Attribute Weighted Naive Bayes Classiﬁer Using a Local Optimization

 Akshay Istev Dais1, Arpitha A2, Deepthi R3, Madhuvanti Puranik4

 *Department of Information science and engineering, Srinivas Institute of Technology*

**Abstract- The Naive Bayes classiﬁer is a popular classiﬁcation technique for data mining and machine learning. It’s been shown to be very eﬀective on a variety of data classiﬁcation problems. However, the strong assumption that all the attributes are conditionally independent given the class is often violated in real world applications. Many methods have been proposed in order to boost the performance of the Naive Bayes classiﬁer by alleviating the attribute independence assumption. However, violation of the independence assumption will increase the expected error. Another alternative is assigning the weights for attributes. In this paper, we propose a novel attribute weighted Naive Bayes classiﬁer by considering weights to the conditional probabilities. An objective function is modelled and taken into account, which is based on the structure of the Naive Bayes classiﬁer and the attribute weights. The optimal weights are determined by a local optimization method using the quasisecant method. In the proposed approach, the Naive Bayes classiﬁer is taken as a starting point. We report the results of numerical experiments on several real-world data sets in binary classiﬁcation, which show the eﬃciency of the proposed method.**

**Keywords— Attribute Weighting, Classiﬁcation, Local Optimization Naive Bayes**

 Akshay Istev Dais is currently studying in Department of Information Science and Engineering at Srinivas Institute of Technology, Valachil, Mangalore

 Arpitha A is currently studying in Department of Information Science and Engineering at Srinivas Institute of Technology, Valachil, Mangalore

 Deepthi R is currently studying in Department of Information Science and Engineering at Srinivas Institute of Technology, Valachil, Mangalore

 Madhuvanti Puranik is currently studying in Department of Information Science and Engineering at Srinivas Institute of Technology, Valachil, Mangalore (E-mail: madhuvantipuranik12@gmail.com)

I. INTRODUCTION

Classiﬁcation is the task of identifying the class labels for instances based on a set of attributes. Learning accurate classiﬁers from pre-classiﬁed data is a very active research topic in machine learning and data mining. Classiﬁcation learning is the process of predicting a discrete class label C ∈ {C1,···,Cm} for a test instance X = (X1,···,Xn). One of the most eﬀective classiﬁers is the Bayesian Network (BN) introduced by Pearl [20]. A BN is composed of a network structure and its conditional probabilities. The structure is a directed acyclic graph where the nodes correspond to domain variables and the arcs between nodes represent direct dependencies between the variables. The classiﬁer represented by the BN can be expressed as:



this rule is called Bayes rule. We can see that for each class, the denominator of equation (1) is the same and it will not interfere in classiﬁcation. So, the BN classiﬁer can be rewritten as:



However, accurate estimation of P(X|Ck) is non trivial. It has been proved that learning an optimal BN is NP-hard [4] [10]. In order to avoid the intractable complexity for learning the BN, the Naive Bayes (NB) classiﬁer has been used. In the NB [15] [22], attributes are conditionally independent given the class.

However, the attribute independence assumption made by the NB classiﬁer harms its classiﬁcation performance when it is violated in reality. In order to relax the attribute independence assumption of the NB classiﬁer while at the same time retaining its simplicity and eﬃciency, researchers have proposed many eﬀective methods. These methods have been proposed in order to improve the performance of the Naive Bayes classiﬁer by alleviating the attribute independence assumption.

Another way to mitigate its attributes independence assumption is assigning weights to important attributes in the classiﬁcation. Since attributes do not play the same role in many real-world applications, some of them are more important than others. A natural way to extend the NB classiﬁer is to assign a weight to each attribute. This is the main idea of the algorithm called attribute weighted NB.

In this paper, we propose a new attribute weighted NB classiﬁer, called AWNB, which assigns more than one weight for each attribute. The number of weights for each attribute is considered as the number of class labels. These weights are written in the form of powers to the conditional attribute-class probabilities. An objective function is constructed based on the NB structure and the attribute weights. The weights, then, are determined by using a local optimization method, which here is the quasisecant method [2]. The initial weights for the quasisecant method are set to unity; this means that the NB classiﬁer is taken as an initial point. More precisely, our aim is improving the NB classiﬁer by modelling a proper objective function and optimizing the attribute weights.

II. LEARNING THE PROPOSED ATTRIBUTE WEIGHTED NAIVE BAYES USING OPTIMIZATION

Good attribute weighting can eliminate the eﬀects of noisy or irrelevant attributes. In this section, we propose a weighting procedure, in which each conditional attribute-class probability has its own power as a weight. The number of weights for each attribute is equal to the number of class labels. The idea of our weighting method is similar to the works in [26] [33], however constructing a proper objective function and utilizing the new weighting procedure are diﬀerent from the existing methods. Let us consider D = {Xi,Ci}, 1 ≤ i ≤ N, where N is the number of instances and Ci ∈ {C1,...,Cm}. Xi is an n-dimensional vector, Xi = (Xi1,Xi2,...,Xin), n is the number of attributes, and Ci is the class label. In this paper, we consider the binary classiﬁcation and assume that the two classes are, 1 and −1. Then, for each attribute, we deﬁne two weights, one corresponding to the class C1 = 1 and another to the class C2 = −1. By considering two weights for each attribute, the attribute weighted NB classiﬁes an instance Xi by selecting:



In equation (3), there are two alternatives for k in wjk. We denote these cases by wj and wj if Xi is allocated to the real class and its counterpart, respectively. Considering that Ck is the real class of Xi, the value of P(Ck|Xi) is expected to be greater than the value of P(Ck|Xi) for the majority of instances, i = 1,...,N, where Ck = −Ck. Then, it is quite natural that the value of



should be maximized, while the value of

to be minimized. The objective function (5) is similar to the objective function presented in [23]. The weights in (5) are considered as positive numbers. Also, we put an upper limit for these weights to prevent large numbers. So, we maximize the above objective function over a hyper box [a,b]. Therefore, the problem (5) can be formulated as a constrained optimization problem:



Diﬀerent methods can be applied to transfer the problem (6) to an unconstrained optimization. One of the well-known methods is the penalty method, which is used here. To ﬁnd the weights in (6), a local optimization method is applied. The NB classiﬁer is taken as a starting point for the quasisecant method. More precisely, we initialize all the weights to unity, then we use the quasisecant method to ﬁnd the attribute weights for further improvement. In other words, we search for an optimal classiﬁer starting with the NB classiﬁer. It is noted that a global optimization is also applicable to ﬁnd the global solution of the problem (6), but the complexity of the problem will increase.

III. EXPERIMENTS

 *A. Data collections*

This paper studies 16 benchmark data sets taken from the literature. A brief description of the data sets is given in Table 1. The detailed description of the ﬁrst eleven data sets used in this experiment can be found in the UCI repository of machine learning databases [1], and the last ﬁve data sets are downloadable on the tools page of the LIBSVM [3]. These data sets have been analysed quite frequently by the current data mining approaches. Another reason for selecting these data sets were that conventional approaches have analysed them with variable success.

*B. Results and discussion*

We conduct empirical comparison for the Naive Bayes (NB), the Tree Augmented Naive Bayes (TAN), the improved Naive Bayes (INB) proposed by Taheri et al. [22], and the attribute weighted Naive Bayes (AWNB) in terms of accuracy. The structure of the TAN and the INB are originated from the structure of the NB, in which each attribute has at most one augmenting edge pointing to it. The relations between attributes in the TAN are found by using the tree procedure [8], while the INB uses conditional probabilities for ﬁnding the correlations [22]. We discretize the values of continuous attributes in data sets using two diﬀerent methods. In the ﬁrst one, we apply Fayyad and Irani’s discretization method [7]. The second one is the discretization algorithm SOAC [28]. For each method, we run 50 trials and then the average accuracy over the 50 runs are calculated. The accuracy of the methods in each run is calculated using 10-fold cross validation with random orders of data records in partitioning training and test data sets to have more reliable results. More precisely, each fold contained 10% of the data set randomly selected (without replacement). For consistent comparison, the same folds, including the same training and test data sets, are used in implementing the methods. The penalty parameter is chosen as µ = 106. We set the lower and upper limits in (9) as a = 0.1,b = 10. Table 2 presents the average accuracy obtained by the NB, the TAN, the INB and the AWNB on 16 data sets, where continuous attributes are discretized by applying Fayyad and Irani’s method [7]. The results presented in this table demonstrate that the accuracy of the proposed method (AWNB) is much better than that of the NB in all data sets. It is also shown a higher accuracy of the AWNB, in general, compared to the results obtained by the TAN and the INB. The proposed method outperforms the both methods (the TAN and the INB) in most of data sets, and the accuracy of this method slightly less or almost ties with the TAN and the INB in a few cases. The results of the average accuracy obtained by the methods on 16 data sets using discretization algorithm SOAC are reported in Table 3. The results show that the accuracy obtained by the proposed method (AWNB) in all data sets are higher than those of the NB. The accuracy of the AWNB is also higher than those of the TAN and the INB in most of data sets, and the accuracy of the AWNB almost ties with those of the TAN and the INB in a few cases. Figure 2 shows the scatter plots comparing the average miss-classiﬁcations of the proposed attribute weighted Naive Bayes, AWNB, with those of the NB, the TAN and the INB using two diﬀerent discretization methods. In these plots, each point represents a data set, where the horizontal axis shows the percentage of miss-classiﬁcations according to the NB, the TAN and the INB and the vertical axis is the percentage of miss-classiﬁcation according to the proposed method, AWNB. Therefore, points below the diagonal line correspond to data sets where the AWNB performs better, and points above the diagonal line correspond to data sets where the other mentioned methods perform better.



Table 1: Brief description of data sets



Table 2: Test set accuracy averaged over 50 runs for data sets using Fayyad and Irani’s discretization method. NB stands for Naive Bayes, TAN for Tree Augmented Naive Bayes, INB for improved Naive Bayes and AWNB for attribute weighted Naive Bayes

According to the results explained above, the proposed attribute weighted Naive Bayes, AWNB, works well in that it improves the results of the NB classiﬁer. Moreover, in general, it outperforms the TAN and the INB so that it’s accuracy in most of the data sets are higher than those of the TAN and the INB. In a few cases, the TAN and the INB perform slightly better than the proposed method, and the results are acceptable as the two methods are also developments on the NB classiﬁer. The complexities of the methods are not compared in this work, since different software’s are used to implement the methods. The proposed method is coded in Matlab, while others are coded in Fortran. It is clear that the complexity of the proposed method is higher than the others due to the complexity of the optimization procedure. A global optimization is also applicable to determine the weights for the attributes. Although it may cause a better accuracy, a higher level of computational eﬀort is required.



Table 3: Test set accuracy averaged over 50 runs for data sets using discretization algorithm SOAC. NB stands for Naive Bayes, TAN for Tree Augmented Naive Bayes, INB for improved Naive Bayes and AWNB for attribute weighted Naive Bayes

III.CONCLUSION

In this paper, we proposed a classiﬁer based on attribute weighted Naive Bayes, AWNB. A novel weighting method for attribute weighted NB classiﬁer was introduced, in which for each attribute we used more than one weight depending on the number of class labels. An objective function consisting of the attribute weights based on the structure of the NB classiﬁer was then modelled to optimize the attribute weights. This objective function was optimized by a local optimization using the quasisecant method. The initial values in the quasisecant method were chosen as one; meaning that the NB classiﬁer was taken as a starting point. We carried out a number of experiments on some data sets obtained from the UCI machine learning repository and LIBSVM. The numerical results demonstrated that the proposed method has positive impact on the NB accuracy as expected. How this attribute weighting for the NB classiﬁer performs in multi class data sets remains an important question for future work.

REFERENCES

[1] A. Asuncion, D. Newman, UCI machine learning repository. School of Information and Computer Science, University of California(2007) <http://www.ics.uci.edu/mlearn/MLRepository.html>.

[2] A. M. Bagirov, A. Nazari Ganjehlou, A quasisecant method for minimizing nonsmooth functions. Optimization Methods and Software, Vol. 25, No. 1, 3-18 (2010)

[3] C. Chang, C. Lin, LIBSVM: A library for support vector machines (2001) Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm.

[4] D.M. Chickering, Learning Bayesian Networks is NP-complete, In: Fisher, D., Lenz, H. Learning from data: Artiﬁcial Intelligence and statistics V, Springer, 121-130 (1996)

[5] P. Domingos, M. Pazzani, On the optimality of the simple Bayesian classiﬁer under zero-one loss, Mach Learn 29, 103-130 (1997)

 [6] J. Dougherty, R. Kohavi, M. Sahami, Supervised and unsupervised discretization of continuous features, In Proceedings of the 12th International Conference on Machine Learning, 194-202 (1995)

[7] U.M. Fayyad, K.B. Irani, On the Handling of Continuous-Valued Attributes in Decision Tree Generation, Machine Learning 8, 87-102 (1993)

[8] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classiﬁers, Machine Learning 29, 131-163 (1997)

[9] M. Hall. A Decision Tree-Based Attribute weighting Filter for Naive Bayes, In Knowledge Based Systems, Vol 20, 120-126 (2007)

[10]. D. Heckerman, D.M. Chickering, C. Meek, Large-Sample Learning of Bayesian Networks is NP-Hard, Journal of Machine Learning Research, 1287-1330 (2004)

[11]. L. Jiang, H. Zhang, Weightily Averaged One-Dependence Estimators, Proceedings of the 9th Biennial Paciﬁc Rim International Conference on Artiﬁcial Intelligence, Guilin, China, 970-974 (2006)

[12]. L. Jiang, D. Wang, Z. Cai, X. Yan, Survey of improving naive Bayes for classiﬁcation, In Proceedings of the 3rd International Conference on Advanced Data Mining and Applications, Springer, vol. 4632, 134-145 (2007)

[13]. E.J. Keogh, M.J. Pazzani, Learning augmented Bayesian classiﬁers: A comparison of distribution-based and classiﬁcation-based approaches, In: Proc. Int. Workshop on Artiﬁcial Intelligence and Statistics, 225-230 (1999)

[14]. R. Kohavi, Scaling up the accuracy of naive-Bayes classiﬁers: a decision-tree hybrid, In: Proc. 2nd ACM SIGKDD Int, Conf, Knowledge Discovery and Data Mining, 202-207 (1996)

[15] P. Langley, W. Iba, K. Thompson, An Analysis of Bayesian Classiﬁers, In 10th International Conference Artiﬁcial Intelligence, AAAI Press and MIT Press, 223-228 (1992)

[16]. P. Langley, and S. Saga, Induction of selective Bayesian classiﬁers, In: Proc. Tenth Conf, Uncertainty in Artiﬁcial Intelligence, Morgan Kaufmann, 399-406 (1994)

[17]. J. Lu, Y. Yang, G. I. Webb, Incremental Discretization for Naive-Bayes Classiﬁer, Springer, Heidelberg, Vol. 4093, 223-238 (2006)

[18]. S. Ozsen, S. Gunecs, Attribute weighting via genetic algorithms for attribute weighted artiﬁcial immune system (AWAIS) and its application to heart disease and liver disorders problems, Expert Systems with Applications, 36, 386-392 (2009)

[19] M.J. Pazzani, Constructive induction of Cartesian product attributes, ISIS: In- formation, Statistics and Induction in Science, 66-77 (1996)

[20] J. Pearl, Probabilistic Reasoning in Intelligent Systems: networks of plausible inference, Morgan Kaufmann (1988)

[21] W. Sun, Y.X. Yuan, Optimization theory and methods: nonlinear programming, Springer (2006)

[22] S. Taheri, M. Mammadov, A. M. Bagirov, Improving Naive Bayes Classiﬁer Using Conditional Probabilities, In the proceedings of Ninth Australasian Data Mining Conference (AusDM 2011), Ballarat, Australia. Vol. 125, (2011)

[23] S. Taheri, M. Mammadov, Tree Augmented Naive Bayes Based On Optimization, Proceeding of 42nd Annual Iranian Mathematics Conference, Vali-e-Asr University of Rafsanjan, Iran (2011)

[24] S. Wang, F. Min, Z. Wang, T. Cao, OFFD: Optimal Flexible Frequency Discretization for Naive Bayes Classiﬁcation, Springer, Heidelberg, 704-712 (2009)

[25] G. I. Webb, J. Boughton, Z. Wang, Not so naive bayes: Aggregating one dependence estimators, Machine Learning, Vol. 58, 5-24 (2005)

[26] J. Wu, Z. Cai, Attribute Weighting via Diﬀerential Evolution Algorithm for Attribute Weighted Naive Bayes (WNB), Journal of Computational Information Systems 7:5, 16721679 (2011)

[27] W. Xindong et al., Top 10 algorithms in data mining, Knowl Inf Syst, 14, 1-37 (2008)

[28] A. Yatsko, A. M. Bagirov, A, Stranieri, On the Discretization of Continuous Features for Classiﬁcation, School of Information Technology and Mathematical Sciences, University of Ballarat Conference,(2010) (http://researchonline.ballarat.edu.au:8080/vital/access/manager/Repository)

[29] Y. Ying, I. Geoﬀrey, Discretization for Naive-Bayes Learning: Managing Discretization Bias and Variance, In Machine Learning, 74(1), 39-74 (2009)

[30] Y. Ying, Discretization for Naive-Bayes Learning, PhD thesis, school of Computer Science and Software Engineering of Monash University (2003)

[31] H. Zhang, S. Sheng, Learning weighted naive Bayes with accurate ranking, In Proceeding of the 4th IEEE International Conference on Data Mining, 567-570 (2005)

[32] F. Zheng, G.I. Webb, Semi-naive Bayesian Classiﬁcation, Journal of Machine Learning Research (2008)

[33] Y. Zhou T. S. Huang, Weighted Bayesian Network for Visual Tracking, Proceedings of the 18th International Conference on Pattern Recognition (ICPR’O6), 0-7695-2521-0106 (2006)

[34] Sona Taheri, John Yearwood, Musa Mammadov, Sattar Seifollahi, Attribute Weighted Naive Bayes Classiﬁer Using a Local Optimization, Neural Computing and Applications (2014)