Chapter 11
Correlation Analysis in Classifiers

Vincent Lemaire
France Télécom, France

Carine Hue
GFI Informatique, France

Olivier Bernier
France Télécom, France

ABSTRACT

This chapter presents a new method to analyze the link between the probabilities produced by a classification model and the variation of its input values. The goal is to increase the predictive probability of a given class by exploring the possible values of the input variables taken independently. The proposed method is presented in a general framework, and then detailed for naive Bayesian classifiers. We also demonstrate the importance of "lever variables", variables which can conceivably be acted upon to obtain specific results as represented by class probabilities, and consequently can be the target of specific policies. The application of the proposed method to several data sets shows that such an approach can lead to useful indicators.

INTRODUCTION

Given a database, one common task in data analysis is to find the relationships or correlations between a set of input or explanatory variables and one target variable. This knowledge extraction often goes through the building of a model which represents these relationships (Han & Kamber, 2006). Faced with a classification problem, a probabilist model allows, for all the instances of the database and given the values of the explanatory variables, the estimation of the probabilities of occurrence of each class target.

These probabilities, or scores, can be used to evaluate existing policies and practices in organizations and governments. They are not always directly usable, however, as they do not give any indication of what action can be decided upon to change this evaluation. Consequently, it seems useful to propose a methodology which would, for every instance in the database, (i) identify the importance of the explanatory variables; (ii) identify

DOI: 10.4018/978-1-60566-906-9.ch011

the position of the values of these explanatory variables; and (iii) propose an action in order to change the probability of the desired class. We propose to deal with the third point by exploring the model relationship between each explanatory variable independently from each other and the target variable. The proposed method presented in this chapter is completely automatic.

This chapter is organized as follows: the first section positions the approach in relation to the state of the art; the second section describes the method at first from a generic point of view and then for the naive Bayes classifier. Through three illustrative examples the third section allows a discussion and a progressive interpretation of the obtained results. In each illustrative example different practical details of the proposed method are explored. Finally we shall conclude.

BACKGROUND

Machine learning abounds with methods for supervised analysis in regression and/ or classification. Generally these methods propose algorithms to build a model from a training database made up of a finite number of examples. The output vector gives the predicted probability of the occurrence of each class label. In general, however, this probability of occurrence is not sufficient and an interpretation and analysis of the result in terms of correlations or relationships between input and output variables is needed.

Furthermore, the interpretation of the model is often based on the parameters and the structure of the model. One can cite, for example: geometrical interpretations (Brennan & Seiford, 1987), interpretations based on rules (Thrun, 1995) or fuzzy rules (Benitez, Castro, & Requena, 1997), statistical tests on the coefficient's model (Nakache & Confais, 2003). Such interpretations are often based on averages for several instances, for a given model, or for a given task (regression or classification).

Another approach, called sensitivity analysis, consists in analyzing the model as a black box by varying its input variables. In such "what if" simulations, the structure and the parameters of the model are important only as far as they allow accurate computations of dependant variables using explanatory variables. Such an approach works whatever the model. A large survey of "what if" methods, often used for artificial neural network, are available in (Leray & Gallinari, 1998; Lemaire, Féraud, & Voisine, 2006).

VARIABLE IMPORTANCE

Whatever the method and the model, the goal is often to analyze the behavior of the model in the absence of one input variable, or a set of input variables, and to deduce the importance of the input variables, for all examples. The reader can find a large survey in (Guyon, 2005). The measure of the importance of the input variables allows the selection of a subset of relevant variables for a given problem. This selection increases the robustness of models and simplifies the understanding of the results delivered by the model. The variety of supervised learning methods, coming from the statistical or artificial intelligence communities often implies importance indicators specific to each model (linear regression, artificial neural network...).

Another possibility is to try to study the importance of a variable for a given example and not in average for all the examples. Given a variable and an example, the purpose is to obtain the variable importance only for this example: for additive classifiers see (Poulin et al., 2006), for Probabilistic RBF Classification Network see (Robnik-Sikonja, Likas, Constantinopoulos, & Kononenko, 2009), and for a general methodology see (Lemaire & Féraud, 2008). If the model is restricted to a naive Bayes Classifier, a state of art is presented in (Možina, Demšar, Kattan, & Zupan, 2004; Robnik-Sikonja & Kononenko,