Neighbourhood Covering Reduction by Ada-Boost Ensemble Learning Approach

Pragya Pandey* , **Deepak Kulhare, Pankaj Sahu**
Comp.Sc & Engg, R.G.P.V. Bhopal
India

Abstract— Predictive accuracy, computational speed, scalability and robustness are criteria for the evaluation of classification and predictive methods. Existing rule learning techniques having some limitation related performance. The known nearest neighbor methods are robust with the variable data set but they are sensitive to input data set. In this work, we introduced ada boost ensemble learning for reduction of randomized attributes. We focused on to nearest neighborhood classifier for defining attribute. Random attribute selection can be used to obtain a reduces representation of the data while minimizing the loss of information content because stream data content large data set with the variations and this approach gives the effective results for high dimensionality of the data. Experimental results show that if we combine the nearest neighbor classification with the randomized selection of attributes for stream data than this will be effective method handling possible uncorrelated errors and noisy data.

Keywords— Classification, prediction, nearest neighbor classification, randomized attribute selection, Ada boost ensemble method.

I. INTORDUCTION

Fix and Hodges [1] presented the simple search which is widely used in the area of pattern recognition. Nearest neighbor classifiers are based on learning by analogy, means it compares the test data with the training tuple. In general nearest neighbor method [2][3] is referred as k-NN classification in which input pattern is classified into the class with most members among the k- nearest neighbor can estimates the Bayes error and its probability of error asymptotically approaches the Bayes error[4].The basic kNN method has three shortcomings-

A. High calculation required when training set contains huge amount of data because in k- nearest neighbor, to find the k-nearest neighbor samples, all the similarities between training samples must be calculated. This technique has complex calculations and more time is needed to deal with this calculations.

B. This algorithm highly depends on training set, thus small change in training sample tends the entire recalculation.

C. All the training samples are considered equal irrespective of their weight. This phenomenon will create problem where data is distributed unevenly among samples.

In nearest neighbor approach each tuple in training set represents a point in an n-dimensional space and every tuple is described by n-attributes thus all of the training tuple are stored in n- dimensional pattern space. K-Nearest neighbor classifier searches the pattern space and finds the k-training tuple that are closest to the unknown tuple. These k- training tuples are the k-"nearest neighbor" of the unknown tuple.

One limitation of attribute reduction in traditional rough sets is that, it is applicable only to databases whose attributes can induce equivalence relations or partitions because traditional rough sets are developed on equivalent relations or partitions. Ensemble learning [5] is a machine learning model where multiple learners are trained to solve the same problem. In dissimilarity to ordinary machine learning approaches which try to learn one premise from training data, ensemble methods try to construct a set of premises and combine them to use. The new objective of ensemble, which is only the enhancement of traditional technique, is to distribute the data for processing several subset of original training set [6][7][8]. There are many effective ensemble methods. The following will brief introduction of three representative methods, Boosting, Bagging and Stacking.

In this paper, we present a nearest neighbor cover reduction though the ensemble method. We described learning methodology by Ada boost approach for cover reduction though rough set theory. Improving robustness, accuracy and understanding ability is the objective of classification modelling. Regarding instability and performance limitation of existing rule learning techniques, we present an ensemble classifier based on randomized neighbourhood covering reduction.

II. RELATED WORK

Attribute reduction, also called feature subset selection is a common problem in pattern recognition, machine leaning and data mining as there usually are many candidate attributes collected to represent recognition problems. Databases expand quickly not only in the rows (objects) but also in the column (attributes) nowadays. Some of attributes are unrelated to the learning or recognition tasks. Experiments show irrelative attributes will decline the performance of the learning algorithms for the curse of dimensionality, increase training and test times.

One way to generate an ensemble with the required properties is to train the classifiers on different sets of data, obtained by sampling from the original training set [9][11]. Breiman's bagging [9] and Freund and Schapire's boosting [10] are well known examples of successful iterative methods for improving the predictive power of classifier learning systems. Bagging uses sampling with replacement. It generates multiple classifiers by producing replicated samples of the data. Randomized attribute subsets are common example of feature partitioning, randomized attribute subsets increases the independence among classifiers [14]. To classify an instance, a vote for each class j is recorded by every classifier that chooses it, and the class with the most votes is chosen by the aggregating scheme. Boosting uses adaptive sampling. It uses all instances at each repetition, but maintains a weight for each instance in the training set that reflects its importance as a function of the errors made by previously generated hypotheses. As for bagging, boosting combines the multiple classifiers by voting, but unlike bagging boosting assigns different voting strengths to component classifiers on the basis of their accuracy.

Experimental evidence proved that both bagging and boosting are quite effective in reducing generalization error, with boosting providing in general higher improvements. This behaviour can be explained in terms of the bias-variance components of the generalization error. The variance component measures the scatter in the predictions obtained from using different training sets, each one drawn from the same distribution. The effect of combination is to reduce the variance that is what both bagging and boosting achieve. In addition, boosting does something more. By concentrating the attention of the weak learner on the harder examples, it challenges the weak learner algorithm to perform well on these harder parts of the sample space, thereby reducing the bias of the learning algorithm.

It turns out that sampling the training set is not effective with NN classifiers [12]. To gain some insights as to why this is the case, let us analyse the conditions under which the bagging procedure is effective. As observed above, bagging reduces the variance component of the generalization error. When the weak learner is unstable with respect to variations in the training set, perturbing the training data can cause significant variability in the resulting predictor. Thus, bagging the ensemble improves accuracy in this case. Suppose the weak learner is the NN classifier. It has been shown that the probability that any given training point is included in a data set bootstrapped by bagging is approximately 63.2%. It follows that the nearest neighbor will be the same in 63.2% of the nearest neighbor classifiers [16]. Thus, errors are highly correlated, and bagging becomes ineffective.

III. ATTRIBUTE REDUCTION THROUGH ADABOOST ENSEMBLE APPROACH

The key idea of Boosting algorithm is to transfer a nearest neighbor classifier to a strong one by integration and train for attribute reduction. Adaboost algorithm is a kind of Boosting algorithms, which is an adaptive Boosting one. Adaboost algorithm can adjust weight distribution of the training samples adaptively, and consistently select the best nearest neighbor classifier of sample weight distribution, to integrate all nearest neighbor classifier and vote by a certain weight to form a strong classifier. Adaboost algorithm [13][15][18] combines nearest neighbor classification with index selection, and reaches the key indexes selection on the basis of forecast accuracy.

1) Given train sample set $S = \{(x_1, y_1), (x_2, y_2), \ldots \ldots \ldots \ldots \ldots (x_n, y_n)\}$ neighbourclassifier space H, $x \in X$, X is a sample space, $y = \{1, 2, 3, \cdots, K\}$ is a class label set. Initiating sample probability distribution $D_t(i) = 1/n, i=1,2,\ldots,n$.

2) For t=1,2……. T, T is the feature numbers needed.
   To every nearest classifier h of *H*, we can do below:

a) Dividing sample space *X*, we can get $x_1, x_2, \ldots \ldots \ldots \ldots \ldots x_m$

b) Under the training sample probability distributionD, we can calculate

$$D(x,y) = \| x - y \| = \sqrt{(x - y) - (x - y)}$$

$$D(x,y) = \sum_{i=1}^{m} ((x_i - y_i)^2)^{1/2}$$

Improving accuracy, robustness and understand ability is the objective of classification modeling. We have presented the method to ensemble multiple classifiers with randomized neighborhood reduction that overcomes the performance limitation of existing rule learning techniques.

III. EXPERIMENTAL RESULT

Below are the results of some of the more significant tests performed using by nearest neighbor classification based on ensemble method called as adaboosting ensemble method.In this section, we first compare the accuracy of these algorithms on different datasets. Then we present the comparative results with some classical feature selection algorithms. Finally, we discuss the influence of parameters in the neighborhood model.

TABLE I
DESCRIPTION OF DATASETS

Dataset	Features	Class	Instances
Wine	14	3	178
Iris	5	3	150
Wdbc	31	2	569
Pima	9	2	768

Furthermore, we collect ten classification tasks from UCI machine learning repository. The description of these data sets is given in Table 1. The accuracies based on Nearest Neighbor rule (NN),Neighborhood Classifier (NEC)[17], Learning Vector Quantization (LVQ) [19] and proposed work are presented in Table 2.

TABLE II
PERFORMANCE OF NEIGHBOURHOOD ATTRIBUTE REDUCTION BASED ADABOOSTINGENSEMBLE LEARNING METHOD

Dataset	Accuracy			
	NEC	LVQ	1NN	Proposed work
Wine	96.6	96.0	94.9	97.6
Iris	96.0	96.6	96.8	96.9
Wdbc	94.6	95.4	98.7	99.0
Pima	76.0	73.3	76.6	76.9

IV. CONCLUSIONS

Reducing redundant or irrelevant attributes can expand classification performance in most of cases and decrease cost of classification. We design a feature evaluating function, called neighbourhood dependency, which reflects the percentage of samples in the decision positive region. Theoretical arguments show that the significance of features monotonically increases with the feature subset. This property is important for integrating the evaluating function into some search strategies. Then adaboosting ensemble feature selection algorithms are constructed based on the dependency function. The future work will be focused on constructing neighborhood classifiers with the proposed model to lay a foundation for neighborhood based learning systems, such as k-nearest neighbor methods and neighborhood counting methods.

REFERENCES
[1] Fix, E. and Hodges, J.: Discriminatory analysis. Nonparametric discrimination: Consistency properties, Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.
[2] Cover, T.M., Hart, P.E.: Nearest neighbor pattern classification. IEEE Trans. On Information Theory (1967) 21–27
[3] Dasarathy, B.V.: Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques.
[4] Duda, R.O. and Hart, P.E.: Pattern classification and scene analysis, New York: Wiley, 1973.
[5] Gertheiss, Jan and Tutz, Gerhard: Feature Selection and Weighting by Nearest Neighbor Ensemble .Technical Report Number 033,2008,University of Munich
[6] Mollineda, R.A., Sotoca, J.M., Sanchez, J.S.: Ensembles of data-reduction-based classifiers for distributed learning from very large data sets:TIC2003-08496 from the Spanish CICYTM.
[7] Chan, P., Stolfo, S. : Learning arbiter and combiner trees from partitioned data for scaling machine learning. In Usama, F., Uthurusamy, R., eds.: Proceedings of International Conference on Knowledge Discovery and Data Mining (KDD-95),Montreal, Canada, AAAI Press (1995) 39–44
[8] Chawla, N., Eschrich, S., Hall, L.: Creating ensembles of classifiers. In: Proceedings of International Conference on Data Mining (ICDM'01), San Jose, Califormia, IEEE (2001) 580–581
[9] Breiman, L.: Bagging predictors. Machine Learning, vol. 24, issue 2, 1996, 123–140.

[10]   Freund, Y. and Schapire, R.: A decision-theoretic generalization of on-line learning and application to boosting, Journal of Computer and System Sciences, Vol. 55, issue 1, August 1997, 119–139.

[11]   Samet, H.: K-Nearest Neighbor Finding Using MaxNearestDist, IEEE Transaction Pattern Analysis Machine Intelligence, vol. 30, issue 2, Feb. 2008, 243-252.

[12]   Hu, Qinghua, Yu, Daren, Xie, Zongxia: Neighborhood classifiers, Expert Systems with Applications (2007), doi:10.1016/j.eswa.2006.10.043.

[13]   Freund, Y., Schapire, R.: Experiments with a new boosting algorithm. In: Proceedings of the Thirteenth International Conference on Machine Learning. (1996) 148–156

[14]   Skurichina, M., Duin, R.P.W.: Bagging and the random subspace methods for redundant features spaces. In: Multiple Classifier Systems, J. Kittler, F. Roli(eds.), LNCS 2096,   Springer-Verlag (2001) 1–10

[15]   Xuan Liu, Xiaoguang Wang, Nathalie Japkowicz, Stan Matwin: An Ensemble Method   Based on AdaBoost and Meta-Learning, Lecture Notes in Computer Science Vol 7884, 2013, pp 278-285.

[16]   Quinlan, J. R.: Bagging, boosting and C4.5, Fourteenth National Conference on Artificial Intelligence, 1996.

[17]   Samworth, Richard J.: Optimal weighted nearest neighbor classifiers, The Annals of Statistics, 2012, Vol. 40, No. 5, 2733–2763.

[18]   Opitz, D. and Maclin, R.: Popular ensemble methods: An empirical study, Journal of Artificial Intelligence Research, 11:169–198, 1999.

Mwebaze, E., Schneider, P., Schleif, F.M., Aduwo, J.R., Quinn, J.A.,Haase,S., Villmann, T., Biehl, M. (2011) :Divergence based classification and Learning Vector Quantization, Neurocomputing, vol. 74, 2011,1429-1435.