An Improvement On The Naïve Bayes Classifier

Hong Choon Ong[1][+], May Yee Khoo[1] and Siew Lin Saw[1]

[1] A School of Mathematical Sciences,
Universiti Sains Malaysia, 11800 USM, Penang, Malaysia

Abstract. Naïve Bayes classifier is a simple and efficient classification method. Classification is a process which classifies data into one of several predefined categorical classes. In Naïve Bayes classifier, all the features are assumed as independent of each other within each class. Structural learning among the features thus helps in the classification problem. Structural learning using a hill-climbing algorithm is used in this study to learn the causal relationship and dependencies among the attributes. The accuracy of prediction by structural learning of the features using hill climbing is better than Naïve Bayes classifier using a tenfold cross validation.

Keywords: Bayesian Networks, Classification, Naïve Bayes classifier, Structural Learning

1. Introduction

1.1. Bayesian Networks

Over the last decade, the Bayesian network has become a popular representation for encoding uncertain knowledge in expert systems [1]. Bayesian networks use conditional independence to provide a concise representation for complex probability distributions. It efficiently represents complex probability distributions and thus it is more easily to figure out the structure of a data [2]. However, methods for exact inference in Bayesian networks are often too expensive [3].

Bayesian network is a graphical model that encodes probabilistic relationships among the variables of interest. It is an approach under Machine Learning, a tool of data mining that helps in summarizing large data, which is a probabilistic graphical model that represents a set of random variables and their conditional independencies via directed acyclic graph (DAG) [4].

Bayesian networks can be defined as a probabilistic graphical model G = (N, A) whose nodes N, and arcs A, represent random variables X, and direct correlations between the variables respectively. They use conditional independence to provide a concise representation for complex probability distributions.

1.2. Classification

Classification is one of the most important tasks in data mining. Classification is a process which classifies data into one of several predefined categorical classes. A classifier is basically expected to achieve classification result with high accuracy. There are many methods for classification analysis such as neural networks, decision trees, support vector machines, and Naïve Bayes classifier.

If the attributes are numeric, discretizing the continuous attributes before carrying the analysis is an option. Although, this may cause the accuracy to slightly decrease but it has the advantage of simplicity and is proved that the Bayesian classifier performs much better than other methods [5].

[+] Corresponding author. Tel.: +60164981802; fax: +6046570910.
 E-mail address:hcong@cs.usm.my.

2. Naïve Bayes Classifier

Naïve Bayes Classifier is one of the classification methods in Bayesian networks. Bayesian networks were considered as classifiers when the Naïve-Bayes was created, Lowd and Domingos proposed Naive Bayes model as an alternative to Bayesian networks for general probability estimation tasks [6]. Naïve Bayes is a simple kind of Bayesian network [7].

In Naive Bayes classifier, all the features are independent of each other within each class [8]. Although the independence assumption of naïve Bayes classifier does not hold for most data sets, the independence assumption does perform well for most data sets. However, the assumption of class independence of the features or attributes allows the Naive Bayes classifier to better estimate the parameters required for accurate classification while using less training data than many other classifiers. This makes it particularly effective for datasets containing many predictors or features. Naive-Bayes classifiers are simple and efficient.

A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naïve) independence assumptions. Generally, the Naïve Bayes classifier classifies data in two steps [9]. Firstly, in training, it estimates the parameters of a probability distribution assuming the features are conditionally independent given the class. Then in prediction, it computes the posterior probability of that sample belonging to each class and classifies the test sample accordingly.

We defined the A as the observed attribute values and C as a class of attributes. These probabilities calculated as:

$$P(C = c|A = a), \tag{1}$$

When attributes $A_1, A_2, \ldots, A_n$ are conditionally independent of each other given the class, then the probability is proportional to:

$$P(C = c) \prod_{i=1}^{n} P(A_i = a_i | C = c) \tag{2}$$

Thus, classifiers using (2) are called Naïve Bayes classifiers [10] and are shown in Figure 1.
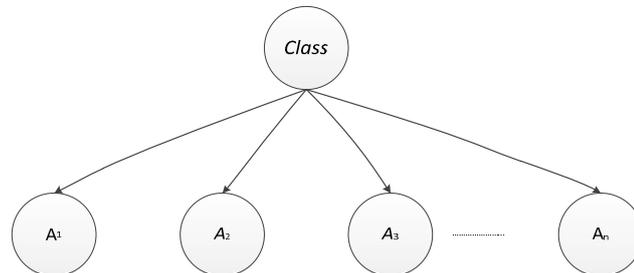

Figure 1: Simple Naïve Bayes Classifier structure. Naïve Bayesian model allows for no arcs between feature nodes.

3. Proposed Method (Structural Learning)

Ge et al. [4] suggested that researchers use structural learning in Bayesian Network to explore the dependencies among the variables. Structural learning using Hill-Climbing (HC) which is a score-based learning algorithm is used in this paper to learn the causal relationship and dependencies among the attributes. It can be determined whether there are any dependencies among the attributes after running the learning algorithm by using the bnlearn package [11]. An arc is constructed to show the relationship between two attributes when there exists some dependencies between them. Figure 2 shows the example of the Naïve Bayes network structure with an added relationship after the structural learning algorithm show a dependency between two attributes A2 and A3.
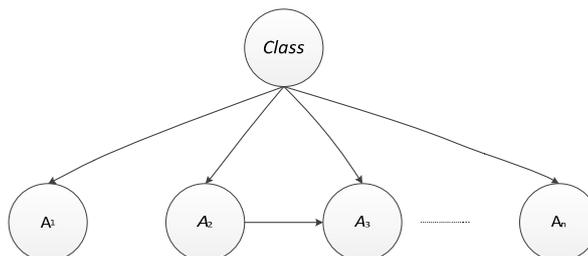

Figure 2: A Naïve Bayes network structure with an added relationship from node A_2 to A_3.

4. Experimental Methodology

Our study uses a score based structural learning algorithm in Bayesian Network to explore the dependencies among the variables. Based on this, we want to show that with the structure learned among the variables in classification data set we have higher classification accuracy than the Naïve Bayes classifier. We test the improvement in the classification of the data that we collected using 10- fold cross validation.

In our study, we used nine sets of data from the UCI Machine Learning Repository [12] and Journal of Statistics Education Data Archive [13]. Table 1 lists the corresponding number of attributes or features, number of instances and number of classes of the nine sets of data.

Table 1: Data sets for classification used

Dataset	No. of Instances	No. of attributes	No. of classes
Breast Cancer	699	10	2
E. coli	336	8	8
Glass Identification	214	10	7
Iris	150	5	3
Liver Disorders	345	7	2
Shuttle Landing	15	6	2
Wine Recognition	178	13	3
Contact Lens	24	4	3
Fish Catch	159	8	7

The first eight sets of data are obtained from UCI Machine Learning Repository. Wisconsin Breast Cancer Database is the database that reflects the chronological grouping of the data. The E. coli database is to classify protein localization sites of E. coli according to attributes. The Glass Identification Database is used to classify the glass into specific classes according to their attributes. The Iris database classify iris into 3 species based on the attributes. The Liver Disorders database determines liver disorders according to their attributes. Space Shuttle Autolanding Domain, generates rules for determining the conditions under which an auto landing would be preferable to manual control of the spacecraft. The Wine Recognition data is the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivation site. For the Database for fitting contact lenses, this data set is used to classify the contact lens to three different groups that is whether the patient should be fitted with hard contact lenses, soft contact lenses, or the patient should not be fitted with contact lenses. The last data set, Fish Catch, which is collected from Journal of Statistics Education-Data Archive, it is to classify 159 fishes that were caught and measured according to attributes and then classified into specific species.

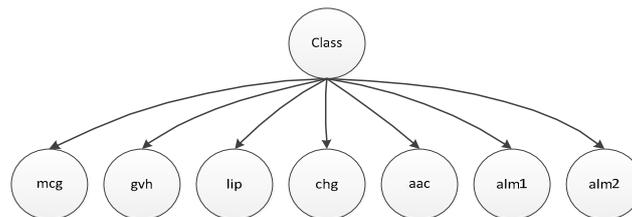An example of Naïve Bayes network structure on the E. coli database is shown in Figure 3.

Figure 3: Naïve Bayes network structure from *E. coli* data set.

The dependencies for E. coli data set are shown in Figure 4 where there are 5 directed arcs among the attributes. Example of dependencies of another data set is shown in Figures 5.
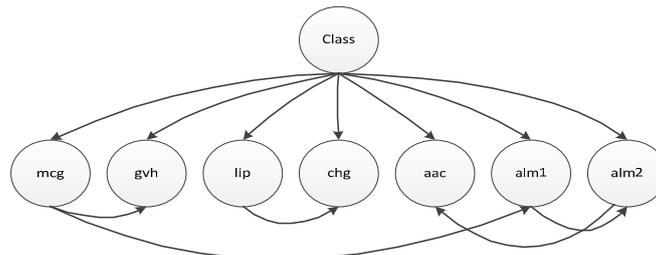

Figure 4: Naïve Bayes network structure with added relationship after structural learning using the hill climbing algorithm on the attributes from *E. coli* data set.

5. Results

The accuracy of Naïve Bayes classifier and the proposed classifier with added relationship in the presence of dependencies among the attributes after structural learning using the hill climbing algorithm are listed in Table 2. 10-fold cross validation is used to make the comparison of accuracy.

By using 10-fold cross-validation method, the data were partitioned into 10 almost equal parts, with nine of ten parts being used as training data and the remaining part used as test data each time. For example, for the first set, parts one to nine were used for the training data and part ten for test data; and for the second set, part one to eight, and ten were used for training data and part nine for test data, and so on. The final predictive accuracy was the average of the ten test samples. These averages are known to be better estimates of the accuracy [15]. A accuracy of successful prediction on the test data of each data set is shown in Table 2.

There is no relationship among the attributes in the Contact Lens data set. Therefore, there is no added relationship in the structure among the attributes nodes. The remaining 8 data sets which have relationship among attributes and have arcs constructed show better accuracy of prediction than Naïve Bayes classifier. The overall results are shown in Table 2 and Figure 8. From the results obtained, when dealing a classification data set using Naïve Bayes classifier given the attributes are not independent it is advisable to apply structural learning on the attributes because it gives a better accuracy of prediction.
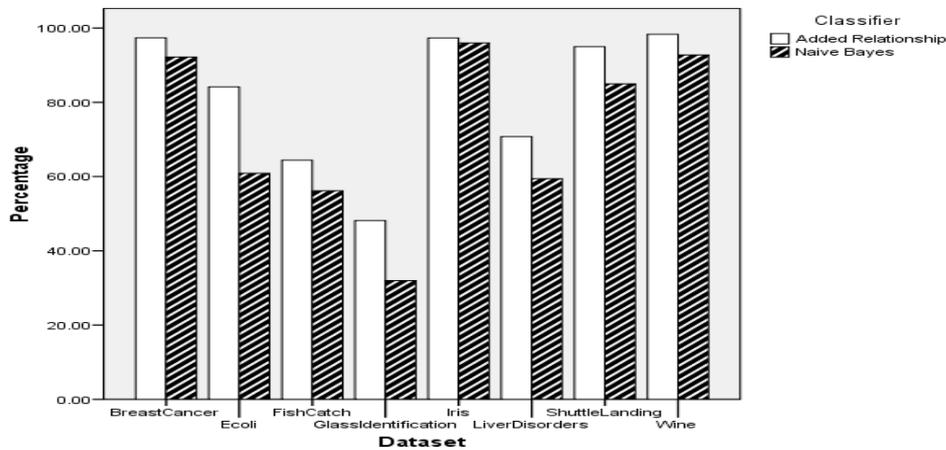

Figure 5: Comparison of accuracy between Naïve Bayes classifier and added relationship through structural learning of the attributes .

Table 2: A comparison between the accuracy of Naïve Bayes classifier and the classifier with added relationship through structural learning of the attributes using 10-fold cross validation.

Dataset	Attributes relation	Accuracy using Naïve Bayes	Accuracy using added relationship among attributes	Improvement
Breast Cancer	Yes	92.21%	97.36%	5.15%
E. coli	Yes	60.92%	84.16%	23.24%
Glass Identification	Yes	31.99%	48.14%	16.15%
Iris	Yes	96.00%	97.33%	1.33%
Liver Disorders	Yes	59.45%	70.76%	11.31%
Shuttle Landing	Yes	85.00%	95.00%	10.00%
Wine	Yes	92.78%	98.33%	5.55%
Contact Lens	No	43.00%	-	-
Fish Catch	Yes	56.17%	64.42%	0.67%

6. Conclusion

In this study, we built a classifier which is better in accuracy to the Naïve Bayes classifier. When the attributes of classification data set are not independent, structural learning is advised to be used for the analysis of the features or attributes. From the results shown, we can concluded that constructing the added relationship and learning the dependencies among the attributes can make an improvement on accuracy of

Naïve Bayes classifier. When the dependency and relationship among the attributes are strong, the improvement on accuracy of Naïve Bayes classifier will be greater.

7. Acknowledgements

8. References

[1] Heckerman, D., A tutorial on learning with Bayesian networks, Technical Report MSR-TR-95-06, Microsoft Research, (1995)

[2] Pearl, J., Probabilistic Reasoning in Intelligent Systems, Networks of Plausible Inference, Morgan Kaufmann Publishers, San Mateo, CA, USA (1988)

[3] Gilks, W.R., Richardson, S., and Spiegelhalter, D.J., Markov chain Monte Carlo in Practice (Eds.), Chapman and Hall, London, UK (1996)

[4] Ge, Y., Li, C., and Yin, Q., Study on Factors of Floating Women's Income in Jiangsu Province Based on Bayesian Networks, Adv. in Neural Network Research & Application, 67, 9, 819-827 (2010).

[5] Dougherty, J., Kohavi, R., and Sahami, M., Supervised and unsupervised discretization of continuous features, In ML-95: Proceedings of the Twelfth International Conference on Machine Learning, 194-202, Morgan Kaufmann, San Francisco, CA (1995)

[6] Lowd, D. and Domingos, P., Naive Bayes Models for Probability Estimation, In Proceedings of the Twenty-second International Conference on Machine Learning, 529-536, ACM Press, Bonn, Germany (2005)

[7] Jamain, A. and Hand, D.J., The Naive Bayes Mystery: A classification detective story, Pattern Recognition Letters, 26, 11, 1752-1760 (2005)

[8] Friedman, N. and Goldszmidt, G., Building classifiers using Bayesian networks, In Proceedings of the Thirteenth National Conference on Artificial Intelligence, 1277–1284, AAAI Press , Menlo Park, CA (1996)

[9] Krishnaveni, M., and Radha, V., Performance Evaluation of Statistical Classifiers Using Indian Sign Language Datasets, International Journal of Computer Science, Engineering and Applications (IJCSEA), 1, 5, 167-175 (2011)

[10] Yang, Y. and Webb, G.I., A Comparative Study of Discretization Methods for Naïve Bayes Classifiers, In T. Yamaguchi, A. Hoffmann, H. Motoda and P. Compton (Eds.), Proceedings of the 2002 Pacific Rim Knowledge Acquisition Workshop (PKAW'02), 159-173, Tokyo, Japan: Japanese Society for Artificial Intelligence (2002)

[11] Scutari, M., bnlearn-An R Package for Bayesian Network Learning and Inference, [Online]. [Accessed on 19th September 2011]. Available from World Wide Web: http://www.bnlearn.com/ (2011)

[12] Frank, A. and Asuncion, A., UCI Machine Learning Repository. (Irvine, CA: University of California, School of Information and Computer Science). [Online]. [Accessed on 19th September 2011]. Available from World Wide Web: http://archive.ics.uci.edu/ml (2010)

[13] Gabrosek, J., An International Journal on the Teaching and Learning of Statistics, Journal of Statistical Education. [Online]. [Accessed on 19th September 2011]. Available from World Wide Web: http://www.amstat.org/publications/jse/ (1993)

[14] Bouckaert, R.R., Choosing between two learning algorithms based on calibrated tests, In Proceedings of 20th International Conference on Machine Learning, 51–58, AAAI Press, Washington, DC USA (2003)