IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB8930

Chapter XI

Bayesian Data Mining and Knowledge Discovery

Eitel J. M. Lauria
State University of New York, Albany, USA
Universidad del Salvador, Argetina

Giri Kumar Tayi
State University of New York, Albany, USA

ABSTRACT

One of the major problems faced by data-mining technologies is how to deal with uncertainty. The prime characteristic of Bayesian methods is their explicit use of probability for quantifying uncertainty. Bayesian methods provide a practical method to make inferences from data using probability models for values we observe and about which we want to draw some hypotheses. Bayes' Theorem provides the means of calculating the probability of a hypothesis (posterior probability) based on its prior probability, the probability of the observations, and the likelihood that the observational data fits the hypothesis.

The purpose of this chapter is twofold: to provide an overview of the theoretical framework of Bayesian methods and its application to data mining, with special emphasis on statistical modeling and machine-learning techniques; and to illustrate each theoretical concept covered with practical examples. We will cover basic probability concepts, Bayes' Theorem and its implications, Bayesian classification, Bayesian belief networks, and an introduction to simulation techniques.

DATA MINING, CLASSIFICATION AND SUPERVISED LEARNING

There are different approaches to data mining, which can be grouped according to the kind of task pursued and the kind of data under analysis. A broad grouping of data-mining algorithms includes classification, prediction, clustering, association, and sequential pattern recognition.

Data Mining is closely related to machine learning. Imagine a process in which a computer algorithm learns from experience (the training data set) and builds a model that is then used to predict future behavior. Mitchell (1997) defines machine learning as follows: a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. For example, consider a handwriting recognition problem: the task T is to recognize and classify handwritten words and measures; the performance measure P is the percent of words correctly classified; and the experience E is a database of handwritten words with given class values. This is the case of classification: a learning algorithm (known as classifier) takes a set of classified examples from which it is expected to learn a way of classifying unseen examples. Classification is sometimes called supervised learning, because the learning algorithm operates under supervision by being provided with the actual outcome for each of the training examples.

Consider the following example data set based on the records of the passengers of the Titanic[1]. The Titanic dataset gives the values of four categorical attributes for each of the 2,201 people on board the Titanic when it struck an iceberg and sank. The attributes are social class (first class, second class, third class, crew member), age (adult or child), sex, and whether or not the person survived. Table 1 below lists the set of attributes and its values.

In this case, we know the outcome of the whole universe of passengers on the Titanic; therefore, this is good example to test the accuracy of the classification procedure. We can take a percentage of the 2,201 records at random (say, 90%) and use them as the input dataset with which we would train the classification model.

The trained model would then be used to predict whether the remaining 10% of the passengers survived or not, based on each passenger's set of attributes (social class, age, sex). A fragment of the total dataset (24 records) is depicted in Table 2.

The question that remains is how do we actually train the classifier so that it is able to predict with reasonable accuracy the class of each new instance it is fed? There are many different approaches to classification, including traditional multivariate statistical

Table 1: Titanic example data set

ATTRIBUTE	POSSIBLE VALUES
social class	crew, 1st, 2nd, 3rd
age	adult, child
sex	male, female
survived	yes, no