览区域内的服务请求模式有很大的不同。在实际生活中，位

置语义不同的空间区域内数据密度也往往是不同的，因此传统挖掘方法在密度多样化数据集中有明显的局限性。论文[10]探索出一个基于 k 近邻图的层次式区域同位模式挖掘，用 k 近邻图替代距离阈值，层次式地为各个空间区域构建邻域关系图。

基于特征语义信息的区域模式

不同的特征会有不同的影响距离，比如例子中，用户请求餐厅服务是一般指就近用餐，而出行服务的影响范围则会更大。生物学中更是如此，不同物种的领地范围存在很大区别的，比如，老虎和野猪在食物链上有紧密的联系，但老虎的领地范围(一般在 100~400 平方公里)远大于野猪的邻地范围(一般 8~12 平方公里)。论文[11]认为要求用户只用单一的阈值参数进行挖掘在一些案例中是无法得到真的频繁模式。因此论文提出采用统计检验的方法，对一个候选模式在所有可能的距离上做显著性检验，挖掘统计意义上真的频繁模式及其对应的距离阈值。最大的特点是：挖掘结果是不同的距离下的频繁模式，挖掘的正确性很高，但效率较低，适合用于对正确率要求较大的应用中。

文献[8] [9] [10] [11]从不同的角度对区域模式进行挖掘研究，综合考虑区域信息，而不局限在给定一个距离阈值的邻近的区域里。

3.3. 压缩模式

空间并置模式挖掘研究的一个重要的挑战在于发现的模式数量巨大。虽然使用距离阈值能够调节控制挖掘的模式数量，但是效果有限，因为如果距离阈值取得过高，得到的只会是一些常识性的模式，而距离阈值取得过低的话，又很可能引起挖掘的模式数量爆炸。解决这个问题的关键思想是找到一种有效的方法来压缩表达并置模式，例如，找到一个高质量的模式可以精确而简洁地概括一组完整的模式。现有的文献中已经探索了 3 种类型的压缩模式：极大模式[12] [13] [14]、闭模式[15] [16]、代表模式[17] [18]。

定义 6 极大并置模式。如果并置模式 F 是频繁的，并且不存在 F 的超集 F'，使得 $F \subset F'$，且 F' 是频繁的，则称 F 是极大并置模式。

定义 7 闭模式。如果并置模式 F 是频繁的，并且不存在 F 的超集 F'，使得 $F \subset F'$，且 $PI(F) = PI(F')$，则称 F 是闭模式。

定义 8 代表模式，ε 覆盖。如果 $F \subset F'$，则 F 可以被 F' 表达。把 F' 能表达的所有模式的集合称为簇，而 F'称为簇的代表模式。

如果有并置模式 F 和 F'，而且有 $F \subset F'$，且模式之间的距离 $D(F,F') \le \varepsilon$，则称 F' ε 覆盖 F。将 F' ε 覆盖的所有并置模式的集合称为 ε 簇。ε 覆盖的定义可用来度量簇的紧密性。

例 4 空间数据库中特征 A、B、C 各有 4 个实例，观察表 1 中的 4 个频繁模式。

如果采用极大并置模式对例子中的模式进行压缩，则输出模式 F_4。如果使用闭模式的话，会返回结果$\{F_1,F_3,F_4\}$。如果 ε 取 0.7，则由于 $D(F_1,F_4) = 2/3$，$D(F_2,F_4) = 0$，$D(F_3,F_4) = 3/4$，代表模式输出为$\{F_3,F_4\}$，如果 ε 取 0.8，输出为$\{F_4\}$。从例子中可以很明显的看出，在 3 类压缩模式中：极大并置模式最有效的控

Table 1. Prevalent patterns
表 1. 频繁模式

编号	模式	对应的表实例	参与度
F_1	$\{A,B\}$	A_1B_1, A_2B_2, A_3B_3	3/4
F_2	$\{A,C\}$	A_1C_1	1/4
F_3	$\{B,C\}$	B_1C_1, B_2C_2, B_3C_3, B_4C_4	1
F_4	$\{A,B,C\}$	$A_1B_1C_1$	1/4

制了挖掘产生的频繁模式的数量，但不能保留参与度信息。闭模式很好的保留了参与度信息但是压缩效果最弱。代表模式能确定被压缩的模式参与度信息的范围区间，并且通过调整 ε 的值改变输出的模式数量。当 ε 取 1 时，代表模式退化为极大模式，当 ε 取 0 时，代表模式退化为闭模式。

以上的这些论文均对巨大数量的挖掘结果进行压缩，表达成用户感兴趣而且简洁的形式，而 top-*k* 模式[19]则是一种简单直接的减少挖掘返回的模式数量的策略。论文[19]在不确定数据集上输出前 *k* 个最有可能频繁的并置模式。使用输出前 *k* 个模式的策略，可以降低对用户的要求，就算用户没有相关的领域知识，不能设置合理的阈值，也能挖掘到感兴趣的模式。

3.4. 近似模式

对实时性要求较高、正确率要求较低的系统中，比如，实时系统、流系统，近似模式是比较好的选择。在允许一定误差的前提下，文献[20]提出了网格微分算法，在网格基础上进一步细分，提高挖掘结果的准确率，在随机分布的实例集上，得到准确率高达 0.95 的近似模式，算法牺牲了准确性但效率上有明显提高，实验表明，可以实时的处理 45 万数量级的数据集。论文[21]提出了一个适用于实时系统的多项式时间复杂度的近似模式挖掘算法。所谓的近似模式指牺牲了一部分准确率以取得效率上的大幅提高，使得挖掘算法的伸缩性提高。

3.5. 稀有模式

例 5 在珠宝首饰销售中，钻石表的销售是稀有的。然而，涉及钻石表销售的记录可能是令人感兴趣的。

例子中描述了一种特征，相对于数据库中的其他特征，这种特征的实例个数特别地少，称这种特征为稀有特征。我们把带稀有特征的并置模式称之为稀有模式。如果采用最小参与度这一衡量标准，会丢失一些稀有模式，为了解决这一问题，文献[22]提出最大参与率的概念，将参与度定义为并置模式中所有特征的参与率的最大值。文献[22]还用理论和实验证明了用最大参与率能识别和度量所有稀有模式。文献[23]在文献[22]的基础上进一步地讨论了这个问题，文献[23]使用加权参与率的方法解决了将伪稀有模式挖掘出来的问题。

3.6. 非频繁模式

非频繁模式是与频繁模式相对而言的，它很自然的定义如下：

定义 9 非频繁模式是一个空间特征的集合 F，其流行程度的度量值 prevalence(F)小于阈值 *min_prev*，即 prevalence(*F*) < *min_prev*。

值得注意的是，按照频繁模式挖掘讨论的，实例的出现比不出现重要，因此，表实例很少的模式(即非频繁模式)不是令人感兴趣的。论文[24]注意到，在实例分布非均匀的情况下，存在一些特别的非频繁模式：尽管它们在全局上非频繁，但其表实例在一些不规则区域内频繁地关联在一起。论文提出自适应的聚类方法来挖掘这类特别的非频繁模式。

例 6 云南省“三江并流”保护区的植被丰富，研究发现，植物“沧江新樟”和“长苞冷杉”常一起出现，“长苞冷杉”和“虫草”也常一起出现，但“长苞冷杉”、“虫草”、“沧江新樟”却不经常出现在一起。

例子中{“长苞冷杉”,“虫草”,“沧江新樟”}是一个非频繁模式，但研究发现，这 3 中植物有一种特别的性质：当“长苞冷杉”和“虫草”同时出现时，基本上不会出现“沧江新樟”。我们称这类模式为负模式。负模式指的是达不到给定阈值而相互之间又有很强负相关性的特征集。非频繁模式和负模式是两个密切相关的概念，它们之间存在一定的共性，如图 3 所示[1]。挖掘负模式的难点有：1) 空间计算十分耗时；

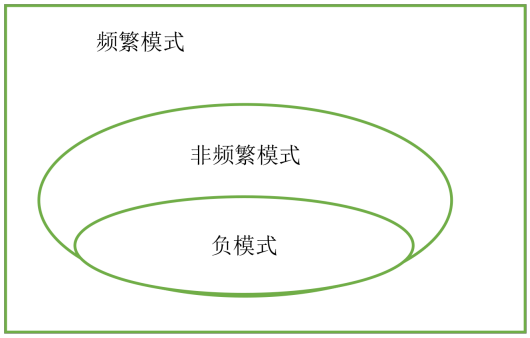

Figure 3. The relationship between non-frequent and negative patterns
图 3. 非频繁模式和负模式的关系

2) 已有算法不能重用；3)生成负模式十分耗时[25]。论文[26] [27]挖掘正、负模式，讨论了算法框架，并应用不同的数据结构、剪枝等策略提高算法效率，这是最早研究负模式的论文。文献[28]则讨论了基于正负并置模式挖掘的城市规划应用，从挖掘结果可以看出同时挖掘正负模式可以更好的解决一般并置模式挖掘在城市规划分析中存在的不足。

3.7. 不确定模式

不确定模式针对的是在数据不确定时挖掘得到的模式。不确定模式挖掘首先要解决的是采用什么样的数据模型来描述空间不确定数据。不确定模式挖掘研究近年来受到越来越多的关注。

文献[29]采用的是 GIS 数据模型，用空间数据的误差分布代替空间数据中的精确点。采用连续函数来描述邻近关系，充分考虑到一个点在其邻居中的影响。

文献[30] [31]观察到实际应用中，实例对模糊特征存在一个大于零的隶属度，因此采用模糊集来描述不确定数据，称这样的数据集为模糊数据，并给出了挖掘的不确定模式的基本定义及合理性证明。

文献[32] [33]采用的是区间数模型。文献[32]从区间数表示的空间数据集上挖掘不确定模式。针对现实应用中精确概率值难以获取的情况，文献[33]使用概率区间描述空间数据集。

文献[34]研究了在位置不确定的数据集上挖掘不确定模式，其空间实例的位置的不确定性由概率密度函数来描述。文献[21]和[35]采用了一个应用最广泛的数据模型-可能世界模型为数据建模，分别挖掘期望上的不确定模式和概率上的不确定模式。

4. 挖掘方法

本文列举了两大类挖掘方法：产生候选方法和模式增长方法。□

产生候选方法

Apriori-like 算法最常用的算法，它采用逐层搜索迭代挖掘模式，其中 $k-1$ 阶频繁模式用于探索 k 阶频繁模式。首先，频繁 1 阶模式的集合记为 L_1，然后，使用 L_1 找出频繁 2 阶模式的集合 L_2，使用 L_2 找出频繁 3 阶模式的集合 L_3，以此类推，直到不能再找到频繁 k 阶模式。huangyan 在论文[3]中提出的基于完全连接(join-based)的算法就是以 Apriori-like 的形式，基于 $k-1$ 阶频繁模式产生 k 阶候选模式，基于 $k-1$ 阶表实例连接产生 k 阶表实例。由于最小参与率概念自然的具有向下闭合性质，因此论文[3]利用向下闭合性质将算法分为“连接”和“剪枝”两步，有效的压缩了搜索空间，提高频繁模式逐层产生的效率。

模式增长方法

与 Apriori-like 算法最大的不同之处在于不产生候选。使用紧凑的前缀树结构组织数据，并直接从该

结构中提取频繁模式而不产生候选。

与传统的关联规则挖掘的 FP-Growth 研究相对应，文献[36]提出了一种基于投影的并置模式挖掘算法，该算法首先为每个空间特征创建一个事务型数据库，将空间数据集转换为传统事务数据集，然后在事务数据集上用基于 FP 树的方法求出最大频繁模式，通过最大频繁模式组合求出所有的频繁模式。

文献[37]提出 CPI-tree 算法。CPI-tree 以树结构物化空间实例间的邻近关系，理论证明了 CPI-tree 的无冗余性和完备性。所有的 co-location 表实例能够通过 CPI-tree 快速生成。采用了与模式增长相类似的分治和递归技术，适于在频繁并置模式数量较多的情况。

5. 应用

并置模式挖掘的应用可在两个方面开展：基于位置语义的应用和基于数据特征的应用。所谓的位置语义指每个经纬度坐标表示的地理位置所代表的地方，比如生活区、商业区等。基于数据特征的应用主要在邻近区域挖掘数据的相关性。以下列举了一些应用：

5.1. 基于位置语义的应用

论文[38]中使用了中国武汉江汉区的犯罪记录和土地使用数据，研究了犯罪和不同功能区之间的关联，实验得到了许多有启发的模式，比如电动车偷盗经常发生在商店附近，因为人们去商店买东西时，往往计划着马上回到电动车旁边，就没锁电动车了。而在酒店、医院附近则比较少出现电动车被盗的案件，因为人们比较少骑着电动车去这些地方的。这些有趣的模式可以帮助警察在不同的功能区有针对性的进行宣传和部署以减少犯罪。

文献[28]从预处理后的昆明市矢量地图中挖掘正负规则。论文中挖掘出许多有意思的规则，部分规则如表 2 所示。

从表 2 中可以看出，昆明市 52%的公司和小区附近都有购物商城和医院药店，46%的学校、医院药店附近没有银行等。对结果进行综合分析，可以为基础设施是否合理、资源是否被浪费、是否会形成交通压力等方面提供证据，以期给够对城区建设的改造优化给出一些建议。从应用中可以看出，并置模式挖掘能提供更为全面的城市规划决策依据，对城市的发展有十分重要的意义。

选址问题是一个商业机构都要面临的重大决策问题之一，它受到各种因素制约，文献[39]利用并置模式的这种特征间“共存”关系，提出了一种基于并置模式的客观评价的地址选择算法，算法基于本体描述空间数据的分类信息，并在本体的指导下对用户感兴趣的兴趣点进行关键挖掘。

5.2. 基于数据特征的应用

文献[2]提出 k 邻近类集在移动服务中，寻找在空间邻近区域上频繁出现的服务，比如，可能发现票务服务和时刻表服务频繁的相互联系，则基于位置的服务台提供商可也为找过时刻表的顾客推销票务服务。

Table 2. The positive and negative rules of mining
表 2. 挖掘的正负规则

编号	规则	参与度	置信度
1	CO (公司和小区)→SP (购物商城)∧ME (医院药店)	65%	52%
2	GO (政府机构)∧CO (公司和小区)→SP (购物商城)	69%	60%
3	CO (公司和小区)∧SC (学校)→PA (休闲娱乐)	63%	54%
4	SC (学校)∧ME (医院药店)→BA (银行)	59%	46%

文献[40]提出了基于 Web 的可视化空间并置模式挖掘原型系统 SCPMiner，实际数据集采用云南“三江并流”自然保护区的植被分布数据。挖掘结果对植物生长特性及植物分类等研究具有重要意义。文献[41]针对深圳市制造业公司专题数据进行数据预处理和并置模式挖掘研究，并依据经济学中导致产业集聚的三种机制对挖掘结果进行定性分析，该工作为进一步的专题数据研究提供了一定的理论基础和实验支撑。

6. 总结

本文简单介绍了几种常见的并置模式类型及其应用实例，包括频繁模式，区域模式，压缩模式和近似模式等。其中，频繁模式是应用研究中最经典最常见的类型模式，区域模式在应用中需要根据地理位置或位置功能进行划分挖掘的模式，而不确定模式则是针对不确定数据集的模式。根据是否产生候选模式，挖掘算法可以分为产生候选方法和模式增长方法两大类。本文还简单介绍了空间并置模式的应用研究现状，并把应用分为基于位置语义的应用和基于数据特征的应用两大类。总的来说，并置模式以不断扩展的应用为驱动，在理论研究和实践上取得一系列的成果。同时，并置模式挖掘研究也受到了一系列新的挑战，并有少量的文献针对这些挑战展开研究：

1) 一些新的空间应用数据的出现带来的挑战，这些新的应用数据集对模式类型、挖掘效率等方面提出了新的要求。比如，文献[42]中关注的带稀疏位置信息的社交网络数据，文献[43]中对带时序的空间数据集展开研究，文献[44]中研究动态数据集上的模式挖掘。

2) 空间并置模式关心空间中的共存现象，但当数据集增大时，数据间的关系往往爆炸增加，如何为海量的应用数据集设计更好的挖掘算法也是一个新的有趣的挑战。比如，文献[5]首次为海量数据集设计了并行挖掘方法。

并置模式挖掘方法的挖掘结果会包含许多无用的或用户不感兴趣的知识，如何挖掘更有效用的模式是并置模式研究的挑战之一。比如，文献[45]从领域驱动的角度提出有趣性度量指标，用以更有效的挖掘用户感兴趣的模式。

资助信息

本文得到云南省高校科技创新团队支持计划资助，云南省教育厅科研基金(No. 2016ZZX304)资助。

参考文献

[1] 王丽珍, 陈红梅. 空间模式挖掘理论与方法[M]. 北京: 科学出版社, 2014.

[2] Morimoto, Y. (2001) Mining Frequent Neighboring Class Sets in Spatial Databases. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 26-29 August 2001, San Francisco, 353-358. https://doi.org/10.1145/502512.502564

[3] Huang, Y., Shekhar, S. and Xiong, H. (2004) Discovering Co-Location Patterns from Spatial Data Sets: A General Approach. *IEEE Transactions on Knowledge & Data Engineering*, **16**, 1472-1485. https://doi.org/10.1109/TKDE.2004.90

[4] Yao, X., Chen, L., Wen, C., *et al.* (2018) A Spatial Co-Location Mining Algorithm That Includes Adaptive Proximity Improvements and Distant Instance References. *International Journal of Geographical Information Science*, **3**, 1-26. https://doi.org/10.1080/13658816.2018.1431839

[5] Jin, S.Y., Boulware, D. and Kimmey, D. (2014) A Parallel Spatial Co-Location Mining Algorithm Based on MapReduce. *IEEE International Congress on Big Data*, **2014**, 25-31.

[6] 吴萍萍, 王丽珍, 周永恒. 带模糊属性的空间 Co-Location 模式挖掘研究[J]. 计算机科学与探索, 2013, 7(4): 348-358.

[7] 江万国, 王丽珍, 方圆, 等. 领域驱动的高效用 co-location 模式挖掘方法[J]. 计算机应用, 2017, 37(2): 322-328.

[8] Dai, B.R. and Lin, M.Y. (2011) Efficiently Mining Dynamic Zonal Co-Location Patterns Based on Maximal Co-Locations. *IEEE International Conference on Data Mining Workshops*, 11-11 December 2011, Vancouver, BC, 861-868. https://doi.org/10.1109/ICDMW.2011.73

[9] Zhao, J., Wang, L., Bao, X., *et al.* (2016) Mining Co-Location Patterns with Spatial Distribution Characteristics. *IEEE International Conference on Computer, Information and Telecommunication Systems*, 6-8 July 2016, Kunming, 1-5. https://doi.org/10.1109/CITS.2016.7546446

[10] Qian, F., Chiew, K., He, Q., *et al.* (2013) Discovery of Regional Co-Location Patterns with k-Nearest Neighbor Graph. *Advances in Knowledge Discovery and Data Mining*, **2013**, 174-186. https://doi.org/10.1007/978-3-642-37453-1_15

[11] Barua, S. (2014) Mining Statistically Sound Co-Location Patterns at Multiple Distances. *International Conference on Scientific & Statistical Database Management*, Aalborg, 30 June 2014, 1-12.

[12] Wang, L., Zhou, L., Lu, J., *et al*. (2009) An Order-Clique-Based Approach for Mining Maximal Co-Locations. *Information Sciences*, **179**, 3370-3382. https://doi.org/10.1016/j.ins.2009.05.023

[13] Jin, S.Y. and Bow, M. (2011) Mining Maximal Co-Located Event Sets. *Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, Shenzhen, 24-27 May 2011, 351-362.

[14] 胡新, 王丽珍, 周丽华, 等. 空间极大 co-location 模式挖掘研究[J]. 计算机科学与探索, 2014, 8(2): 150-160.

[15] Jin, S.Y. and Bow, M. (2011) Mining Top-k Closed Co-Location Patterns. *IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services*, Fuzhou, 29 June-1 July 2011, 100-105.

[16] Wang, L., Bao, X., Chen, H., *et al*. (2018) Effective Lossless Condensed Representation and Discovery of Spatial Co-Location Patterns. *Information Sciences*, **436**, 197-213. https://doi.org/10.1016/j.ins.2018.01.011

[17] Liu, B., Chen, L., Liu, C., *et al*. (2015) RCP Mining: Towards the Summarization of Spatial Co-Location Patterns. In: *Advances in Spatial and Temporal Databases*, Springer International Publishing, Berlin, 451-469.

[18] Wang, L., Bao, X. and Zhou, L. (2018) Redundancy Reduction for Prevalent Co-Location Patterns. *IEEE Transactions on Knowledge & Data Engineering*, **30**, 142-155. https://doi.org/10.1109/TKDE.2017.2759110

[19] Wang, L., Han, J., Chen, H., *et al*. (2016) Top-k Probabilistic Prevalent Co-Location Mining in Spatially Uncertain Data Sets. *Frontiers of Computer Science*, **10**, 1-16. https://doi.org/10.1007/s11704-015-4196-9

[20] 姚华传, 王丽珍, 陈红梅, 等. 面向海量数据的空间 co-location 模式挖掘新算法[J]. 计算机科学与探索, 2015, 9(1): 24-35.

[21] Wang, L., Wu, P. and Chen, H. (2013) Finding Probabilistic Prevalent Co-Locations in Spatially Uncertain Data Sets. *IEEE Transactions on Knowledge & Data Engineering*, **25**, 790-804. https://doi.org/10.1109/TKDE.2011.256

[22] Huang, Y., Pei, J. and Xiong, H. (2006) Mining Co-Location Patterns with Rare Events from Spatial Data Sets. *Geoinformatica*, **10**, 239-260. https://doi.org/10.1007/s10707-006-9827-8

[23] 冯岭, 王丽珍, 高世健. 一种带稀有特征的空间 co-location 模式挖掘新方法[J]. 南京大学学报(自然科学), 2012, 48(1): 99-107.

[24] Deng, M., Cai, J., Liu, Q., *et al*. (2017) Multi-Level Method for Discovery of Regional Co-Location Patterns. *International Journal of Geographical Information Science*, **31**, 1846-1870. https://doi.org/10.1080/13658816.2017.1334890

[25] Jiang, Y., Wang, L., Lu, Y., *et al*. (2010) Discovering Both Positive and Negative Co-Location Rules from Spatial Data Sets. *International Conference on Software Engineering and Data Mining*, Chengdu, 23-25 June 2010, 398-403.

[26] 吴凤萍. 正负 co-location 规则挖掘算法研究[D]: [硕士学位论文]. 昆明: 云南大学, 2008.

[27] 胡彩平, 秦小麟. 一种新的正负空间同位规则挖掘算法[J]. 小型微型计算机系统, 2008, 29(1): 80-84.

[28] 赵楠. 基于正负空间 co-location 模式挖掘的城市规划分析[D]: [硕士学位论文]. 昆明: 云南大学, 2012.

[29] Sheng, C., Hsu, W. and Lee, M. (2008) Discovering Spatial Interaction Patterns. Springer, Berlin, 95-109. https://doi.org/10.1007/978-3-540-78568-2_10

[30] 欧阳志平, 王丽珍, 陈红梅. 模糊对象的空间 Co-Location 模式挖掘研究[J]. 计算机学报, 2012(10): 1947-1956.

[31] Ouyang, Z., Wang, L. and Wu, P. (2017) Spatial Co-Location Pattern Discovery from Fuzzy Objects. *International Journal on Artificial Intelligence Tools*, **26**, Article ID: 1750003. https://doi.org/10.1142/S0218213017500038

[32] Wang, L., Chen, H., Zhao, L., *et al*. (2010) Efficiently Mining Co-Location Rules on Interval Data. *International Conference on Advanced Data Mining and Applications*, Chongqing, 19-21 November 2010, 477-488. https://doi.org/10.1007/978-3-642-17316-5_45

[33] Wang, L., Guan, P., Chen, H., *et al*. (2013) Mining Co-Locations from Spatially Uncertain Data with Probability Intervals. In: *WAIM* 2013 *Workshops*, Springer-Verlag, Berlin, LNCS 7901, 301-314.

[34] 陆叶, 王丽珍, 张晓峰. 从不确定数据集中挖掘频繁 Co-Location 模式[J]. 计算机科学与探索, 2009, 3(6): 656-664.

[35] 陆叶, 王丽珍, 陈红梅, 等. 基于可能世界的不确定空间 co-location 模式挖掘研究[J]. 计算机研究与发展, 2010, 47(z1.): 215-221.

[36] Huang, Y., Zhang, L. and Yu, P. (2005) Can We Apply Projection Based Frequent Pattern Mining Paradigm to Spatial Co-Location Mining? *Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, Hanoi, 18-20 May 2005, 719-725.

[37] Wang, L., Bao, Y., Lu, J., *et al*. (2008) A New Join-Less Approach for Co-Location Pattern Mining. *International Conference on Computer and Information Technology*, Sydney, 8-11 July 2008, 197-202.

[38] Yue, H., Zhu, X., Ye, X., *et al*. (2017) The Local Colocation Patterns of Crime and Land-Use Features in Wuhan, China. *International Journal of Geo-Information*, **6**, 307. https://doi.org/10.3390/ijgi6100307

[39] 包旭光, 王丽珍, 陈红梅. 基于 co-location 模式和本体的地址选择算法[J]. 计算机工程与应用, 2017, 53(24): 15-22.

[40] Wang, L., Bao, Y., Lu, J., *et al*. (2009) A Web-Based Visual Spatial Co-Location Patterns' Mining Prototype System. *International Conference on Cyberworlds*, Bradford, 7-11 September 2009, 675-681.

[41] 田晶, 王一恒, 颜芬, 等. 一种网络空间现象同位模式挖掘的新方法[J]. 武汉大学学报(信息科学版), 2015, 40(5): 652-660.

[42] Weiler, M., Schmid, K.A., Mamoulis, N., *et al*. (2015) Geo-Social Co-Location Mining. In: *International ACM Workshop on Managing and Mining Enriched Geo-Spatial Data*, ACM, New York, 19-24.

[43] Wang, L., Wu, P., Fan, G., *et al*. (2013) Extracting Prevalent Co-Location Patterns from Historic Spatial Data. In: *International Conference on Web-Age Information Management*, Springer, Berlin Heidelberg, 287-300.

[44] Lu, J., Wang, L., Fang, Y., *et al*. (2018) Mining Strong Symbiotic Patterns Hidden in Spatial Prevalent Co-Location Patterns. *Knowledge-Based Systems*, **146**, 190-202. https://doi.org/10.1016/j.knosys.2018.02.006

[45] Wang, L., Jiang, W., Chen, H., *et al*. (2017) Efficiently Mining High Utility Co-Location Patterns from Spatial Data Sets with Instance-Specific Utilities. In: *International Conference on Database Systems for Advanced Applications*, Springer, Cham, 458-474.