Naïve Bayes vs. Support Vector Machine: Resilience to Missing Data

Hongbo Shi and Yaqin Liu

School of Information Management, Shanxi University of Finance and Economics
030031 Taiyuan, China
shb710@163.com, liuyaqin2003@126.com

Abstract. The naïve Bayes and support vector machine are the typical generative and discriminative classification models respectively, which are two popular classification approaches. Few studies have been done comparing their resilience to missing data. This paper provides an experimental comparison of the naïve Bayes and support vector machine regarding the resilience to missing data on 24 UCI data sets. The experimental results show that when the missing rate is very small (e.g. 1%), the resilience of the naïve Bayes classifiers to missing data are approximately similar to that of support vector machine classifiers. With the increase of the missing rate, however, the resilience of the naïve Bayes classifiers to missing data are slowly decreased and that of support vector machine classifiers to missing data are rapidly decreased. This demonstrates that the naïve Bayes classifiers have better resilience to missing data than support vector machine classifiers.

Keywords: missing data, the naïve Bayes, SVM, resilience.

1 Introduction

Missing data is a common problem that appears in many real world situations. For example, sensor failures in industrial control processes, omitted entries in databases and non-response in questionnaires [1]. Many scientific, industrial, business and economic decisions are related to the information available at the time of making decisions. In these applications, if we merely ignore the incomplete instance or handle inappropriately missing values, it may lead to biased results in statistical modeling. Therefore, it is essential to research on the problem of missing data.

Many researchers engaged in a serious study of missing data. In order to identify the reason why data are missing, Little and Rubin define three different types of missing data mechanisms [2]: missing completely at random, missing at random and not missing at random. To take advantage of missing data, some common methods handling missing data, which used before learning algorithms, are proposed, for example, case deletion, attribute deletion, mean imputation, multiple imputation and so on. The most representative classification algorithms which are able to deal with missing values were investigated, such as decision trees[3], fuzzy approaches[4], Bayes approaches[5] and support vector machines[6]. In addition, [7] examined the
effect of missing data to different classification algorithms, including two rule inducers, a nearest neighbor method, two decision tree inducers, a naïve Bayes inducer, and linear discriminant analysis. They found that the naïve Bayes method was by far most resilient to missing data.

Generative and discriminative approaches are two different paradigms for solving classification problems, which have different thoughts and frameworks. The discriminative approaches look for an optimal decision function \( f(x) \) or the probability \( p(y|x) \) of \( x \) being the class \( y \) to separate the data from data with the other class label, whereas a generative model often captures the generation process of \( x \) by modeling \( p(x|y) \) and tries to represent the true density of the data. The naïve Bayes classifier and support vector machine (SVM) are the typical generative and discriminative models, respectively. In this paper, we compare the naïve Bayes with support vector machine for examining their reliance to missing data. We select these two particular algorithms for several reasons. First, they are popular with data analysts, machine learning researchers, and statisticians. Second, the naïve Bayes and support vector machine are the generative and discriminative approach, respectively. Third, they often are applied to handle higher dimension data, for instance, text data.

2 Naïve Bayes Classifier vs. Support Vector Machine Classifier

2.1 Naive Bayes Classifier

The naïve Bayes classifier is a typical generative classifier, which can be regarded as a special case of Bayesian network classifiers [8]. In general, Bayesian network classifier models first the joint distribution \( p(x,y) \) of the measured attributes \( x \) and the class labels \( y \) factorized in the form \( p(x|y)p(y) \), and then learns the parameters of the model through maximization of the likelihood given by \( p(x|y)p(y) \). Due to there is a fundamental assumption that the attributes are conditionally independent given a target class, the naïve Bayes classifier in fact learns the parameters of the model through maximization of the likelihood given by \( p(y)[\prod_j p(x_j|y)] \).

Since the naïve Bayes classifiers optimize the model over the whole dimensionality, and are capable of learning even in the presence of some missing values. Furthermore, the naïve Bayes classifier is a stable, and its classification result is not significant changed due to noises or corrupted data.

2.2 SVM Classifier

The SVM [9] classifier is a typical discriminative classifier. Different from generative classifier, it mainly focuses on how well they can separate the positives from the negatives, and does not try to understand the basic information of the individual classes. The SVM classifier maps first the instance \( x \) in a training set into a high dimensional space via a function \( \Phi \), then computes a decision function of the form \( f(x) = \langle w, \Phi(x) \rangle + b \) by maximizing the distance between the set of points \( \Phi(x) \) to the hyperplane or set of hyperplanes parameterized by \( (w, b) \) while being consistent on the training set.

The SVM classifier builds a single model for all classes and hence it requires simultaneous consideration of all other classes. Moreover, the SVM classifier