Mining for Image Classification Based on Feature Elements

Yu-Jin Zhang
Tsinghua University, Beijing, China

INTRODUCTION

Motivation: Image Classification in Web Search

The growth of the Internet and storage capability not only increasingly makes images a widespread information format on the World Wide Web (WWW), but it also dramatically expands the number of images on WWW and makes the search of required images more complex and time-consuming. To efficiently search images on the WWW, effective image search engines need to be developed.

The classification of images plays an important role both for Web image searching and retrieving, as it is time-consuming for users to browse through the huge amount of data on the Web. Classification has been used to provide access of large image collections in a more efficient manner, because the classification can reduce search space by filtering out images in an unrelated category (Hirata, 2000).

The heterogeneous nature of Web images makes their classification a challenging task. A functional classification scheme should take the contents of images into consideration. The association rule mining, first proposed by Agrawal (1993), is an appropriate tool for pattern detection in knowledge discovery and data mining. Its objective is to extract useful information from very large databases (Renato, 2002). By using rules extracted from images, the content of images can be suitably analyzed, and the information required for image classification can be obtained.

Highlights of the Article

A novel method for image classification based on feature element through association rule mining is presented. The feature elements can capture well the visual meanings of images according to the subjective perception of human beings. In addition, feature elements are discrete entities and are suitable for working with rule-based classification models. Different from traditional image classification methods, the proposed classification approach, based on a feature element, does not compute the distance between two vectors in the feature space. This approach just tries to find associations between the feature elements and class attributes of the image. Techniques for mining the association rules are adapted, and the mined rules are applied to image classifications. Experiments with real images show that the new approach not only reduces the classification errors but also diminishes the time complexity.

The remaining parts of this article are structured as follows:

- **Background:** (1) Feature Elements vs. Feature Vectors; (2) Association Rules and Rule Mining; (3) Classification Based on Association
- **Main Thrust:** (1) Extracting Various Types of Feature Elements; (2) Feature Element Based Image Classification; (3) Database Used in Test; (4) Classification and Comparison Results
- **Direction of Future Research.**
- **Conclusion.**

BACKGROUND

Feature Elements vs. Feature Vectors

Traditionally, feature vectors are used for object identification and classification as well as for content-based image retrieval (CBIR). In object identification and classification, different features representing the characteristics of objects are extracted first. These features mark out an object to a point in the feature space. By detecting this point in the space, the object can be identified or classified. In CBIR, the procedure is similar. Features such as color, texture, and shape are extracted from images and grouped into feature vectors (Zhang, 2003). The similarity among images is measured by distances between corresponding vectors.

However, these feature vectors often are different from the representation and description adapted by human beings. For example, when people look at a colorful image, they hardly figure out its color histogram but rather are concerned about what particular colors are contained in certain components of the image. In fact, these color components play a great role in perception and represent useful visual meanings of images. The pixels belonging to these visual components can be taken to form perceptual primitive units, by which human beings could identify the content of images (Xu, 2001).

The feature elements are defined on the basis of these primitive units. They are discrete quantities, relatively independent of each other, and have obvious intuitive visual senses. In addition, they can be considered as sets of items. Based on feature elements, image classification becomes a process of counting the existence of representative components in images. For this purpose, it is required to find some association rules between the feature elements and the class attributes of image.

Association Rules and Rule Mining

The association rule can be represented by an expression $X \Rightarrow Y$, where X and Y can be any discrete entity. As we discuss image database, X and Y can be some feature elements extracted from images. The meaning of $X \Rightarrow Y$ is: Given an image database D, for each image $I \in D$, $X \Rightarrow Y$ expresses that whenever an image I contains X then I probably will also contain Y. The support of association rule is defined as the probability $p(X \subseteq I, Y \subseteq I)$, and the confidence of association rule is defined as the conditional probability $p(X \subseteq I \mid Y \subseteq I)$. A rule with support bigger than a specified minimum support and with confidence bigger than a specified minimum confidence is considered as a significant association rule.

Since the introduction of the association rule mining by Agrawal (1993), many researches have been conducted to enhance its performance. Most works can be grouped into the following categories:

1. Works for mining of different rules, such as multidimensional rules (Yang, 2001).
2. Works for taking advantage of particular techniques, such as, tree projection (Guralnik, 2004), multiple minimum supports (Tseng, 2001), constraint-based clustering (Tung, 2001), and association (Cohen, 2001).
3. Works for developing fast algorithms, such as algorithm based on anti-skew partitioning (Lin, 1998).
4. Works for discovering a temporal database, such as discovering temporal association rules (Guimaraes, 2000; Li, 2003).

Currently, the association rule mining (Lee, 2003; Harms, 2004) is one of the most popular pattern discovery methods in knowledge discovery and data mining. In contrast to the classification rule mining (Pal, 2003), the purpose of association rule mining is to find all significant rules in the database that satisfy some minimum support and minimum confidence constraints (Hipp, 2000). It is known that rule-based classification models often have difficulty dealing with continuous variables. However, as a feature element is just a discrete entity, association rules can easily be used for treating images represented and described by feature elements. In fact, a decision about whether an image I contains feature element X and/or feature element Y can be properly defined and detected.

Classification Based on Association

Classification based on associations (CBA) is an algorithm for integrating classification and association rule mining (Liu, 1998). Assume that the data set is a normal relational table that consists of N cases described by distinct attributes and classified into several known classes. All the attributes are treated uniformly. For a categorical attribute, all the possible values are mapped to a set of consecutive positive integers. With these mappings, a data case can be treated as a set of (attribute, integer value) pairs plus a class label. Each (attribute, integer value) is called an item. Let D be the data set, I the set of all items in D, and Y the class labels. A class association rule (CAR) is an implication of the form $X \Rightarrow y$, where $X \subseteq I$, and $y \in Y$. A data case $d \in D$ means d contains a subset of items; that is, $X \subseteq d$ and $X \subseteq I$. A rule $X \Rightarrow y$ holds in D with confidence C if C percentages of cases in D that contain X are labeled with class y. The rule $X \Rightarrow y$ has support S in D if the S percentages of cases in D are contained in X and are labeled with class y.

The objective of CBA is to generate the complete set of CARs that satisfy the specified minimum supports and minimum confidence constraints, and to build a classifier from CARs. It is easy to see that if the right-hand-side of the association rules is restricted to the (classification) class attributes, then such rules can be regarded as classification rules to build classifiers.

MAIN THRUST

Extracting Various Types of Feature Elements

Various types of feature elements that put emphasis on different properties will be employed in different applications. The extractions of feature elements can be carried out first by locating the perceptual elements and then by determining their main properties and giving them suitable descriptions. Three typical examples are described in the following.