Pattern Recognition Letters 80 (2016) 165–171

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Beyond accuracy: Learning selective Bayesian classifiers with minimal test cost ☆

Ganggang Kong [a], Liangxiao Jiang [a,b,*], Chaoqun Li [c,**]

[a] Department of Computer Science, China University of Geosciences, Wuhan 430074, China
[b] Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430074, China
[c] Department of Mathematics, China University of Geosciences, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 7 December 2015
Available online 5 July 2016

Keywords:
Naive Bayes
Test-cost sensitive learning
Test cost
Classification accuracy
Greedy search

ABSTRACT

Some existing test-cost sensitive learning algorithms are about balancing act of the misclassification cost and the total test cost, and the others focus on the balance between the classification accuracy and the total test cost. By far, however, few works reduce the total test cost, yet at the same time maintain the high classification accuracy. In order to achieve this goal, this paper modifies the backward greedy search strategy employed in selective Bayesian classifiers (SBC), which is a state-of-the-art improved naive Bayes algorithm pursuing the high classification accuracy but ignoring the total test cost. We call the resulting model test-cost sensitive naive Bayes (TCSNB). TCSNB conducts a modified backward greedy search strategy to select an optimal attribute subset with the minimal total test cost, yet at the same time maintains the high classification accuracy that characterizes SBC. Extensive empirical study validates its effectiveness and efficiency.

1. Introduction

Existing classification algorithms focus either on improving classification accuracy or on reducing different types of cost such as the misclassification cost and test cost [38]. Focusing on the former, a mass of classification algorithms have been proposed to scale up the classification accuracy. For example, in Bayesian learning, of numerous approaches to improving the classification accuracy of naive Bayes (NB), the attribute selection approach has demonstrated remarkable performance. The attribute selection approach [12,14,19,21,28,33] selects an optimal attribute subset from the whole attribute space and then builds naive Bayes classifiers on the selected attribute subset only. Focusing on the latter, called cost-sensitive learning, it has received increasing attention in recent years and is hailed as one of the top 10 challenging problems in data mining research [42]. In cost-sensitive learning, researchers always pay their attention to the various kinds of cost [38]. Cost-sensitive learning algorithms are always designed to yield classifiers that minimize the total cost instead of the misclassification errors. When building a classifier on training data or performing classification on a new input instance, we often consider the total test cost when the attribute values must be obtained through physical '*tests*' which incur cost themselves.

Test-cost sensitive learning is more practical than simple traditional classification in many applications such as intelligent medical diagnostic systems [24]. As an example, considering the task of medical diagnosis that doctors examine incoming patients' illness, finding a proper tradeoff between the accuracy (the proportion of patients diagnosed correctly) and efficiency (the total cost of measuring attributes/features) during diagnosis is an important part of the problem. Before diagnosing the patient, his certain information, such as the blood test, cardiogram and X-ray test, may not yet be known which may provide different informational values towards improving the classification accuracy, but performing them will incur extra cost.

However, some existing test-cost sensitive learning algorithms are about balancing the act of the two types of cost, namely the misclassification cost and the test cost, to determine which test will be done [3,23,37,40], and the others [1,2,26,27,30,31,35] focus on the balance between the classification accuracy and the minimal test cost directly, and determine which prior '*test*' should be done for attribute values. In this paper, we focus our attention on the task that covers not only reducing the total test cost but also improving the classification accuracy. Inspired by the success of selective Bayesian classifiers (SBC) [21], we modify the backward greedy search method for attribute selection to minimize the

☆ This paper has been recommended for acceptance by Punam Kumar Saha.

* Corresponding author at: Department of Computer Science, China University of Geosciences, Wuhan 430074, China. Tel:./Fax: +86 27 67883716.
** Corresponding author. Tel:./Fax: +86 27 67883716.
E-mail addresses: ljiang@cug.edu.cn (L. Jiang), chqli@cug.edu.cn (C. Li).

http://dx.doi.org/10.1016/j.patrec.2016.06.018
0167-8655/

total test cost, yet at the same time maintain the high classification accuracy that characterizes SBC. We call our proposed algorithm test-cost sensitive naive Bayes, simply denoted by TCSNB. While selecting an optimal attribute subset for naive Bayes, TCSNB gives consideration not only to the total test cost but also to the classification accuracy. The extensive empirical studies on a large suite of UCI datasets validate the effectiveness of the proposed TCSNB in terms of the total test cost and the classification accuracy.

The rest of the paper is organized as follows: In Section 2, some work related to improved naive Bayes and test-cost sensitive learning are introduced. In Section 3, we incorporate a modified backward greedy search for attribute selection into naive Bayes and propose test-cost sensitive naive Bayes (TCSNB). In Section 4, we design two series of experiments to validate the effectiveness of the proposed TCSNB. In Section 5, we draw conclusions.

2. Related work

Because learning an optimal Bayesian network is NP-hard [6], learning naive Bayes (NB) and its variants has attracted much attention from researchers. Given a test instance x, represented by an attribute vector $\langle a_1, a_2, \ldots, a_n\rangle$, NB uses Eq. (1) to classify x.

$$c(x) = \arg\max_{c\in C} P(c)\prod_{i=1}^{n} P(a_i|c), \tag{1}$$

where a_i $(i = 1, 2, \ldots, n)$ is the value of the ith attribute A_i, C is the set of all possible class labels c, and $c(x)$ is the class label of x predicted by NB.

NB is the simplest form of Bayesian networks due to its attribute independence assumption. In order to weaken its attribute independence assumption, many approaches have been presented. The related work can be broadly divided into five main categories: (1) structure extension; (2) attribute weighting; (3) attribute selection; (4) instance weighting; (5) instance selection, also called local learning. In this paper, we would like to focus our attention on the attribute selection approach.

The attribute selection approach selects an attribute subset from the whole attribute space and only uses the selected attribute subset to build naive Bayes. To execute the attribute selection process, many state-of-the-art algorithms are proposed and fall into two broad categories: filters and wrappers. Filters [12] use the general data characteristics to evaluate the selected attribute subset before the learning algorithm is run. For example, [12] proposed a correlation-based filter algorithm. More examples include the rough set-based algorithm [28], the decision tree-based algorithm [33], and so on. According to the reference [20], a filter would not always work well with naive Bayes even if it perfectly selects strongly and weakly relevant features, because in some cases the performance of naive Bayes improves with the removal of relevant features. Different from filters, wrappers [25] use the learning algorithm itself as a black box to evaluate the selected attribute subset. For example, Hall [21] proposed a greedy search-based wrapper algorithm. More examples include the genetic search-based algorithm [19], the random search-based algorithm [14], the forward greedy search-based algorithm [13], and so on. Generally, filters are more efficient while wrappers are more effective. Recently, Chandrashekar and Sahin [4] provided an overview of feature selection methods and gave a generic introduction to filter, wrapper and embedded methods. Embedded methods are to reduce the computation time which is done for reclassifying subsets in wrapper methods.

Cost-sensitive learning concerning how to reduce different types of cost (e.g., the misclassification cost and the total test cost) but ignoring the classification accuracy, can be generally categorized into two main categories [22]: direct cost-sensitive learning and cost-sensitive meta-learning. Examples of the former include ICET [37], cost-sensitive decision tree [9], cost-sensitive iterative Bayes [11], and cost-sensitive selective naive Bayes [13]. Examples of the latter can be further categorized into thresholding, sampling, weighting, and cloning. Thresholding adjusts thresholds based on misclassification cost and posterior probabilities, while traditional classifiers predict the class of a test instance in terms of a default, fixed threshold 0.5. Thresholding includes MetaCost [8], CostSensitiveClassifier [41] and Empirical Thresholding [34] etc. Sampling [5,15,32] modifies the class distribution of training data and then applies cost-insensitive classifiers on the sampled data directly. Weighting [16,36] assigns a weight to each instance based on its misclassification cost. Cloning [17] clones each minority class instance to rebalance the original training data.

As an attractive hot spot of studies in cost-sensitive learning, the test-cost sensitive learning, has attracted increased attention from researchers in recent years. Turney [37] presented an induction algorithm called ICET using a genetic algorithm to build a decision tree to minimize the sum of the misclassification cost and the test cost. Chai et al. [3] proposed a test-cost sensitive naive Bayes algorithm for designing a classifier that minimize the sum of misclassification cost and test cost via sequential test strategies and batch test strategies. Ling et al. [23] utilized a cost-based splitting criterion for attribute selection measure in order to build a decision tree that minimizes the sum of the misclassification cost and the test cost. Weiss et al. [40] proposed the CASH algorithm cost-sensitive feature selection using histograms.

Of the previous works considering the test cost alone [1,2,26,27,30,31,35], Núñez [30] used background knowledge to provide information for economic induction and used EG2 to generate a decision tree with attribute economic optimization. Tan [35] exploited realistic measurement cost of attributes in the process of incremental learning and proposed a learning-from-examples framework that trades off accuracy and efficiency during learning. Besides, based on the rough set theory, Min et al. [26] proposed a test-cost-sensitive attribute reduction algorithm and pointed out that when tests must be undertaken in parallel, attribute reduction is mandatory in dealing with reducing the test cost. Also in face of the test cost constraint problem, Min et al. [27] proposed the backtracking algorithm to solve the problem on small and medium-sized datasets, and developed the heuristic algorithm which is scalable to large datasets, using the covering rough set. Bolón-Canedo et al. [1] proposed a general framework which takes into consideration not only the relevance of the feature but also their associated cost, and applied this framework to two filter methods (simply CFS and mRMR). This approach allows user to reduce the cost without compromising the classification error through the manual input parameter. At the same time, Bolon-Canedo et al. [2] extended the well-known ReliefF method for feature selection, which adopts a manual parameter to adjust the influence of the cost respect to the influence of the relevance. To address the problem about evaluation functions on cost-sensitive heterogeneous, Qian et al. [31] proposed a multi-criteria based evaluation function to find candidate features with the minimal total cost and the same information as the whole feature set.

3. Test-cost sensitive naive Bayes

Our research starts from our comments to the attribute selection approach for naive Bayes. The attribute selection approach selects an attribute subset from the whole attribute space and only uses the selected attribute subset to build naive Bayes. Why does the attribute selection approach work well? Let us see the following example [18]. Assume that the attribute set $A = \{A_1, A_2, A_3, A_4\}$, in which A_2 and A_4 are completely dependent on A_1 and A_3 respectively (that is, $A_1 = A_2$, $A_3 = A_4$), and A_1 and A_2 are completely