PERGAMON

Expert Systems with Applications 22 (2002) 303–311

Expert Systems with Applications

www.elsevier.com/locate/eswa

Rule induction in data mining: effect of ordinal scales

Helen M. Moshkovich[a], Alexander I. Mechitov[a], David L. Olson[b,*]

[a]*Michael E. Stephens College of Business, University of Montevallo, Montevallo, AL 35115 USA*
[b]*Department of Management, University of Nebraska, Lincoln, NE 68588-0491, USA*

Abstract

Many classification tasks can be viewed as ordinal. Use of numeric information usually provides possibilities for more powerful analysis than ordinal data. On the other hand, ordinal data allows more powerful analysis when compared to nominal data. It is therefore important not to overlook knowledge about ordinal dependencies in data sets used in data mining. This paper investigates data mining support available from ordinal data. The effect of considering ordinal dependencies in the data set on the overall results of constructing decision trees and induction rules is illustrated. The degree of improved prediction of ordinal over nominal data is demonstrated. When data was very representative and consistent, use of ordinal information reduced the number of final rules with a lower error rate. Data treatment alternatives are presented to deal with data sets having greater imperfections. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Data mining; Ordinal data; Classification

1. Introduction

Data mining may be viewed as the extraction of patterns and models from observed data (Berry & Linoff, 1997). The area of data mining is very broad as it incorporates techniques and approaches from different research disciplines. Data mining is usually used to answer two main types of application questions (Edelstein, 1997): (1) *generate predictions* on the basis of available data or (2) *describe behavior* captured in the data. Examples of the first type of tasks include banking (Kiesnoski, 1999), interested in the success of prospective loans; insurance (Goveia, 1999) interested in the probability of fraud, marketing (Peacock, 1998) interested in identifying the best prospects for direct-mail list campaigns. Examples of the second type include finding out which products are sold together, what infections are connected with surgery, in what time ranges which group of customers use a service. In this article we will concentrate on the first type of tasks.

To answer the first type of question, three main approaches are used (Edelstein, 1997): (1) classification, (2) regression, and (3) time-series. These models are differentiated on the basis of what we want to predict. If we want to forecast continuous values of the output attribute, regression analysis is mostly used (time-series if we are concerned with distinctive properties of time). If we have to predict a categorical value for a specific data item (categorical data fits into a small number of discrete categories such as ‘good credit history’ or ‘bad credit history’), we have a classification task to solve. Examples of this type of tasks are medical or technical diagnostics, loans’ evaluation, bankruptcy prediction, etc.

Classification is one of the most popular data mining task. There are many different methods, which may be used to predict the appropriate class for the objects (or situations). Among the most popular one are: logistic regression, discriminant analysis, decision trees, rule induction, case-base reasoning, neural networks, fuzzy sets, and rough sets (Kennedy, Lee, Van Roy, Reed, & Lippman, 1997). Other methods are used as well.

The majority of data mining techniques can deal with different data types. The traditional types of data mentioned in applications are continuous, discrete, and categorical. Among these three categories continuous scales are usually assumed to be numerical, while categorical and discrete data involve variety. Categorical information may be either ordinal (e.g. ‘high’, ‘medium’, ‘low’), or nominal (e.g. ‘blue’, ‘yellow’, ‘red’). Discrete data is an uncertain data type as different models can treat this type of data differently. The majority of data mining models (e.g. regression analysis, neural networks, etc.) will consider discrete data as numeric and apply models suitable for numeric data (Lippmann, 1987). In other cases, discrete data can be treated as categorical, viewing it as numeric codes for nominal data as done in decision trees or rough sets approaches (Quinlan, 1990; Slowinski, 1995). In the latter case, it can be ordinal as well, e.g. if we describe cars, we

* Corresponding author. Tel.: +1-402-472-4521; fax: +1-402-472-5855.
E-mail address: dolson3@unl.edu (David L. Olson).

0957-4174/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0957-4174(02)00018-0

can have an attribute 'number of doors' with possible values of 2, 4, 5, 6. It can be considered numeric (with the more doors we have the better), or '4 doors' can be the most desirable characteristic, while others are less attractive.

Use of numeric information usually provides possibilities for more powerful analysis than ordinal data. On the other hand, ordinal data allows more powerful analysis when compared to nominal data. It is therefore important not to overlook the knowledge about ordinal dependencies in the data sets. Although not as popular in the area of data mining, the qualities of ordinal data were rather thoroughly examined in the area of decision analysis and expert systems (Ben-David, 1992; Larichev & Moshkovich, 1994, 1997; Mechitov, Moshkovich, Olson, & Killingsworth, 1995; Mechitov, Moshkovich, Bradley, & Schellenberger, 1996; Yager, 1981). Implementation of this knowledge may be useful in some classification problems.

The rest of the paper will investigate additional data mining support available from ordinal data. The effect of inclusion of the information on ordinal dependencies in the data set on the overall results of constructing decision trees and induction rules will be illustrated. Possibilities for using ordinal properties to evaluate quality of the training data set will be discussed.

2. Classification task and data mining methodology

In many cases, previous experience in business decisions is coded in the form of a classification. Classification predicts a categorical value for a specific data item and uses a small number of discrete categories (e.g. 'good credit history', 'bad credit history', etc.). Classification is one of the most popular knowledge discovery tasks applied to a variety of fields. Methods from different research areas are devoted to the analysis of such a problem.

Logistic regression is a traditional approach to a classification problem with numeric scales. It has the same assumptions as regression analysis, but allows use of discrete values for classes. Berry and Linoff (1997) comment that for data sets consisting entirely of continuous variables, regression is probably a very good method. Regression analysis identifies an additive function which links case values with the outcome class with the least error. Regression models can be designed to reflect non-linearities, such as the interactions across the variables describing cases or objects. Once the model is obtained it can be used to find corresponding class for new objects.

Artificial neural networks act much the same way, except that they try many different coefficient values in the additive functions to fit the training set of data until they obtain a fit as good as the modeler specifies. Artificial neural network models have the added benefit of considering variable interactions, giving it the ability to estimate training data contingent upon other independent variable values. (This could also be done with regression, but would require a tremendous amount of computational effort.) Artificial neural network models are based on automatic fitting of a model including non-linear combinations of variables. These models are self-adjusting, in that they train on a given set of input data. During the training stage, if the current model correctly predicts the next observation of data, it continues on to the next observation. However, if the current model is incorrect, the model adjusts to more accurately predict the current observation. This may lose the accuracy developed for past learning, so the system iterates through the data until an acceptable level of accuracy is obtained.

Neural networks have been applied to almost all types of data mining applications. They have the feature that the model remains hidden to the analyst, so that while the computer can quickly give its prediction for a variable, the analyst cannot take the model apart to understand why. Therefore, if explanation of results is very important, usually other methods, especially rule induction systems, are used instead of neural networks.

While logistic regression, discriminant analysis, and neural networks are best to use in cases when all attributes are presented with continuous scales, decision trees usually can deal with continuous and categorical data simultaneously (Quinlan, 1990). Decision trees are useful because they allow the decision process to be unveiled from the data (Linoff, 1998). Each branch in the tree represents a decision made on a particular attribute. The algorithm automatically determines which attributes are most important. The method has relative advantage over neural network and genetic algorithms in that a reusable set of rules are provided, thus explaining model conclusions (Michie, 1998). There are quite a number of systems for tree construction. Almost all data mining systems include tools for this type of data analysis. One of the most popular decision tree systems is C4.5 (Quinlan, 1993). It was used in a very large number of applications as well as the basis for comparison of new machine learning algorithms. The system also generates rules, which can be applied in production rule expert systems.

Fuzzy and rough sets approaches try to take into account the uncertainty of data. The assumption in traditional decision trees and rule induction systems is that the distinctions between classes and attribute levels are clear and crisp (e.g. in that somebody is either old, middle-aged, or young, with distinct and well-defined limits). However, someone aged 34 is not all that young, nor is someone aged 36 all that much older than someone 34. Fuzzy logic (Zadeh, 1965) considers degrees of membership in categories. Someone who is 36, for instance, might have a membership function of 0 for the category old, a 0.9 membership function for the category middle-aged, and a 0.3 membership function for the category young. A membership function provides a relative rating between 0 and 1 of membership in a set. As a result, fuzzy analysis provides for each instance a distribution of membership functions for all classes (or can choose a class with the highest membership as the outcome class).

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات