An Algorithm of Fast Mining Frequent Neighboring Class Set

Gang Fang, Hong Ying, and Jiang Xiong

College of Math. and Computer Science, Chongqing Three Gorges University
Chongqing 404000, P.R. China
cqwzjsjfg@163.com, dcs-yh@263.net, xjcq123@sohu.com

Abstract. For these problems that present frequent neighboring class set mining algorithms have more repeated computing and redundancy neighboring class sets, this paper proposes an algorithm of fast mining frequent neighboring class set, which is suitable for mining frequent neighboring class set of objects in large spatial data. The algorithm uses the approach of going back to create database of neighboring class set, and uses the approach of generating proper subset of neighboring class set to compute support by descending search, it only need scan once database to extract frequent neighboring class set. The algorithm improves mining efficiency by two ways. One is that it needn’t generate candidate frequent neighboring class set, the other is that it needn’t repeated scan database when computing support. The result of experiment indicates that the algorithm is faster and more efficient than present algorithms when mining frequent neighboring class set in large spatial data.

Keywords: neighboring class set, proper subset, going back, descending search, spatial data mining.

1 Introduction

Mining spatial association rules from Geographic Information Databases are known as discovery spatial co-location pattern as in [1], which are some implicit rules expressing construct and association of spatial objects in Geographic Information Databases, and also expressing hierarchy and correlation of different subsets of spatial association or spatial data in Geographic Information Databases as in [2]. Nowadays, in spatial data mining, there are mainly three kinds of mining spatial association rules as in [3], such as, layer covered based on clustering as in [3], mining based on spatial transaction as in [2, 4, and 5] and mining based on non-spatial transaction as in [3]. The first two methods may be also used to mine frequent neighboring class set, the spatial association as in [4, 5] is quite single, because they only express spatial association among these objects which are all close to objective. However, neighboring class set expresses another spatial association among these objects which are close to each other. MFNCS as in [2] uses the similar method of Apriori to search frequent neighboring class set, and so the algorithm has some repeated computing and superfluous neighboring class set, its efficiency is not efficient. Hence, this paper
proposes an algorithm of fast mining frequent neighboring class set, denoted by FMFNCS, which may avoid repeated computing and scanning database.

2 Definition and Problem Description

Every object in spatial domain constitutes spatial data set, which is expressed as data structure, denoted by <Object Identify, Class Identify and Spatial Location>. Here, identify of different class in spatial data set is denoted by Class Identify, identify of different object instance in the same class is denoted by Object Identify, location coordinate of object is denoted by Spatial Location. We regard an object as an instance of corresponding class, and so spatial data set is made up of these instances of spatial Class Identify. Sets of Class Identify are thought as a class set, denoted by \( C = \{C_1, C_2, \ldots, C_m\} \) means there are \( m \) different classes.

**Definition 1.** Neighboring Class Set, it is a subset of class set in spatial data set, which is expressed as \( \{C_{t_1}, C_{t_2}, \ldots, C_{t_k}\} \) \((t_k \leq m)\) denoted by NCS. Let \( I = \{i_{t_1}, i_{t_2}, \ldots, i_{t_k}\} \) be an instance of neighboring class set denoted by NCS = \( \{C_{t_1}, C_{t_2}, \ldots, C_{t_k}\} \), here, \( i_{t_j} \) is an instance of \( C_{t_j} \) \((j \in 1, 2, \ldots k)\).

Example, let \( \{D, E, G\} \) be a NCS, and \( I = \{D_3, E_1, G_2\} \) is an instance of NCS.

**Definition 2.** Neighboring Class Set Length, its value is equal to the number of class set contained in neighboring class set. If the length of NCS is equal to \( k \), it is denoted by \( k \)-NCS.

**Definition 3.** Right Instance of Neighboring Class Set, let \( I = \{i_{t_1}, i_{t_2}, \ldots, i_{t_k}\} \) be an instance of NCS, if \( \forall i_p \text{ and } i_q \left( i_p, i_q \in I \right), \text{ and } \text{dist} (i_p, i_q) \leq d \), and then we think \( I \) be an right instance of NCS. Here, \( d \) is the minimal distance used by deciding two spatial objects are close to each other, Euclidean distance is expressed as \( \text{dist} (i_p, i_q) \).

**Definition 4.** Neighboring Class Set Support, it is equal to the number of right instance of neighboring class set, which is denoted by support (NCS).

**Definition 5.** Frequent Neighboring Class Set, its support is not less than the minimal support given by user.

As above knowledge, mining frequent neighboring class set is expressed as follows:

Input: (1) Class set is denoted by \( C = \{C_1, C_2, \ldots, C_m\} \), instance set is denoted by \( I = \{i_1, i_2, \ldots, i_n\} \), each \( i_k \) \((i_k \in I)\) is expressed as above mentioned data structure. (2) Minimal distance is denoted by \( d \). (3) Minimal support is denoted by \( s \).

Output: Frequent neighboring class set.

3 The Algorithm of Fast Mining Frequent Neighboring Class Set

3.1 Using the Method of Going Back to Create NCS Database

To find corresponding NCS of right instance in spatial data set, the algorithm uses the method of going back to create NCS database used by a specifically data structure.