DOI 10.3233/BME-151476
IOS Press

Fuzzy Naive Bayesian for constructing regulated network with weights

Xi Y. Zhou[a] , Xue W. Tian [b] and Joon S. Lim[c,*]
[a, c] *IT College, Gachon University, Seongnam, South Korea*
[b] *Department of Educational Technology, Teachers College, Qingdao University, QingDao, China*

Abstract. In the data mining field, classification is a very crucial technology, and the Bayesian classifier has been one of the hotspots in classification research area. However, assumptions of Naive Bayesian and Tree Augmented Naive Bayesian (TAN) are unfair to attribute relations. Therefore, this paper proposes a new algorithm named Fuzzy Naive Bayesian (FNB) using neural network with weighted membership function (NEWFM) to extract regulated relations and weights. Then, we can use regulated relations and weights to construct a regulated network. Finally, we will classify the heart and Haberman datasets by the FNB network to compare with experiments of Naive Bayesian and TAN. The experiment results show that the FNB has a higher classification rate than Naive Bayesian and TAN.

Keywords: Naive Bayesian, Tree Augmented Naive Bayesian, Fuzzy Naive Bayesian, fuzzy neural network, weights

1. Introduction

Classification is based on the existing classified data to learn and train a classification function or construct a classification model (classifier). Then, using the model to classify the unknown datasets. Classification for massive data training and testing contains: decision tree, neural networks, Bayesian networks and so on. The Bayesian network is one of the most commonly used algorithms.

However, learning Bayesian network without any restrictions is very time consuming. Restricted Bayesian classifier is based on Bayesian classifier with appropriate restrictions, for example, naive Bayesian classifier is a Bayesian network restriction on each attributes only depending on the class attribute, and Tree Augmented Naive Bayesian classifier is limited by each attributes only depending on one-parent node, besides class attribute [1]. This paper proposes algorithm FNB (Fuzzy Naive Bayesian) which is the similar structure with TAN.

Currently, TAN classifier is recognized as one of the best improved Naive Bayesian classifier [2]. This paper focuses on a directed graph based on Naive Bayesian classifier, a more detailed description on the different methods to construct tree classifier. Finally, we introduce this new method to construct Naive Bayesian classifier named FNB (Fuzzy Naive Bayesian) with weights and no restriction for regulated relations.

[*] Address for correspondence: Joon S. Lim, College, Gachon University, Seongnam, South Korea. Tel.: 82-031-750-5750; Fax: 82-031-750-5662; E-mail: jslim@gachon.ac.kr.

1.1. Naive Bayesian network

Naive Bayes classifier is a single-layer structure tree, the root node of the tree is the class node, which also is the class attribute. All other attributes nodes are dependent on the class attribute nodes, in addition, other attributes are mutual independence [3]. Because of its highly conditional independence assumption, the Naive Bayesian classifier makes entire classifier training, and makes the classification process simplified, and its structure fixed. The problem of strong conditional independence assumption is difficult to solve, although the vast majority of data sets' attributes are correlated to each other, so sometimes the performance is not good.

1.2. Tree Augmented Naive Bayesian network

In TAN classifier, each attribute not only depends on class attributes, but also depends on one other parent attribute. This structure improves the strong conditional independence assumptions in Naive Bayesian classifier [4].

TAN classifier algorithm mainly includes the following four steps:

(1) Input the training set.

(2) Add class node as a parent of every attribute.

(3) Mutual information calculated for each of the attributes in a given class attribute conditions.

(4) Learn the parameters and output the Tan network.

2. Experiment

2.1. Method: Fuzzy Naive Bayesian network

Because in TAN method, the number of the parent node is fixed, and it is not very fair to each attribute as in reality, not every attribute nodes only corresponds to the parent node. To improve the Naive Bayesian and TAN, we use a neural network with weighted membership function (NEWFM) based on fuzzy neural network to extract relation between attributes and its parent attributes. At this stage, we rank the relations based on their weights and pick several largest relations [5]. Finally, we use the BSWFM method to obtain the attribute distributions. Therefore, we named the new algorithm as Fuzzy Naive Bayesian (FNB) algorithm.

FNB classifier algorithm mainly includes four steps below:

(1) Data normalization.

Because we need to use NEWFM method to select features and extract relation and the normalized dataset can obtain a better result, so we firstly do data normalization on each attribute with sigmoid function to make the values on the scale of [0, 1].

$$\alpha_i = \frac{1}{1 + e^{-(a_i - \min(\mathrm{A}_i))/(\max(\mathrm{A}_i) - \min(\mathrm{A}_i))}}, \forall\, i = 1,2,\ldots,s \tag{1}$$

(2) Feature selection.

If the number of attributes is huge, we must spend much time in training data to extract relation between different attributes. Therefore, we must do feature selection to decrease the number of attributes.

During this phase, we use the non-overlap area distribution measurement method (NADM) method in NEWFM to select features [5, 6].

(3) Extract the parent node with NEWFM and calculate the weights.

In this phase, we just input the local classification dataset into NEWFM, which is using each attributes' mean values to separate the original dataset. Also we can use the NADM to extract the relation between different attributes. Then we use Eq. (2) to rank all the weights of different attributes. Finally, we select the largest several weights and corresponding parent nodes. The selected number of weights is equal to the number of attributes. So, FNB has no restriction for each attributes' number of parent nodes.

$$w_{i\leftarrow p} = \frac{h(A_p)}{t} * a(A_i), \forall j = 1,2,\dots,s \tag{2}$$

(4) Classify the dataset with FNB network.

At this stage, we use the BSWFM method to obtain each attributes' distributions. Then, we combine attribute distribution with the weights obtained in Eq. (3) to classify the dataset.

Eq. (3) is to obtain the probability that this sample belongs to C.

$$P(\mathrm{b}_1, \dots, b_n, C) = P(C)\sum_{i=1}^{n} P(b_i|b_p, C) \tag{3}$$

In Eq. (4), $P(b_{\mathrm{i}}|C)$, $P(b_{p1}|C)$ and $P(b_{pn}|C)$ are calculated from bounded sum of weighted fuzzy membership functions (BSWFMs) method [7]. The $P(b_{p1}|C)$ and $P(b_{pn}|C)$ are parent nodes of $P(b_{\mathrm{i}}|C)$.

$$\begin{aligned} P(b_i|b_b, C) = P(b_{\mathrm{i}}|C) &+ \left(P(b_{p1}|C) - P(b_{\mathrm{i}}|C)\right) * \mathtt{weight}_1 \dots \\ &+ \left(P(b_{pn}|C) - P(b_{\mathrm{i}}|C)\right) * \mathtt{weight}_{\mathrm{n}} \end{aligned} \tag{4}$$

In Eq. (5), $c(S)$ is to obtain the sample S which belongs to Class A or Class B. If $P(c)\sum_{i=1}^{n} P(\mathrm{b}_i|b_p, A)$ is bigger than $P(c)\sum_{i=1}^{n} P(\mathrm{b}_i|b_p, B)$, that means sample S belongs to Class A.

$$c(S) = \arg\max_{c\in C} P(c)\sum_{i=1}^{n} P(\mathrm{b}_i|b_p, c) \tag{5}$$

2.2. Dataset

In this experiment, we selected two datasets, both of which were downloaded from the UCI machine learning website. The first dataset describes a heart disease (angina), 270 samples in total [8]. The se

Table 1
Heart's Selected Attributes Information

Attributes	Description
CPT	chest pain type
FBS	fasting blood sugar > 120 mg/dl
MHR	maximum heart rate achieved
EA	Exercise-induced angina
DER	ST depression induced by exercise relative to rest
SPE	the slope of the peak exercise ST segment
NMV	number of major vessels colored by fluoroscopy
TH	thal

Table 2
Haberman Survival's Attributes Information

Attributes	Description
AGE	Age of patient at time of operation
YEAR	Patient's year of operation
NAN	Number of positive axillary nodes detected

cond dataset describes a doctor named Haberman and the survival of patients who had undergone surgery for breast cancer: the patient survived 5 years or longer (225 patients) and the patient died within 5 years (81 patients) [9]. Details about these two datasets are shown in Tables 1 and 2.

Because the number of heart's attributes is so big that it will waste time and efficiency. We use the NADM method to select features. After training the dataset, we finally pick up CPT, FBS, MHR, EA, DER, SPE, NMV and TH.

2.3. Subtraction of FNB network

When we use NEWFM to extract relations and calculate the weights of each attribute, then we can construct the network, like Figure 1. It is very obvious that CPT node has three-parent nodes, except

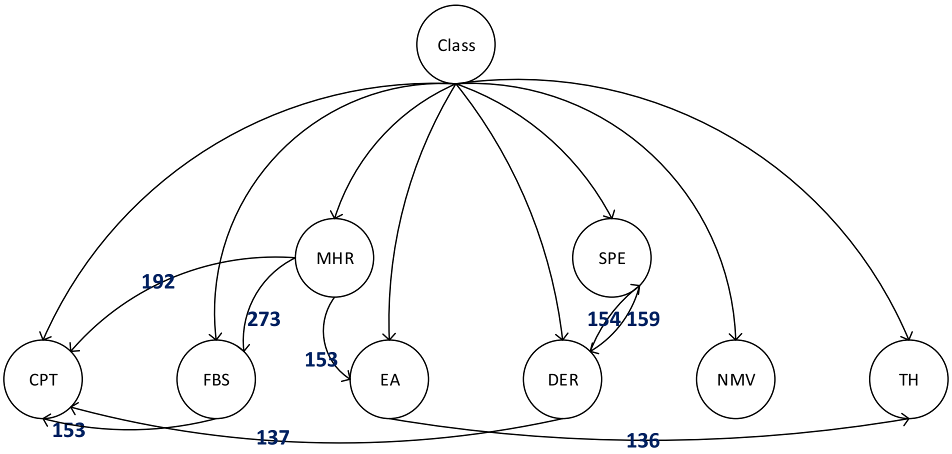

Fig. 1. FNB network of heart dataset.

Table 3
Classification Results

Dataset	NB	TAN	FNB
Heart	72.51	73.52	**83.7**
Haberman Survival	72.9	72.9	**75.5**

the class node, and EA node and NMV node just have the class node. So FNB algorithm is more flexible and fairer to extract the relations.

2.4. Classification results

Table 3 shows the results of these three different algorithms, in which results of Naive Bayesian (NB) and Tree Augmented Naive Bayesian (TAN) are experimented by other papers [5, 6]. The result of Fuzzy Naive Bayesian (FNB) is operated by us. The accuracy rate shows the method has a better performance. It is obvious that the FNB has a higher accuracy than NB and TAN in Heart disease and Haberman Survival datasets.

3. Conclusion

The main reasons why FNB network has a better classification performance have two parts. The first reason is that the FNB has flexible relations between attributes, that means no restriction for number of parent nodes. This is fairer to each attribute than Naive Bayesian and TAN. The other reason is the FNB network has an effective method to calculate the weight memberships. And the weights play an important role in classifying the datasets. All of these make the FNB network have a better classification results.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology. (2012R1A1A2044134).

References

[1] Z. Zheng and G.I. Webb, Lazy learning of Bayesian rules, Machine Leaming **41** (2000), 53–84.

[2] T. Dietterich, Statistical tests for comparing supervised classification learning algorithms, Technical Report, ftp://ftp.cs.orst.edu/pub/tgd/papers/stats.ps.gz, Feb. 2014.

[3] A. Perez, P. Larranaga and I. Inza, Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive Bayes, International Journal of Approximate Reasoning **43** (2006), 1–25.

[4] Eamonn J. Keogh and Michael J. Pazzani, Learning Augmented Bayesian classifiers:A comparison of distribution-based and classification-based approaches, International Journal on Artificial Intelligence Tools **11** (1999), 587–601.

[5] J.S. Lim, D. Wang, Y.-S. Kim and S. Gupta, A neuro-fuzzy approach for diagnosis of antibody deficiency syndrome, Neurocomputing **69** (2006), 969–974.

[6] S.-H. Lee and J.S. Lim, Minimized stock forecasting features selection by automatic feature extraction method, Korean Institute of Intelligent Systems **19** (2009), 206–211.

[7] J.S. Lim, Finding features for real-time premature ventricular contraction detection using a fuzzy neural network system, IEEE Transactions on Neural Networks, 2009, 522–527.

[8] David W. Aha, UCI Repository of Machine Learning, https://archive.ics.uci.edu/ml/datasets/Statlog+(Heart), May 2014.

[9] Tjen-Sien Lim, UCI Repository of Machine Learning, https://archive.ics.uci.edu/ml/datasets/Haberman's+Survival, Jun. 2014.