CSCanada
Management Science and Engineering
Vol. 14, No. 2, 2020, pp. 28-32
DOI:10.3968/11920

ISSN 1913-0341 [Print]
ISSN 1913-035X [Online]
www.cscanada.net
www.cscanada.org

Various Bayesian Classifiers and Their Application Fields

DOU Xiaodan[a],*

[a] Hangzhou Foreign Languages School, Cambridge A-Level Centre, Hangzhou, China.
*Corresponding author.

Received 19 September 2020; accepted 6 December 2020
Published online 26 December 2020

Abstract
The background and basic principle of Bayesian classification algorithm are briefly introduced at first, following with some different kinds of Bayesian classifiers and their application fields or scopes. Due to its simple calculation, high accuracy and high efficiency, Bayesian classification has been applied in variety of fields to replace manual or other methods. Through literature research, several typical examples are found and listed, including the usage of various classifiers in economics, finance, medicine, agriculture and so on. Meanwhile, the classification effect and future development of Bayes are also analyzed.

Key words: Bayesian classifiers; Classification; Customer classification; Character recognition; Weather prognosis

Dou, X. D. (2020). Various Bayesian Classifiers and Their Application Fields. *Management Science and Engineering*, *14*(2), 28-32. Available from: URL: http://www.cscanada.net/index.php/mse/article/view/11920
DOI: http://dx.doi.org/10.3968/11920

1. INTRODUCTION
In statistical application, there are many ways of classification. Bayes algorithm is one of the simplest but effective methods, which was firstly presented by a British mathematician Thomas Bayes in 18th Century.

Bayes algorithm is a kind of method of classification in statistics, which especially popular for the naïve Bayesian classification. Compared to other algorithms, Bayes (formula (1)) is much easier to operate, but still achieve the same high standard of accuracy and rate.

$$P(C_1, C_2, \dots C_m | X) = \frac{P(X | C_1, C_2, \dots C_m) \times P(C_1, C_2, \dots C_m)}{P(X)} \quad (1)$$

This is the basic formula of Bayes, which uses the knowledge of conditional probability. In the formula, X represents the attribute set of training samples, , while C represent categories, from to , which means there are m categories in total. Also, we can know that the probability of X and happening at the same time is equal to the of and X 's.

However, due to the hypothesis of attribute independence, Bayes often has some limitation to the features of its data sets. To expand its range of applications, some additional Bayes classifiers also invented, for instance, TAN (Tree Augmented Bayes Network), BAN and GBN, also with some other algorithms such as ARCS (Association Rule Clustering System), CBA (Classification Based on Association) and ADT (Association Decision T rec) .

2. LITERATURE REVIEW
Apart from this, through the literature collection of Bayesian classifier, it is found that Bayes has a variety of categories and have been used in the different fields, such as finance, medicine, agriculture, text and words recognition, AI and so on.

2.1 Bayes Classifiers
Fan and Shi studied about how to use hierarchical naïve Bayesian classification algorithm in make water quality

re-inspection system. To prove it, they substitute the corresponding information into the formula, construct new models and put it into the improvement of annealing algorithm . The final result tells us that Hierarchical naïve Bayes classifier can not only achieve the classification accuracy, but also obtain the classification rules of corresponding instances through clustering nodes, which reflects the superiority of the classifier. However, the computational efficiency of the construction algorithm is still not high enough now.

There are many kinds of Bayesian classifiers. In this passage, Deng and F u list some of them and give some introduction for us, which include naïve Bayes, TAN, BAN and GBN. In order to reflect their different performance, they have done some tests and gather all the data in tables to see the differences among these. There into, naïve Bayes and TAN are more suitable to those smaller data sets, while the other two show better abilities in data sets that with larger size and stronger correlation between attributes. This phenomenon indicates that we can select a most suitable classifier which with higher classification accuracy and best classification effect according to the size of the data set and the degree of correlation between attributes.

In order to relax the constraint of attribute independence and improve the generalization of naïve Bayes classifier, Sui, Jiang and Zhou have conducted some research and introduce a new learning algorithm called FISC (frequent item sets classifier) (formula (2)).

$$\arg_y \max \hat{P}(y|X) \propto \arg_y \max \hat{P}(y, X) \quad (2)$$

In the passage, authors give some explanation for the created formula, and put it into actual tests, while one of them can compare the error rates of different algorithms (including FISC, NB, AODE, LBR, TAN, LB), and the others shows the differences of t-test at 0.05 significance level among these algorithms. According to the tests, we can generally judge that FISC has relative high availability and accuracy. However, since the parameter min-sup has great influence on the performance of FISC algorithm, there is no proper technology nowadays to determine the appropriate t value for FISC except cross-validation.

Selective classifiers have been proved to be a kind of algorithms that can effectively improve the accuracy and efficiency of classification by deleting irrelevant or redundant attributes of a data set. In this essay, Chen, Huang and Tian…have introduce some different selecting Bayesian classifiers that used to analyze incomplete data, mainly about SRBC classifier (formula (3)), CBSRBC classifier (formula (4)) and RBC classifier. After introducing these classifiers, they shows some contrast tests and compare the data results. In general, CBSRBC is superior to SRBC in classification accuracy and operation efficiency, while SRBC algorithm requires less pre-specified thresholds. Therefore, we can say that the execution of the SRBC algorithm will be easier.

$$a_s = arg \max_{0 \le t \le N} \{f(\{a_i\})\}, f_{max} = f(\{aCs\}) \quad (3)$$

$$\mathrm{Chi}(A, C) = \sum_{I=1}^{m} \sum_{j=1}^{c} \frac{(e_{ij} - f_{ij})^2}{e_{ii}} \quad (4)$$

2.2 Financial Field

Zhang and Zhao have learned how to use Bayesian classifiers to identify and classify financial transactions. By analyzing customers information and the transaction data of financial institutions, they create an algorithm according to Bayes algorithm (formula (5)):

$$P(X_1|X) = \frac{P(X|X_1) \times P(X_1)}{P(X)} = \frac{\prod_{k=1}^{n} p(X_K|X_1) \times P(X_1)}{\prod_{k=1}^{n} P(X_1)} \quad (5)$$

to identify suspicious transactions and then find clues of money laundering. In order to verify, the authors did some tests to check the classification of “suspicious cluster”, “suspicious wavelet” and “suspicious link” respectively, which represent three different parameters. The results prove that, on the basis of synthesizing three suspicious financial transaction detection methods, the overall determination method is effective by applying Bayes criterion, which is better than any single kind of detection method.

Xiao and He presented a new Bayesian structural learning algorithm based on SODM (Self-Organize Data Mining) (formula (6)) and a classifier called GMBC -BDE. This will combine tow Bayesian network structure learning ideas based on dependency analysis and scoring search, then use Bayesian score as an external criterion for screening intermediate models so that it can complete the adaptive modelling process on different data sets. Through the classifier construction and comparative experiments and analysis, they can finally get the effectiveness of this method. GMBC -BDE has a relative high classification performance, and strong stability and anti-noise interference ability, which can be a new practical method for classifying customers. Meanwhile, GMBC -BDE cannot be the best one, since it only shows good performance with suitable data sets. But we can still say that the results of various classification methods are high complementary to each other.

$$y = a_0 + \sum_{i=1}^{n} a_1 x_1 + \sum_{i=1}^{n} \sum_{j=0}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \ldots \quad (6)$$

In order to help with the development of today’s retail industry, Wei, Li u and Li use naïve Bayes classification to find potential customers from customer database. By classifying the features of different customers groups, the classifier can then determine the most suitable groups for a certain product. For example, male or female, young

people or the elderly… the authors also did tests to verify the practicability of the classifier, and the results show that this method provides accurate reference information and scientific decision-making basis for the retail industry to gain insight into first opportunity and adjust the effective customer service strategy in the fierce market competition. All in all, we can say that Bayes classifier is really a suitable method for customer prediction.

To get more returns in the market, Qian and H u use naïve Bayes (formula (7)) to pick stocks. Be given the fundamental accounting and price information of stocks trading on the Shanghai Stock Exchange, the classifier can then identify stocks that are likely to outperform the market by having exceptional returns. In reality, Bayes classifiers do show us good performance in the actual tests and all the stocks chosen have higher returns than the benchmark, especially the one in the second half of 2003. However, this method is still slightly inferior to other methods that based on time series, since the correlation between stock trading returns and financial structure and accounting information of the issuing company is usually weak. Therefore, Bayesian classifier still need improvement even if it already has great utility in stock picking.

$$P(H|X) = \frac{P(X|H)\times P(H)}{P(X)} \quad (7)$$

2.3 Medical Field

The traditional pulse diagnosis needs the doctors to press the wrist to scratch the arterial pulse for pulse identification, which has strong subjectivity and fussiness, and its accuracy and reliability will be relies on the feeling and experience accumulation of doctors. Therefore, Wang and Xu find a new method that use Bayes classifier to establish pulse model and can then identify pulse in an automatic way. With pulse feature extraction and the process of unbalanced data sets, they did some tests on the classifier to verify its utility. This method can obtain knowledge directly from data, fully excavate the correlation between pulse characteristics and the causal relationship between it and pulse types, then determine the key characteristics related to pulse type, and effectively assist experts to make decisions.

Sun, Ning and Lu wrote an article about Chinese traditional medical clinical diagnosis for coronary heart disease, which is based on the Bayes classification. They use Bayesian network and naïve Bayes respectively (formula (8)), to construct models, which can more accurately classify the diagnostic information in medical cases. After introducing the algorithms, they combine them with actual medical attribute and data so that we can find out the final results of the classification. This also proves that the established models have good classification performance, in the other words, it has important practical significance, which will be helpful to improve the ability of clinical syndrome differentiation and even discover other new svndrome differentiation elements in the future.

$$P(c_j|a_1, a_2, \ldots, a_n) = \frac{P(a_1,a_2,\ldots,a_n|c_j)\times P(c_j)}{P(a_1,a_2,\ldots,a_n)} \quad (8)$$

2.4 Agricultural Field

Based on the techniques of Bayes classification, Chen, Dong and Wu did some research on how to forecast product quality for agriculture goods according to their growth characteristics. In the passage, they describe the steps in detail and show the tests they did for the classifiers. Through the “predicted results” of the tests, they find out that Bayes classifiers do have the ability to have good classification performance in product quality forecast for agriculture goods. What’s more, the classification prediction of Bayesian method has the advantages of simple form and easy to explain. High accuracy of prediction results are also easy to be promoted in different fields. However, the algorithm is still not effective for the prediction of low frequency events, which needs to be further improved.

He, Kong and Zhao have done some research on weather prognosis based on Bayes classification. They use naïve Bayes to predict rainfall problems and create an algorithm named learn-and-classify-rainfall (training set, example, V). To construct the model, they made some training sets of rainfall patterns over the years. Then on the sets, the prior probability of each prediction target and the conditional probability of each prediction factor are studied. Through the tests, they find out that Bayesian inference learning algorithm has shown good results in many practical fields and can be used in a wider rage of applications. While in the aspect of weather prognosis, the function of Bayesian classifier remains to be explored and studied in the future.

2.5 Text and Words Recognition

Lin and Guo have done some work on the recognition of handwritten Chinese characters with Bayesian classifier. They introduce Bayesian algorithms and, to verify its feasibility, they did some tests on both Bayesian classifiers and Manhattan distance function (formula (9)).

$$d_i=\sum_{j=1}^{k}\left|x_j - u_j^i\right| \quad (9)$$

Compare with the latter, Bayesian classifier reaches a higher average recognition rate at 89.21%, which can still be improved with lower rejection rate. Also, the performance of the classifier is closely related to the independence of the features selected, which means a better selection of feature may further improve the performance of the classifier.

Li, Sun, Z hang…have studied how to categorize different texts by using Bayesian classifiers. Except

naïve Bayes, the document type (formula (10)) and the word type (formula (11)), they also use EM algorithm (expectation maximum) to get more general training text database thereby expanding its application, and getting higher precision. They did some experiments and compare the classification results of different texts. The results tell us that the application of naïve Bayes classifier can be enhance by using EM algorithm. Also, naïve Bayes is a very practical classifier with high classification accuracy and there is no difference between single classifier and multi - classifiers.

$$P(w_j|c_i) = \frac{1+N(doc(w_j)|c_i)}{2+|D_c|} \quad (10)$$

$$P(w_j|c_i) = \frac{1+TF(w_j|c_i)}{|V|+\sum_{k=1}^{|V|} TF(w_j,c_i)} \quad (11)$$

2.6 Artificial Intelligence

Li and Cai studied to use Bayes classifiers to help robots obstacle avoidance, which requires the classifier to classify unknown class samples. In this case, picture processing should be done first before the classification, so that the classifier can do it work through the different features in the pictures. They also did tests for the classifier and robots, while the results show that the naïve Bates classifier has high classification accuracy and there is no difference between single classifier and multi-classifier, so it is a very practical classifier.

DISCUSSION

In general, we can see that the methods related to Bayesian classification become really practical and precise ways in many applications, but still have diverse features.

For instance, in financial and other mathematical ways, Bayes classifiers show particularly superior performances in both classification and prediction results, which successfully provide high-quality information for the users, so that they can find out the most suitable objects in those situations.

In medicine, both two cases effectively help the doctors to identify patience' illness and show good performances, which achieve meaningful and significant effect even if it is still not the best and need improvement in technical way in the future.

While turn to natural aspects, Bayes classification shows a relatively lower performances compare to others. In my own opinions, these are reasonable as the natural changes are much more uncontrollable than those activities or techniques that invented by human ourselves, which would be easier to predict and classify, while natural phenomenon are always fickle.

To sum up, Bayes classification is a method with multi- application and usually show us a high classification accuracy, despite it is easy to express and utilize. However, there are still many shortcomings in this approach that need to be improved, so that it can be able to adapt to those complex and uncontrollable changes.

CONCLUSION

According to the above information, we can know that there are variety of kinds of Bayes classifiers.

Naïve Bayes, TAN, BAN, GBN are some of the known Bayes classifiers, while first two of which are especially suitable to those smaller data sets, and BAN, GBN usually show better abilities in data sets that with larger size and strong correlation between attributes. There is another classifier named hierarchical naïve Bayes, which is a construction algorithm used to realize the clustering hierarchical relationship among attribute variables by introducing potential nodes. It once contributed to build an evaluation model for eutrophication of water quality and had good apply effect.

On account of its range of abilities, Bayes classifiers have been used in many fields.

In economic field, it once be used to classify the financial transactions and look for the suspicious trades, which reached the highest classification accuracy at 88%. Also, the problem of customer classification has been solved with a new Bayesian structural learning algorithm based on SODM (Self-Organize Data Mining) and a classifier called GMBC -BDE. Bayes even helps the retail industry to adjust the effective customer service strategy in the fierce market competition. Naïve Bayes can be used to select stocks as well so that people can earn more returns in the stock market.

In terms of medicine, Bayes classifiers also achieve great contribution. For instance, to improve the diagnosis accuracy, some models have been constructed based on Bayesian algorithms to help doctors with pulse identification. Similar methods using Bayes network and naïve Bayes also realized in Chinese traditional medical clinical diagnosis for coronary heart disease.

For the prediction of product quality for agriculture goods, it has the advantages of simple form and easy to explain, however, will get stuck for the prediction of low frequency events, which needs to be further improved. And when it is carrying out for weather prognosis, the predicted results often contain some errors in time period or even inaccurate information, which demonstrate that the functions of Bayesian classifier in this field remain to be explored and studied.

Combined with other mathematical algorithms, Bayes classifiers can also classify the written words and texts. With Manhattan distance function, Bayesian classifier reaches a higher average recognition rate at 89.21% on the recognition of hand written words. The application of naïve Bayes classifier can also be enhanced by using EM algorithm when categorizing different texts. Other

than that, the combination of different kinds of Bayes classifiers can also bring some complementary results and so improve the final accuracy of the classification.

ACKNOWLEDGEMENT

Finally, I would like to express my sincere thanks to Dr. Zhimin Chen for his professional suggestions and criticism to my paper. The essay would not have been possible without Dr. Chen's support.

REFERENCES

Chen, C., Dong, Q., & Wu, Y. J. (2011). Research on Crop product Quality Mining based on Bayesian Classification. *Journal of Anhui Agricultural Sciences,* (02), 7448-7449.

Chen, J. N., Huang, H. K., Tian, F. Z., & Fu, S. J. (2007). Selective bayes classifier for incomplete data. *Journal of computer Research and Development,* (08), 1324-1330.

Deng, S., & Fu, C. H. (2008). Four Bayesian classifiers and their comparison. *Journal of Shenyang Normal University (Natural science edition), 26*(1), 31-33.

Fan, M., & Shi, W. R. (2010). Construction algorithm and application of hierarchical naïve Bayesian classifier. *Journal of instrumentation, 31*(004), 776-781.

He, W., Kong, M. R., & Zhao, H. Q. (2007). Research on weather prediction cased on Bayesian Classifier. *Computer Engineering and Design, 28*(015), 3780-3782.

Li, J. M., Sun, L. H., Zhang, Q. R., & Zhang, C. S. (2003). The utility model relates to a naïve Bayes classifier in text processing. Journal of Harbin Engineering University, *24*(001), 71-74.

Lin, Z. Q., & Guo, J. (2002). The application of Bayesian Classifier in Handwritten Chinese Character Recognition. *Journal of Electronics*.

Obstacle avoidance of mobile robot based on Bayesian Classifier. *Control Engineering of China, 11*(004), 332-334.

Qian, Y. N., & Hu, Y. F. (2007). Stock selection using naïve Bayesian classification. *Computer Applications and Software*, (06), 90-92.

Sui, J. M., Jiang, Y., & Zhou, Z. H. (2007). Bayesian classification algorithm based on frequent itemset mining. *Journal of Computer Research and Development, 44*(008), 1293-1300.

The application of Bayesian Classification algorithm in the diagnosis of TCM clinical syndromes of coronary heart disease. *Computer Applications and Research*, (11), 164-166.

Wang, H. Y., & Xu, S. (2009). An automatic pulse recognition method based on Bayesian Classifier. *Chinese Journal of Biomedical Engineering, 028*(005), 735-742.

Wei, X. Q., Liu, H. L., Li, M. D. (2008). Application of Bayesian Classification mining technology in retail industry. *Journal of Science and Technology of Western China, 007*(027), 28-29.

Xiao, J., & He, C. Z. (2008). Structural Learning of Bayesian Classifier based on SODM and its application in customer classification. *Journal of Scientific Management,* (04), 56-62.

Yu, K. X. (0). Improvement and application of Naïve Bayes classification algorithm. (Doctoral dissertation)

Zhang, C. H., & Zhao, Xiao.Hu. (). Research on Identification of Suspicious Financial Transactions based on Bayesian Classification. *Journal of Finance and Economics.*

