Available online at www.sciencedirect.com
ScienceDirect
Information Sciences 177 (2007) 1593–1608
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

Mining spatial association rules in image databases

Anthony J.T. Lee *, Ruey-Wen Hong, Wei-Min Ko, Wen-Kwang Tsao, Hsiu-Hui Lin

Department of Information Management, National Taiwan University, No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan, ROC

Received 25 October 2005; received in revised form 18 September 2006; accepted 23 September 2006

Abstract

In this paper, we propose a novel spatial mining algorithm, called 9DLT-Miner, to mine the spatial association rules from an image database, where every image is represented by the 9DLT representation. The proposed method consists of two phases. First, we find all frequent patterns of length one. Next, we use frequent k-patterns ($k \geqslant 1$) to generate all candidate $(k+1)$-patterns. For each candidate pattern generated, we scan the database to count the pattern's support and check if it is frequent. The steps in the second phase are repeated until no more frequent patterns can be found. Since our proposed algorithm prunes most of impossible candidates, it is more efficient than the Apriori algorithm. The experiment results show that 9DLT-Miner runs 2–5 times faster than the Apriori algorithm.

Keywords: Spatial data mining; Spatial association rule; 9DLT string; Image database

1. Introduction

Many data mining methods have been proposed such as association rule mining [9,14,23,27,39,43], sequential pattern mining [16,26], calling path pattern mining [24], text mining [22,35], temporal data mining [25,28], spatial data mining [12,15,18,21,29,31,33,37,38,45], etc. Because of the advances in information technology, vast numbers of images have accumulated on the Internet, or in entertainment, education, and other multimedia applications. Therefore, how to mine interesting spatial patterns from image databases has attracted more and more attention in recent years. Spatial data mining refers to the extraction of implicit knowledge, spatial relationships, or other interesting patterns stored in spatial databases. It has broad applications, including geographic information systems (GIS) [4,21,29], architectural images [15], medical images [30,34,36], and multimedia information systems [41].

A general data mining method [1,2], called Apriori, was extended towards mining spatial association rules [21]. In addition to the Apriori method, other methods have been used for spatial association mining such as

* Corresponding author.
E-mail addresses: jtlee@ntu.edu.tw (A.J.T. Lee), d90004@im.ntu.edu.tw (R.-W. Hong), r92030@im.ntu.edu.tw (W.-M. Ko), d93725001@ntu.edu.tw (W.-K. Tsao), d8725003@im.ntu.edu.tw (H.-H. Lin).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2006.09.018

statistical cluster analysis [33], spatial clustering [11,20], predicting locations and finding neighboring class sets [29], and co-locations [18,37]. Moreover, Megalooikonomou [30] presented an evaluation framework to evaluate the performance of association mining methods in 3-D medical image databases in which they use Bayesian networks [13,42] to model the associations among the spatial regions of interest in the images. However, most of these approaches focus on discovering the spatial relationships among neighboring data sets.

Several image representation methods have been proposed. In the attributed relational graphs (ARGs) [3,7], the image content is represented by attributed relational graphs in which objects in the image are represented by graph nodes, and the relationships between them are represented by arcs between nodes. Both nodes and arcs are labeled by attributes corresponding to properties of objects and relationships, respectively. Based on the attributed relational graphs, Petrakis et al. [36] proposed a method to perform similarity search in medical image databases. Moreover, one of the most important methods for discriminating images is by using the spatial relations between objects in the images. The spatial representation of an image can be classified into two categories. One is based on positional information, such as the 2D string representation [8,41]. The other is based on directional information, such as the 9DLT representation [5,6]. Kim et al. [19] proposed the SRR representation, which combines both positional and 9DLT's directional information to represent the spatial relations between objects. Huang et al. [17] presented the 9D-SPA representation, which uses directional and topological information to capture the spatial relationships between objects in a symbolic image. Representing an image with the 9D-SPA representation is more powerful than any other representation schemes based on minimum bounding rectangles or the centroids of objects.

Based on the 2D string representation [8], Huang [18] developed a spatial mining method; however, the relationships between the objects are vague. Hsu et al. [15] proposed the viewpoint mining method that uses the distances and orientations between objects to describe the spatial relationships among objects, but, the mining result may be very detailed.

Directional information is one of the most important types of information in an image database, and the 9DLT representation is fundamental in this method. Therefore, in this paper, we propose a novel spatial mining algorithm, called 9DLT-Miner, to mine spatial association rules in an image database, where every image is represented by the 9DLT representation [6]. Our proposed method consists of two phases. In the first phase, we find all frequent patterns of length one. In the second phase, we use the frequent k-patterns ($k \geqslant 1$), whose lengths are equal to k, to generate all candidate $(k+1)$-patterns. For each candidate $(k+1)$-pattern generated, we scan the database to count the pattern's support and check whether it is frequent. The steps in the second phase are repeated until no more frequent patterns can be found.

The contributions of this paper are:

- We take advantage of the 9DLT representation to design a complementary matrix that enumerates all possible candidates. While generating candidate $(k+1)$-patterns from frequent k-patterns using the complementary matrix, we prune most impossible candidate patterns.
- We propose a novel spatial mining algorithm, called 9DLT-Miner, to mine frequent patterns for the images represented by the 9DLT representation. Compared to the Apriori algorithm [1,2], the 9DLT-Miner algorithm avoids generating impossible candidates, and therefore is more efficient in terms of the execution time.

The remainder of this paper is organized as follows. We introduce the problem definitions and preliminary concepts in Section 2. In Section 3, we describe our proposed method in detail. We present the performance evaluations in Section 4. Finally, our conclusions and suggestions for future work are given in Section 5.

2. Problem definition

In this section, we introduce the problem definitions and preliminary concepts.

Definition 1. Let $I = (i_1, i_2, \ldots, i_w)$ be a set of items, each of which is an object in an image. An *itemset* is a subset of I.