An Improved Naive Bayesian Classification Model Based on Attribute Weighting

Xi Yue\textsuperscript{a}, Mengxuan Tang\textsuperscript{b}

Software of Engineering Chengdu University of Information Technology Chengdu, China

Software of Engineering Chengdu University of Information Technology Chengdu, China

\textsuperscript{a}E-Mail: yuexi@cuit.edu.cn

\textsuperscript{b}Email: 1062091655@qq.com

Abstract. Naive Bayesian model has good classification accuracy and efficiency, which makes it show good performance in many fields, especially in data mining and artificial intelligence. However, the traditional Naive Bayesian classification model ignores the attributes' independence, resulting in the reduction of classification accuracy. For this reason, an improved model based on attribute fusion and weighting (AWNBC) is proposed, in which data fusion is realized by Spearman coefficient and weighting is realized by average confidence and ReliefF coefficient. The experiment classifies the selected data in UCI database. The result of experiments show that the improved classification model has good classification accuracy and efficiency.

1. INTRODUCTION

With the deepening study of big data, data mining plays an increasingly important role, and also makes the data have new content. Through the continuous accumulation and mining of a large amount of original data, more valuable information can be obtained. Therefore, in the current situation of data explosion, how to conduct efficient and accurate classification model research is obviously important\cite{1}. Naïve Bayes Classifiers (NBC), one of the ten classical data mining algorithms\cite{2}, has been widely applied in the field of data mining. However, the precondition of NBC is that it is often difficult to realize the independence among attributes.

There are two main reasons why NBC prerequisite is often difficult to achieve: 1. the attributes are not always independent of each other, and their dependencies are more or less among attributes; 2. some data in the data set have different influences on the classification results. In particular, the redundant attributes in the data set have little influence on the classification itself, but greatly reduce the performance of the classification.

In order to alleviate its conditional independence assumption, and improve the efficiency and accuracy of model classification, different researchers have provided different optimization schemes for this problem. So different scenarios of the improved scheme also differ in thousands ways, but to sum up, there are the following ideas: 1. the use of Bayesian networks in the dependencies of the said method, structure and extend the naive Bayes model to weaken the purpose of the original classification model conditional independence assumption, the typical representative is TAN classifier and SP-TAN
classifier[3]; 2. adopt attribute selection or attribute weighting to improve the efficiency and accuracy of NBC. Redundant attributes in the data will not only increase computation, cause waste of computer hardware and software resources and affect the efficiency of classification, but also reduce the accuracy of classification. Therefore, attribute selection is usually used as a method to improve the performance of classifier[4]. However, there are still influence differences in the classification process that cannot distinguish different attributes. The calculation of attribute weight can not only eliminate some redundant attributes in the data set, but also reflect the influence of each attribute on the classification result through the weighted form. Therefore, the optimization scheme proposed in this paper is based on attribute weighting. 3. based on the idea of the first one, it expands the undirected network dependence and transforms the structure between attributes into a dissolvable Markov network [5], so that the optimized classification model can more efficiently distinguish the dependencies between attributes and improve the performance of classification. Overfitting caused by classification can also be avoided by adjusting the threshold.

In fact, the impact of attributes for classification effect is not necessarily the same. Some attributes have little influence on the results of classification, however, they can exert great influence on the efficiency and accuracy of classification[6]. The improved model based on attribute weighted naive Bayes classification is put forward based on the idea of combining grouping and attribute weighted. The model can effectively reflect the correlation among attributes, and significantly improve the performance of classification model.

The main contributions of this paper are as follows: 1. calculate the spearman correlation coefficient between any two different attributes, and eliminate redundant attributes by comparing with the set threshold, so as to reduce the impact on classification. 2. calculate the average confidence and ReliefF coefficient among attributes to realize the weighting of classification attributes, improve the performance of classification, and establish the attribute weighting model.

2. RELATED WORK

2.1. Naive Bayesian model

First, The mathematical basis of NBC model is Bayesian formula in mathematical statistics[7]. After training the data of training set, the classification model is obtained, and the data of test set is input for classification, and the classification results are obtained.

Suppose \( X=\{a_1, a_2, ..., a_m\} \) is a data to be classified, and \( a_i \) is a characteristic attribute of the data to be classified, and there is a category set \( C=\{y_1, y_2, ..., y_n\} \) Calculate \( P(y_1|x), P(y_2|x), ..., P(y_n|x) \) according to the bayesian formula:

\[
P(y_i|x) = \frac{P(x|c)P(c)}{P(x)}
\]

where each attribute is assumed to be conditionally independent, and:

\[
P(x|c) = \prod_{j=1}^{m} P(a_j|c)
\]

bringing Eq. 2 back to Eq. 1 deformation.

\[
P(y_k|x)P(c) = \arg\max_{y_k \in y_n} P(c|x)
\]

and Eq. 2’s calculation need samples in m properties are independent of each other, but it is hard to meet the conditions in the process of the practical application. There is a common link between things. It is unlikely to be truly independent of each other[8]. In order to weaken the link between them, the attributes that are closely related to each other can be converged into a new attribute, and then get a new type of factor weighted formula through the Eq. 2:

\[
P(x|c) = \prod_{j=1}^{m} P(a_j|c)^{w_j}
\]

The main purpose of the paper is to find the fusion method and the weight \( W \).
2.2. Data discretization
When the attribute types of the classified feature attributes are discrete, the frequency of each category in the training set is only needed to be calculated as the classified conditional probability. However, when the attribute data type is continuous, it needs to be converted into discrete data by means of discretization.

The number of continuous type is discretized according to gaussian segmentation, and the information of non-digital type is converted into Numbers before discretization. Gaussian distribution discretization requires the variable to obey a certain probability distribution. Based on this, the key to discretization is to obtain the segmentation point. Data discretization (DD) can be depicted as Algorithm 1.

Algorithm 1: DD(x)

Input: a test instance x.
Output: A instance x that has been discretized.

1. Randomly select a column of attributes A in the dataset.
2. Attribute A is classified as $A_1, A_2, A_3$ according to the training set.
3. Assume that the value of $A_1, A_2, A_3$ comply with the gaussian distribution, calculate the mean values of $\mu_1, \mu_2, \mu_3$ and the variance value of $\sigma_1^2, \sigma_2^2, \sigma_3^2$ to get the probability density function $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
4. Two adjacent sets of data are calculated, and the specific value corresponding to the intersection of the probability density functions of $A_1, A_2$ and $A_3$ is the segmentation point.
5. The intersection of element values in the feature attribute column is classified according to the intersection point, and the same value is used for the same category, that is, the discretization of continuous data.

3. AWNBC

3.1. Laplace calibration
In order to avoid the problems of zero probability and overfitting, this paper adopts a more effective Laplace method to improve the prior probability formula. The calculation of Eq. 2 is as follows:

$$P(a_j|c) = \frac{\sum_{i=1}^{n} \delta(a_{ij}, a_i) \delta(c_i, c) + 1}{\sum_{i=1}^{n} \delta(c_i, c) + N}$$

where $n$ is the number of training instances, $N$ is the number of classes, $c_i$ is the class label of the $i$th training instance, $a_{ij}$ is the $j$th attribute value of the $i$th training instance, and $\delta(\cdot)$ is a binary function, which is one if its two parameter are identical and zero otherwise.

3.2. Attribute fusion
In statistics, Spearman coefficient can be used to quantify the correlation strength between two variables, which can be used to represent the correlation among variables in the form of monotonic function[8].

$$r_R = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

where $D$ is the grade difference of the pair of attributes in the two columns of classification attributes, and $N$ is the number of grades or the total number of data pairs. The calculation process is as follows: firstly, the data of two variables are sorted in reverse order, and the position after sorting is recorded. The value of the position is the rank, and then the difference between them is calculated as $D$, and finally, the correlation coefficient is obtained by substituting into Eq. 6.

the Spearman correlation coefficient $r_R$ of any two column attributes is calculated by Eq. 6, and its value is between -1 and 1. When $r_R$=0, $X$ and $Y$ are not related. When $|r_R|=1$, $X$ and $Y$ are completely correlated. When $|r_R|<1$, changes in $X$ cause partial changes in $Y$. The larger the value of $|r_R|$ the higher the correlation. By calculating the correlation coefficient of attributes, a certain threshold is set. If it is larger than the threshold, the two columns of attributes are combined to achieve the purpose of dimensionality reduction. If less than or equal to the threshold, the attributes are not merged. If more than
one group of data is larger than the threshold, the largest group is selected. After obtaining the attribute group, the two column attributes are merged into new attributes, and the original data set is updated to obtain the new training set.

3.3. Attribute weighting

Attribute weighting is usually regarded as a special form of attribute selection, which can weaken the assumption of conditional independence of attributes[9]. The author of literature[10] proposed to take the average confidence score as an attribute weighted weight, and the calculation method is as follows:

$$w_1 = \frac{AC(i) \times m}{\sum_{i=1}^{m} AC(i)}$$

(7)

where $AC(i)$ is the average confidence of $a_i$. In addition, the ReliefF coefficient can also be used for attribute weighting.

$$w_2 = \frac{Re(i) \times m}{\sum_{i=1}^{m} Re(i)}$$

(8)

where

$$Re(i) = W(A) = W(A) - \sum_{j=1}^{k} \frac{diff(A,R,H)}{(m+k)} + \sum_{C \in class(R)} \left[ \frac{p(C)}{1-p(class(R))} \right] \frac{\sum_{i=1}^{k} diff(A,R,M,H)}{(m+k)}$$

(9)

and $diff(a_{ij}, a_{in})$ is the sample difference between $a_{ij}$ and $a_{in}$ on $a_i$. $k$ is the number of the nearest sample.

In order to comprehensively consider the influence of the average confidence score and the correlation score on the classification results, this paper further integrates the above coefficients to obtain the new weighting coefficient, and its calculation formula is as follows:

$$w_i = \frac{w_1 + w_2}{2}$$

(10)

Now, the detailed learning algorithm for attribute weighted Naive Bayes Classifiers (AWNBC) can be depicted as Algorithm 2.

Algorithm 2: AWINBC (D, x)

Input: a training dataset D, a test instance x.

Output: the class label of the test instance x.

1. Estimate the prior probability of each class $c$.
2. Estimate the Spearman coefficient by Eq. 6 to eliminate redundant attributes
3. Estimate the conditional probability $P(a_j | c)$ of each attribute value $a_j$ given by the class $c$ by Eq. 5.
4. Estimate the coefficient by Eq. 7 and Eq. 8, and obtain the matrix $w$ by Eq. 10.
5. Estimate the class membership probability by Eq. 4
6. Predict the class label of the test instance $x$ using Eq. 3.
7. Return the predicted class label.

Please note that, we assume that all attribute weights fall into the interval [0, 1].

4. EXPERIMENTS AND RESULTS

4.1. Experimental data and platform

The experimental data set consists of three groups, all of which are Teaching Assistant Evaluation, hayes-roth and Balance Scale from UCI Machine Learning Repository. For the convenience of presentation, the Teaching Assistant Evaluation of the data set is called dataset A, the data set hayes-roth is called dataset
B, and the Balance Scale of the data set is called dataset B. Data set A contains 151 objects, and each object contains 5 classification attributes and 1 category attribute, which are divided into 3 categories. Data set B contains 160 objects, and each object contains 5 classification attributes and 1 category attribute, which can be divided into 3 categories. Data set C contains 625 objects, and each object contains 4 classification attributes and 1 category attribute, which are divided into 3 categories.

The experiments were all completed on IntelliJ IDEA platform, which was configured as: Intel I7 8550U 2.0ghz CPU; 8 GB of memory; Win10 operating system; Java programming language.

4.2. Experimental analysis

The advanced model proposed three improved methods and the integrated classification model. In order to record the improvement effect of each method and the classification accuracy of the improved model, the experimental statistics were conducted on the classification results and comprehensive classification results of each improved method, with a total of 5 results in each data set. The experimental results are shown in table 1 - table 3.

**Table 1. Experimental results for dataset A**

| Laplace calibration | Attribute fusion | Attributing weighting | Classification accuracy (%) |
|---------------------|-----------------|----------------------|----------------------------|
| 0                   | 0               | 0                    | 83.846                     |
| 1                   | 0               | 0                    | 84.105                     |
| 0                   | 1               | 0                    | 88.741                     |
| 0                   | 0               | 1                    | 88.079                     |
| 1                   | 1               | 1                    | 90.066                     |

**Table 2. Experimental results for dataset B**

| Laplace calibration | Attribute fusion | Attributing weighting | Classification accuracy (%) |
|---------------------|-----------------|----------------------|----------------------------|
| 0                   | 0               | 0                    | 83.125                     |
| 1                   | 0               | 0                    | 83.750                     |
| 0                   | 1               | 0                    | 88.750                     |
| 0                   | 0               | 1                    | 86.875                     |
| 1                   | 1               | 1                    | 89.375                     |

**Table 3. Experimental results for dataset C**

| Laplace calibration | Attribute fusion | Attributing weighting | Classification accuracy (%) |
|---------------------|-----------------|----------------------|----------------------------|
| 0                   | 0               | 0                    | 79.200                     |
| 1                   | 0               | 0                    | 82.760                     |
| 0                   | 1               | 0                    | 86.400                     |
| 0                   | 0               | 1                    | 86.880                     |
| 1                   | 1               | 1                    | 87.840                     |

From the above experimental results, it can be seen that through three different training sets, each improved method has improved the accuracy. Due to the different characteristics of the data set itself and the different correlations between attributes, it also has different influences on the experimental results.
The experimental classification results of the improved Laplace calibration optimization are shown in FIG 1. It can be seen that there is a positive linear correlation between the classification result and the number of attributes in the Laplace calibration optimization. The more attributes there are, the more obvious the classification effect will be. In data set A and B, the improvement of 0 point optimization was 0.259 and 0.625, respectively, with little improvement. However, in data set C, the range of improvement reached 3.56, and the effect was significantly improved. This is because the more attributes, the greater the probability of 0, and the more obvious the improvement effect on the model.

According to the experimental classification results improved by attribute fusion, it can be seen that attribute fusion significantly improves the classification effect. Compared with Laplace calibration, the correlation between the accuracy and the amount of sample data is positively linear.

The experimental classification results with improved attribute weighting are shown in the figure. It can be seen that attribute weighting also has a good promotion effect. Due to the different characteristics of the attributes of the samples, the effect of the promotion on classification varies.

As can be seen from the figure above, due to the different characteristics of the attributes of the samples, the improvement of classification accuracy rate is not very stable for the single improved point.
classification model. However, compared with the original naive Bayesian classification model, the improved model has a higher classification accuracy and better classification stability.

5. CONCLUSIONS AND FUTURE WORK

This paper proposes an improved classification model based on attribute weighting for the conditional independence assumption of naive Bayesian classification model. The improved model with the method of attribute weighted calculates spearman correlation coefficient to eliminate the redundant attributes and to improve the classification accuracy and efficiency significantly. However, by calculating ReliefF as weighted factor and the average confidence coefficient, it also caused the complexity of the algorithm. The research will continue to solve this problem.

REFERENCES

[1] HUANG Chun-hua, CHEN Zhong-wei, LI Shi-jun, Bayesian Decision Tree Method Applying in Admission Data Mining[J], Computer technology and development, 2016(4) : 114-118.

[2] WANG Hui, ZHANG Fan, LI Yu-jie, Semi-naive bayesian classifier research based on greedy choice algorithm [J], Journal of Northeast Normal University(Natural Science Edition), 2018(2) : 79-82.

[3] FRIENDMAN N, GEIGER D, GOLDSZMIDT M. Bayesian network classifier[J],Machine Learning,1997,29(2/3):131-163.

[4] AHA D W. Tolerating noisy, irrelevant and novel attributes in in-stance-based learning algorithms[J]. International Journal of Man-Machissne Studies,1992,36(2):267-287.

[5] WANG Shuangcheng, GAO Rui, DU Ruijie Integrated regression model of bayesian networks with superparent nodes [J]. Computer journal,2017,40(12): 2748-2761.

[6] CHANGRASHEKAR G, SHAHIN F. A survey on feature selection methods[j]. Computers & Electrical Engi-neering,2014,40(1): 16-28.

[7] GUYON I,ELISSEEFF A. An introducetion to variable and feature selection[J]. Journal of Machine Learning Research, 2003,3(6): 1157-1182.

[8] Guo Yingming, Li Hongli. Research on Weighted Naive Bayesian Classification Algorithm Based on Speelman Coefficients [J]. China Computer & Communication,2018(13):54-56.

[9] TUSNOO E, TZANETAISY G,ONOG N.et al. Audio genre classification using percussive pattern clustering combined with timbral features[C] IEEE International Confere-nce on Multimedia and Expo. Washington, DC:IEEE Computer Society,2009:382-385.

[10] NING Ke, SUN Tongjing, ZHAO Haoqiang. Naive Bayesian Classification Algorithm Based on Attribute Association. Computer Engineering,2018,44(6):18-22.