Pertanika J. Sci. & Technol. 24 (1): 205 - 211 (2016)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Classification Using the General Bayesian Network

Sau Loong Ang*, **Hong Choon Ong and Heng Chin Low**

School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia

ABSTRACT

Naive Bayes (NB) is a simple but powerful tool for data classification. It is widely used in classification due to the simplicity of its structure and its capability to produce surprisingly good results for classification. However, the independence assumption among the features is not practical in real datasets. Attempts have been made to improve the Naive Bayes by introducing links or dependent relationships between the features such as the Tree Augmented Naive Bayes (TAN). In this study, we show the accuracy of a General Bayesian Network (GBN) used with the Hill-Climbing learning method, which does not impose any restrictions on the structure and better represents the dataset. We also show that it gives equivalent performances or even outperforms Naive Bayes and TAN in most of the data classification.

Keywords: Naive Bayes, classification, Tree Augmented Naive Bayes, General Bayesian Network

INTRODUCTION

Classification is the organisation of patterns that require the construction of a classifier, which is a function that does grouping based on shared attributes (Madden, 2009; Ahmed *et al*., 2014). Classification problems can be found in various fields ranging from medical, information technology to chemistry (Mishra *et al*., 2011; Chung *et al*., 2013; Ahmed *et al*., 2014). There are many approaches to solve various classification problems, including decision trees, decision lists, neural networks and decision graphs (Friedman *et al*., 1997). However, the focus here is on the improvement of Naive Bayes and TAN classification to achieve accuracy and reliability of structure.

The Bayesian Network has become one of the most effective classifiers (Elgammal *et al*., 2003; Lerner, 2004; Madden, 2009). The Bayesian Network is a directed graphical model that expresses the joint distribution between multiple interacting nodes of interest based on their probabilistic relationship (Pearl, 1998; Neopolitan, 2004). By applying Markov Chain-Rule, the joint probability distribution of the nodes in Bayesian Network can be decomposed as shown in Equation [1] (Pham & Ruz, 2009).

Article history:
Received: 3 June 2015
Accepted: 10 July 2015

E-mail addresses:
dellphin1223@gmail.com (Sau Loong Ang),
hcong@.usm.my (Hong Choon Ong),
hclow@.usm.my (Heng Chin Low)
*Corresponding author

ISSN: 0128-7680

$$P_B(X_1,\ldots,X_n)=\prod_{i=1}^{n}P(X_i\mid Pa_i) \qquad [1]$$

where $P_B(X_1,\ldots,X_n)$ is the joint probability distribution over a set of n random variables $X = \{X_1,\ldots,X_n\}$ and Pa_i is the parent of X_i in a Bayesian Network. With the representation of joint distributions as a product of conditional distributions, the dependency relationship between the nodes in Bayesian Network can be identified (Pham & Ruz, 2009).

In practice, we can compute the conditional probability of one node, given the values assigned to other nodes (Cheng and Greiner, 2001). Therefore, a Bayesian Network can be used as a classifier that gives the posterior probability distribution of the class node given the conditional probability of other attributes from the training data (Friedman *et al*., 1997; Cheng & Greiner, 2001). The classification is performed based on the highest posterior probability distribution that we obtain in the class node. Therefore, Bayesian Networks are used in classification as it allows a fast and intuitive understanding among the interactive nodes (Friedman *et al*., 1997; Madden, 2009).

BAYESIAN NETWORKS CLASSIFIERS

Naive Bayes

A Naive Bayes Network is a simple probabilistic model to classify data into specific classes based on different data features (Friedman *et al*., 1997; Ong, 2011). Naive Bayes has become a core method used in classification for a variety of data ranging from medical, computer network and text recognition due to its simplicity, effectiveness and capability in capturing the data reasoning in a graphical model as shown in Fig.1 (Abraham *et al*., 2009; Zhan & Gao, 2011; Mukherjee & Sharmaa, 2012). The arcs are linked to the nodes based on the conditional probability of class C given the attributes $(X_1,\ldots,X_n)$.

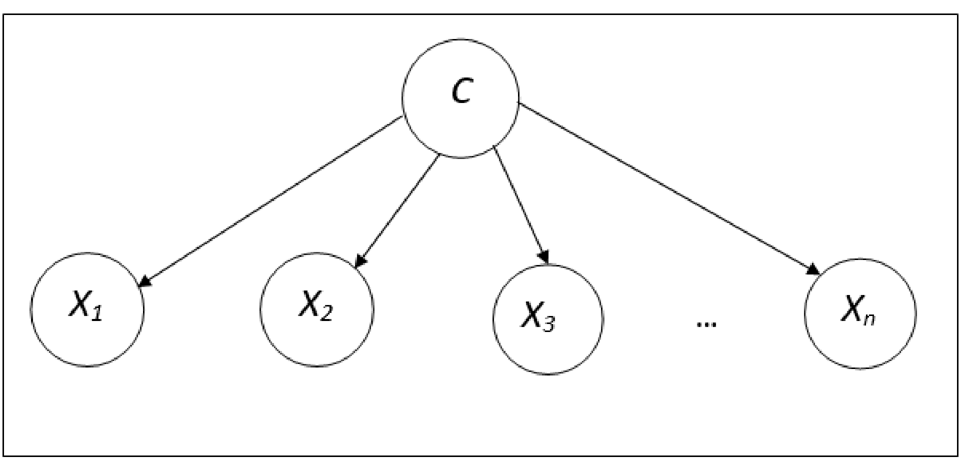

Fig.1: Graphical model of Naive Bayes.

In Naive Bayes, the classifier is set up by the assumption where the relationships between the variables ($X_1, X_2,...,X_n$) are independent given the class name C (Friedman *et al*., 1997). Under the independence assumption, the cost of the joint probability factorisation is reduced to the simplest form as shown in Equation [2] (Liew & Ji, 2009).

$$P(C \mid X) \ \alpha \ P(C)\prod_{i=1}^{n} p(X_i \mid Pa_i) \qquad [2]$$

where $p(X_i \mid Pa_i)$ is the probability of X_i given its parent Pa_i . In general, Bayesian Networks classification is based on a process to obtain the maximum value of the posterior probability of $P(C \mid X)$ as given in Equation [3].

$$P(C \mid X) = \arg_{\max} KP(C)\prod_{i=1}^{n} p(X_i \mid Pa_i) \qquad [3]$$

where K is a normalising constant.

Despite the high accuracy of classification and simplicity in data representation, Naive Bayes suffers from lack of sensitivity in showing the real relationship between the variables, which may not be totally independent (Friedman *et al*., 1997). Questions arise from researchers as to whether a modification in the strong independence assumptions can produce better results compared to Naive Bayes. One of the ways to fix the issue in the Naive Bayes is with added relationship as suggested by Ong (2011). The structure with added relationship is almost similar to Naive Bayes except for the condition where extra arcs are allowed between the variables. Extra arcs that link the dependent variables increase the validity of the graphical model in representing the data. This not only increases the reliability of the structure, but also contributes to higher accuracy of the classification as compared to Naive Bayes.

Tree Augmented Naive Bayes (TAN)

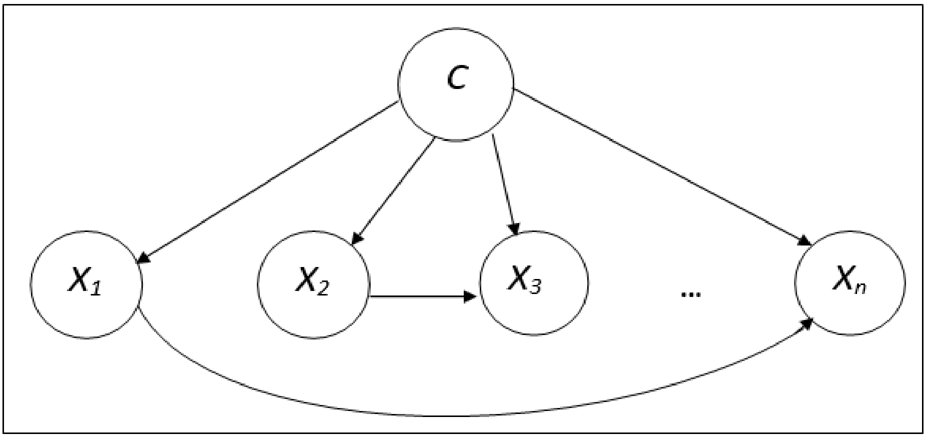

Fig.2: Graphical model of TAN.

Another prominent finding to enhance Naive Bayes is the Tree Augmented Naive Bayes (TAN) by Friedman *et al*. (1997). The structure of the Tree Augmented Naive Bayes is almost similar to the Naive Bayesian Network except for the condition where extra arcs are allowed between the variables to reduce the influence of the strong independence assumption that is made in Naive Bayes. Extra arcs that link the dependent variables increase the validity of the graphical model in representing the data as shown in Fig.2.

However, those models suffer from the aspect of reasoning and relationship between variables whereas in real datasets, the connections of the variables can be complicated and are not restricted. Even with good performance in classification, neither Naive Bayes nor TAN with added relationships is capable of capturing the topology of the Bayesian Network for classification.

General Bayesian Network in Classification

Contrary to the Naive Bayes and Tree Augmented Naive Bayes, the General Bayesian Network (GBN) offers more flexibility in forming the structure with a classifier. Firstly, there is no restriction in setting all the nodes $X_1, X_2, \ldots, X_n$ to be the child of the parent, which is the class C. Secondly, the number of parents can be more than one. With the advantage of being flexible, the relationship between all nodes including the class nodes can be captured in the structure of GBN as shown in Fig.3. However, the searching space and the parameter learning can grow exponentially if the number of parents is not controlled. Thus, the number of parents is restricted to five in order to run the classification without overloading the Bayesian Network. We apply Hill Climbing, which is a score-based structural learning method to search for the structure of the GBN. Setting the initial structure to be random, Hill Climbing adds and deletes the arc until an optimum Bayes score is achieved (Hall *et al*., 2009). To estimate the conditional probability from the learnt structure, we use the Simple Estimator in Weka (Bouckaert, 2004). GBN is a better way to perform classification since having the unrestricted way to link the variables and the class, the structure learning tends to form a Bayesian Network, which is closer to the model required by expert knowledge.

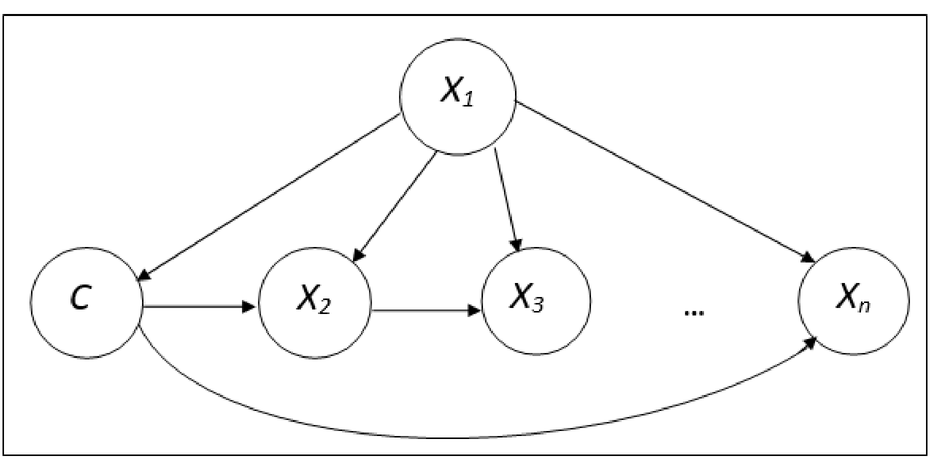

Fig.3: Graphical model of GBN.

This research is extended to various Bayesian Networks to meet different objectives based on the sizes of the dataset, accuracy, computational cost and level of simplicity in the structure (Cheng & Greiner, 1999).

RESULTS AND DISCUSSION

To measure the performance of GBN against the Naive Bayes and TAN, we used seven nominal datasets with the absence of missing values for comparative purposes. These nomimal datasets were taken from the UCI Machine Learning Repository (Lichman, 2013) and they were fed into the Naive Bayes, TAN and GBN for classification with ten-fold cross validation in WEKA software. We purposely selected seven datasets that differed in the size of the rows and number of attributes to determine the stability of Naive Bayes, TAN and GBN in classification. To show our findings, we tabulated the size of datasets, accuracy and time needed for classification for each classifier in Table 1.

TABLE 1: The Results of Naive Bayes, TAN and GBN

No.	Dataset	No. of Rows	No.of Attributes	Accuracy			Time		
				NB (%)	TAN (%)	GBN (%)	NB (s)	TAN (s)	GBN (s)
1	Vote	435	17	90.12	94.94	95.17	0	0.01	0.14
2	Breast Cancer	286	10	71.68	69.58	74.47	0	0.02	0.05
3	Fitting Contact Lenses	14	5	70.83	66.67	83.33	0	0	0
4	Chess	3196	37	87.89	92.05	94.39	0.02	0.32	2.37
5	Nurse	12960	9	90.36	94.26	94.71	0.05	0.13	0.47
6	Mushroom	8124	23	95.83	100	100	0.04	0.37	5.24
7	Balance	625	5	91.36	86.56	91.36	0.01	0	0.02

Table 1 shows the accuracy achieved by different models of the Bayesian networks, which are the Naive Bayes (NB), TAN and GBN. The time taken by those models in producing the outputs of classification is stated respectively based on the number of rows and the number attributes of the datasets.

In the first five datasets, GBN recorded the highest accuracy as compared to NB and TAN. For the remaining two datasets, which were Mushroom and Balance, GBN performed equal or better than the other two Bayesian Networks. GBN allows the nodes and the target class to form a structure without restriction as compared to NB and TAN. This gives an advantage in accuracy of the classification and also the representation in causal relationship. However, we can see that the time used to form the structure of the GBN was the longest among the Bayesian Network variants in Table 1 for all datasets except for the fitting contact lenses dataset. This is due to the structure of NB, TAN and GBN, which are almost the same for this study case. GBN consumes more time in giving results due to the complexity of the datasets especially for those that have a higher number of rows or number of attributes or both. Even with the obstacles, the eligibility and usability of GBN were still better than Naive Bayes and TAN as the effect of time consumption was minimum. We summarised the advantages and disadvantages of the three Bayesian Networks in Table 2.

TABLE 2 : Brief Comparison of Naive Bayes, TAN and GBN

Type of Bayesian Classifier	Advantages	Disadvantages
NB	-Simple in structure -Moderately stable in classifying different sizes of training datasets and producing moderately good results of classification -Fastest in producing classification results	-Not practical to represent the causal relationships of the training datasets due to the fact that assumption of independence of features can be false.
TAN	-Better connectivity among the nodes -Better classification accuracy when the training datasets are relatively large	-Restricted structure, which limits the real representation of the training datasets -Not stable if the training datasets are small
GBN	-Less restriction set in forming the structure. Good graphical structure representation of datasets. -Stable and consistently good performance to classify large and small training datasets.	-Time consuming -Stability of classification reduced when handling small training datasets

CONCLUSION

In this study, GBN was proposed as a better option compared to Naive Bayes and TAN in terms of accuracy and the reasoning of node topology in its structure. The results showed GBN capability in dealing with different ranges of complexities of the datasets in providing high accuracy of classification consistently. The learning structure of the GBN can also be applied using different learning structure methods that can be constraint-based or score-based. A full structure of GBN with all the nodes linking to each other may also increase the chance of scoring better classification results.

REFERENCES

Abraham, R., Simha, J. B., & Iyengar, S. S. (2009). Effective discretization and hybrid feature selection using Naive Bayesian Classifier for medical data mining. *International Journal of Computational Intelligence Research, 5*(2), 116-129.

Ahmed, I., Guan, D., & Chung, T. C. (2014). SMS classification based on Naïve Bayes classifier and apriori algorithm frequent itemset. *International Journal of Machine Learning and Computing, 4*(2), 183-187.

Bouckaert, R. R. (2004). *Bayesian network classifiers in Weka*. Hamilton, New Zealand: University of Waikato.

Cheng, J., & Greiner, R. (1999). Comparing bayesian classifiers. *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp.101-108.

Cheng, J., & Greiner, R. (2001). Learning Bayesian belief network classifiers: Algorithms and system. *AI '01 Proceedings of the 14th Biennial Conference of the Canadian Society on Computational Studies of Intelligence: Advances in Artificial Intelligence*, pp.141-151.

Chung, D., Lee, K. C., & Seong, S. C. (2013). General Bayesian network approach to health informatics prediction: Emphasis on performance comparison. *Procedia- Social and Behavioral Sciences, 81*, 465-468.

Elgammal, A., Duraiswami, R., & Davis, L. S. (2003). Efficient kernel density estimation using the Fast Gauss Transform with applications to color modeling and tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*(11), 1499-1504.

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers, *Machine Learning, 29*(2-3), 131-163.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations, 11*(1), 10-18.

Lerner, B. (2004). Bayesian fluorescence in-situ hybridization signal classification. *Artificial Intelligence in Medicine, 30*(3), 301-316.

Lichman, M. (2013). *UCI machine learning repository*. Irvine, CA: University of California, School of Information and Computer Science. Retrieved from http://archive.ics.uci.edu/ml

Liew, W., & Ji, Q. (2009). Learning Bayesian network parameters under incomplete data with domain knowledge. *Pattern Recognition, 42*(11), 3046-3056.

Madden, M. G. (2009). On the classification performance of TAN and General Bayesian Networks. *Knowledge-Based Systems, 22*(7), 489-295.

Mishra, M., Potetz, B., & Huan, J. (2011). Bayesian classifiers for chemical toxicity prediction. *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine,* pp.595-599.

Mukherjee, S., & Sharmaa, N. (2012). Intrusion detection using Naive Bayes Classifier with feature reduction. *Procedia Technology, 4*, 119-128.

Neapolitan, R. E. (2004). *Learning Bayesian networks.* Upper Saddle River, New Jersey, United States of America: Prentice Hall.

Ong, H. C. (2011). Improved classification in Bayesian networks using structural learning. *World Academy of Science, Engineering and Technology, 5*(3), 1073-1077.

Pearl, J. (1998). *Probabilistic reasoning in intelligent systems: Networks for plausible inference.* San Francisco: Morgan Kaufman.

Pham, D. T., & Ruz, G. A. (2009). Unsupervised training of Bayesian networks for data clustering. *Proceedings of the Royal Society A-Mathematical Physical and Engineering Sciences, 5*, 2927-2948.

Zhan, W., & Gao, F. (2011). An improvement to Naive Bayes for text classification. *Procedia Engineering, 15,* 2160-2164.