Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 4– No.3, January 2016 – www.caeaccess.org

An Incremental Ensemble of Classifiers based on Hypothesis Strength and Ambiguity Grade

P. R. Deshmukh, PhD
Professor,
Dept. of Computer Science and Engg,
Sipna College of Engineering and Technology,
Amravati, Amravati University, Maharashtra,

Roshani Ade
Research Scholar,
Dept. of Computer Science and Engg
Sipna College of Engineering and Technology,
Amravati, Amravati University, Maharashtra

ABSTRACT
The massive amount of raw student's data in the education organization can be converted into the information and buried knowledge can be taken out of it for the purpose various applications related to students. As the student's data in the educational systems is increasing day by day, so instead of classical batch learning algorithm, incremental learning algorithm tries to forget unrelated information while training fresh examples. Now a days, combining classifiers is nothing but taking more than one opinion contributes a lot, to get more accurate results. Therefore, a suggestion is an incremental ensemble of two classifiers namely Naïve Bayes, K-Star using voting scheme based on hypothesis strength and ambiguity grade. The voting rule proposed in this paper is compared with the existing majority voting rule for the student's data set.

General Terms
Machine Learning.

Keywords
incremental learning, ensemble, voting rule.

1. INTRODUCTION
Now a day's choosing right career is one of the most important aspects of the students learning process, and it is difficult to choose the right career option when the number of options are available to choose. It is very much important to consider the interest, talent, expected progress in a specific area, before choosing a career. It is commonly seen that, many of the students have their deprived educational record because of choosing their career without considering their own capabilities and it will cause waste of time and the money, so it is important to choose the right career in the first place. It is also observed that there is an impact of psychological parameters for choosing a right career option [1-5]. The psychometric test [6-7] is conducted on the students and the students are classified for choosing their right career option. In this research, only supervised learning methods has been considered . There are some terminologies used throughout this paper. The set of instances (student samples), each described with many attributes, also called as features. The attributes are independent observable variables, which are numeric or nominal. Each object has been assigned a single value, ie , a value of the dependent (target) variable, which is a function of the independent variables. Thus, the input data for a learning task is a collection of records. Each instance, also known as an example, characterized by a tuple (x,), where x is the attribute set and y is the target class, nothing but class label. For learning of target function, the learner L is presented with a set of training examples, each consisting of an input vector x_i from x, along with its target function value $y=f(x)$. The function to be learned represents a mapping from the attribute space X to the space of real values Y, ie, $X \rightarrow Y$. It is assumed that the training examples are created at random according to the probability distribution D. In general, D can be any distribution and is not known to the learner. Given a set of training examples of the target function $t: X \rightarrow Y$ the problem faced by the learner is to hypothesize, or estimate, f. The symbol H has been used to denote the set of all possible hypotheses that the learner may consider when trying to find the true identity of the target function. H is determined by the set of all hypothesis generated from different base classifiers over the instance space X. After observing a set of training examples of the target function t, L must output some hypothesis h from H, which is its estimate of f. A fair evaluation of the success of L assesses the performance of h over a set of new instances drawn randomly from X, Y according to D, the same probability distribution used to generate the training data.

The paper is ordered as follows, section 2 gives brief introduction of combining multiple hypothesis. Section 3 describes the concept of incremental ensemble of classifiers. The voting scheme based on hypothesis strength is given in section 4. The section 5 talks about the experimentation and results of the proposed method. Finally, section 6 gives the conclusion of the work and its future scope

2. COMBINING MULTIPLE HYPOTHESIS
In ensemble learning, the multiple hypothesis, which supports the final decision making process are combined together to make a final decision. The accuracy of the model can be improved with ensemble learning strategy and the robust model can be built by using ensemble learning concept compared to the model which generates single hypothesis, therefore it has attracted increasing interest in the machine intelligence society. Research in the ensemble learning is expanding rapidly, with many creative ideas, the work includes the combined classifier systems [8-9], experts mixture [10], stacked generalization approach [11], combining of multiple classifiers [12] and bootstrap inspired techniques [13]. The second key component of an ensemble system is the strategy for combining the output of various classifiers, as the output of multiple classifiers combined together to reach to the final decision. Assuming there is a hypothesis set (HS) containing, N hypotheses.

$HS=\{h_1,\ldots, h_N\}$ (1)

For every testing example x_t, each hypothesis can vote an estimate of a posterior probability across all the possible class labels y_j,

$P_i(y_j|x_t)$, i=1,…, and $y_j\ \epsilon\{1,\ldots,C\}$ (2)

It is required to find an ensemble strategy based on individual $P_i(y_j|x_t)$, from each hypothesis h_i. The voting methods are described in detail [8]. There are various rules for combing multiple hypothesis namely Geometric average (GA) rule, Arithmetic average (AA) rule, Median (Med) Rule, Majority (Maj) Rule, Max Rule, Min Rule, Borda Count (BoC) Rule, Weighted Arithmetic Average Rule and Weighted Maj Voting Rule

3. INCREMENTAL ENSEMBLE OF CLASSIFIERS

Instance-based learning is a machine learning method which classifies new examples by comparing them to those already seen and in memory. Instance-wise incremental learning, also called as online learning and have a very restricted choice of classifiers [14-16] and it may require a huge number of instances to run. In case of online learning, the algorithms continually modify its hypothesis as it receives the samples. In this case, the model frequently receives a sample, then prediction will be done and the consequently hypothesis will be updated. The instance-wise incremental learning is useful in many applications like, computer security, market basket analysis, intelligently acting user interfaces and many more. In the students classification system, instance-wise data handling is required in the applications like, online distance-based education system. Some of the important characteristics of an instance-wise incremental learning algorithms are:

1. When training, it should require some constant time per sample.
2. There should be only one sample at a time in memory, so the some memory will be used.
3. It will build the model by just scanning the database only once.

It assumes infinite stream of data, but process it under finite time and memory resources.

When multiple incremental classifiers are combined using the voting technique, called as an incremental ensemble of classifiers. The results can be improved based on the belief that the majority of the experts are more likely to be correct in their decision when they agree with their opinion. The incremental online algorithms available in the literature which are able to handle data incrementally are as follows. The datasets which has been used for this research are having numeric attribute values only. So it is not possible to use the ensemble of incremental classifiers which cannot handle numeric values. Fig. 1 shows an incremental ensemble of classifiers. Based on this structure, the number of incremental learning algorithms are combined to reach to the final decision. When multiple classifiers are combined using different voting methods, the good output can be expected, considering the number of classifiers are more capable to be correct in their decision when they agree in their approximation. The incremental ensemble of naïve bayes and K-star is shown in Fig. 1. The hypothesis of both the classes has been combined using voting rules.

3.1 Naïve Bayes Updatable

Naive Bayes is an extensively known instance-based classifier, it simply updates the internal counter with each instance and uses these counter to assign a class in a probabilistic manner to the new item from the stream data.

It is an incremental form of Bayesian networks, as it assumes that each attributes are not dependent on the remaining attributes. The naïve Bayes algorithm usually used for a batch learning, because when algorithm handles each training example separately, it could not perform its operations well, described in [2, 17]. As per the features of the incremental learning algorithm, the naïve Bayes algorithm can be trained by using one pass only as per the stages given below:

1. Initialize count and total=0
   a. Go through all the training samples, one sample at a time
   b. Each training sample, □ (□,) will have its label associated with it.
   c. Increment the value of count, as it goes through the particular training sample.
2. The probability is calculated by dividing the individual count by the set of training data samples of the similar class attribute.

Compute the previous probabilities (y) as the portion of entirely training samples which are in class y.

3.2 K-star

The KStar (K*) algorithm can be defined as a method which partitions the number of observations into k groups. It is an instance based learner, which used the entropy based distance function. It can handle the real value attributes, attributes having symbolic values, missing values. It is a instance-based learner, where the test instance case is decided by using the class label of training samples based on some kind of similarity function. It uses the entropy based distance function [18], based on the probability of transforming one example into another by randomly choosing between all possible transformations and turns out to be much better than Euclidean distance for classification.

4. VOTING SCHEME

The voting scheme [20] based on hypothesis strength has been used with incremental ensemble of classifiers. Let, there is an ensemble system with m hypotheses having hypothesis strength HSj as a condition related to posterior probability $P(y_j|x_t)$. Each hypothesis is having its hypothesis strength and the ambiguity grade associated with it. There are two types of certainty associated with the classification problem lower-most certainty and topmost certainty. For two class classification problem, P_j=0.5, is a lowermost certainty, means out of the two classes each one is equally likely. For topmost certainty, P_j=0 or P_j=1, means hypothesis are certain about the class label. If the same concept has been applied to the multiclass classification problem, then it will be transferred to the two class classification problem. Let, given a class label y_i, the predicted label y_t of any test instance x_t can be represented using Boolean type. yt=Yi or $yt\epsilon\bar{Y}$, where $\bar{Y}i=\{Yl,l\neq i\}$. The hypothesis strength, HSm can be represented as $|Pj-0.5|$ and the ambiguity grade

AGm=0.5−HSm. The combined hypothesis strength and ambiguity grade in the ensemble voting system is defined using equation 3 and 4.

HS=$w1HS1$+$w2HS2$+⋯+$wmHSm$=Σ$wlHSlml$=1 (3)

AG=$w1AG1$+$w2AG2$+⋯+$wmAGm$=Σ$wlAGlml$=1 (4)

wj is normalized so that Σwj=1mj=1

The knowledge level of hypothesis j is represented by hypothesis strength HSj and ambiguity grade AGj.

Weight assigning strategy of the algorithm:
Higher weights will be assigned to the classifiers having higher hypothesis strength and lower ambiguity grade, means they are more certain of their decisions. Lower weights will be assigned to the classifiers having lower hypothesis strength and higher ambiguity grade, means they are more uncertain about their decisions. The relation between the same are defined in equation (5). The weight wj should be proportional to the ratio of hypothesis strength to ambiguity grade.

wj∝αj=$HSjAGj$ (5)

After obtaining the decision profile, which is nothing but the voting probability from each hypothesis for each testing instance across all possible class identity labels, the hypothesis strength HSj and the ambiguity grade AGj for each element in the decision profile is defined.

HSj=|Pj−0.5| (6)

AGj=0.5−HSj (7)

Where, [0,1], HSj∈[0.0,0.5], AGj∈[0,0.5]

The hypothesis strength and its direction has been shown with

$H\tilde{S}j$=Pj−0.5 Where, $H\tilde{S}j$∈[−0.5,0.5] (8)

The ratio of hypothesis strength to ambiguity grade with considering its direction is defined by using equation (9).

$\tilde{\alpha}j$=$H\tilde{S}jAGj$=$H\tilde{S}j$0.5−|$H\tilde{S}j$| (9)

Using equation (4) and (5), $\tilde{\alpha}$, from various hypothesis for each class label Yi can be combined, the aggregate hypothesis strength to ambiguity grade ratio with its direction, denoted by $\tilde{\alpha}out$was calculated using equation (10).

$\tilde{\alpha}out$=Σ$wlH\tilde{S}mml$=1Σ$wl\ A\tilde{G}mml$=1 (10)

According to equation (5.9),

$\tilde{\alpha}out$=Σ$\alpha lH\tilde{S}mml$=1Σ$\alpha l\ A\tilde{G}mml$=1 (11)

$Pout$=$H\tilde{S}out$ + 0.5 (12)

$Pout$ gives final voting probability for each class label. The maximum output of all the decisions will be final decision.

The steps are summarized in Algorithm below.

Input: $a)$ $A\ set\ of\ classifiers\ m.$

$h_j, j = 1, \dots, m, each\ trained\ on\ the\ training\ data\ Tr\ by\ subspace\ method$

$b)$ $A\ testing\ instance\ x_t \in Te, where\ Te\ is\ a\ testing\ set$

Procedure:

1) $Apply\ each\ testing\ instance\ x_t\ to\ each\ classifier\ h_j\ and\ get\ decision\ P(Y_i|x_t), where\ Y_i = 1, \dots, C, where\ C\ is\ a\ class\ labels.$

2) $Based\ on\ each\ column\ of\ the\ P(Y_i|x_t)\ , the\ hypothesis\ strength\ HS_{Y_i}, the\ ambiguity\ grade\ AG_{Y_i}\ for\ each\ class\ label\ is\ calculated$.

$$HS_{Y_i} = |\ P(Y_i|x_t) - 0.5|$$

$$\widetilde{HS}_{Y_i} = P(Y_i|x_t) - 0.5$$

$$AG_{Y_i} = 0.5 - HS_{Y_i}$$

3) $Calculate\ \tilde{\alpha}_{Y_i}$

$$\tilde{\alpha}_{Y_i} = \frac{\widetilde{HS}_{Y_i}}{AG_{Y_i}}$$

4) $Calculate\ \tilde{\alpha}_{out}(Y_i)$

$$\tilde{\alpha}_{out}(Y_i) = \frac{\sum_{l=1}^{m} \tilde{\alpha}_{Y_i}(l)\widetilde{HS}_m}{\sum_{l=1}^{m} \tilde{\alpha}_{Y_i}(l)\ \widetilde{AG}_m}$$

5) $Calculate\ \widetilde{HS}_{out}(Y_i)$

$$\widetilde{HS}_{out}(Y_i) = \frac{\tilde{\alpha}_{out}(Y_i)}{2(1 + |\tilde{\alpha}_{out}\ (Y_i)|)}$$

6) $Final\ voting\ probability\ calculation$

$$P(Y_i|x_t) = \widetilde{HS}_{out}(Y_i) + 0.5$$

Output: $predicted\ class\ label\ y_i$

$$x_t \to y_i\ satisfy\ \max_{Y_i} P(Y_i|x_t)$$

Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 4– No.3, January 2016 – www.caeaccess.org

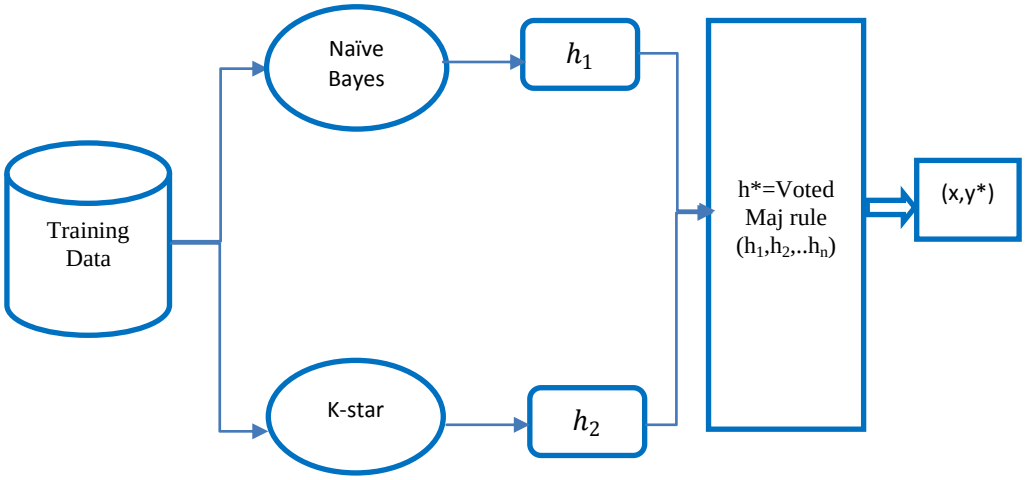

Fig. 1. Incremental ensemble of Naïve Bayes and K-star classifiers

5. EXPERIMENTS AND RESULTS

For the purpose of this study, the data set has been created by conducting some tests on students. The attributes of the tests are shown in Table I. There is no natural order in the dataset. The accuracy is calculated by dividing the whole training set into equal proportion ten sets, ie 10 cross validation is used.

The experimentation has been done by using the free available source code by Witten and Frank[19]. Table II compares the output of different algorithms of first experiment and shows that the proposed algorithm gives good accuracy by using voting rule descried in section 4. The results are compared with the majority voting rule is stated in equation 13.

$$x_t \rightarrow y_j \, satisfy \, {}^{max}_{y_j} \sum_{i=1}^{L} \Delta_i \; (y_j | x_t) \qquad (13)$$

Table 1. Table captions should be placed above the table

	A Self Awareness	B Empathy	C Self Motivation	D Emotional Stability	E Managing Relations	F Integrity	G Self Development
Mean	7.708	18.468	11.159	7.569	8.922	8.73	7.447
Std. Dev.	4.096	10.405	5.94	3.977	4.737	4.971	4.055
Class 1	1 to 2	1 to 4	1 to 3	3 to 4	3 to 5	6 to 6	5 to 6
Class 2	3 to 3	5 to 9	4 to 6	1 to 2	1 to 2	1 to 3	1 to 2
Class 3	5 to 6	10 to 14	7 to 9	5 to 6	6 to 8	4 to 5	3 to 4
Class 4	7 to 8	15 to 20	10 to 12	7 to 8	9 to 10	7 to 9	7 to 8
Class 5	9 to 10	21 to 25	13 to 15	9 to 10	11 to 12	10 to 12	9 to 10
Class 6	11 to 12	26 to 30	16 to 18	11 to 12	13 to 14	13 to 15	11 to 12
Class 7	13 to 14	31 to 35	19 to 21	13 to 14	15 to 16	16 to 18	13 to 14
	H Value Orientation	**I Commitment**	**J Altruistic Behavior**	**K Sub-I**	**L Sub-II**	**M Sub-III**	**N Sub-IV**
Mean	10.998	13.189	17.666	12.542	10.926	7.536	10.189
Std. Dev.	6.475	6.821	9.976	6.747	6.299	3.966	5.475
Class 1	7 to 9	1 to 5	1 to 5	1 to 5	4 to 5	1 to 2	3 to 6
Class 2	1 to 3	6 to 8	6 to 10	6 to 7	7 to 9	3 to 4	1 to 3
Class 3	4 to 6	9 to 12	11 to 15	8 to 10	11 to 12	5 to 6	7 to 9
Class 4	10 to 12	13 to 15	16 to 20	11 to 15	13 to 14	7 to 8	10 to 12
Class 5	13 to 15	16 to 18	21 to 25	16 to 18	15 to 19	9 to 10	13 to 14
Class 6	16 to 17	19 to 21	26 to 30	19 to 20	1 to 3	11 to 12	15 to 16
Class 7	20 to 24	22 to 24	31 to 35	21 to 23	22 to 22	13 to 14	17 to 18

Table 2. Accuracy of algorithms and voting scheme used

The proposed ensemble algorithm	Naïve Bayes	K-star
90.8 Majority Voting	89.6	89.2
92.3 Voting Rule		

Table 3 Comparing the Proposed ensemble with Adaboost.SVM, SVM, Multilayer Perceptron

Proposed ensemble algorithm with voting rule	Adaboost.SVM	SVM	Multilayer Perceptron
92.3	90.4	88.1	91.3

Table 4. Training time (in sec) of a proposed algorithm by using a dual core 2 GHz system with 2GB memory

Proposed ensemble algorithm with voting rule	Adaboost. SVM	SVM	Multilayer Perceptron
6.51	4.13	0.35	5.32

6. CONCLUSION AND FUTURE SCOPE

It has been observed that the use of voting scheme improves the result of the ensemble of incremental learning. The time required for training is more as compared to the methods in the literature so there is scope to reduce the complexity of the proposed voting scheme. There is scope to reduce the complexity of the system.

7. REFERENCES

[1] Roshani Ade, Dr. P. R. Deshmukh, "Classification of Students using Psychometric Tests with the help of Incremental Naïve Bayes Algorithm", International Journal of Computer Application, Foundation of Computer Science, USA, vol 89, no. 14, pp. 27-31 March,2014.

[2] Roshani Ade, Dr. P. R. Deshmukh, "An incremental ensemble of classifiers as a technique for prediction of student's career choice", published in IEEE International conference on network and soft computing, pp. 427-430, ICNSC, July 2014.

[3] Roshani Ade, Dr. P. R. Deshmukh, "Classification of students by using an incremental ensemble of classifiers", published in 3rd IEEE International Conference On Reliability, Infocom Technologies and optimization, pp. 61-65, ICRITO- 8-10 Oct 2014.

[4] Roshani Ade, Dr. P. R. Deshmukh, "Instance based vs Batch based incremental learning approach for Students Classification, International Journal of Computer Application, Foundation of Computer Science, USA, vol. 106, no. 3, pp. – Nov 2014.

[5] Roshani Ade, Dr. P. R. Deshmukh, "Efficient Knowledge Transformation System Using Pair of Classifiers for Prediction of Students Career Choice", International Conference on Information and Communication Technologies, ICICT Dec 2014.

[6] Garrett H.E., Vakils, Feffer and Simons, "Statistics in Psychology and Education," 1981.

[7] Mayer, JD and Salove P., "The intelligence of Emotional Intelligence," pp. 433-442, 1993.

[8] Robi Polikar, "Ensemble based syste.ms in decision making," IEEE Circuits and systems magazine, vol. 6, no. 3, pp. 21-45, Third Quarter, 2006.

[9] C. Ji and S. Ma, "Combination of weak classifiers," IEEE Trans. Neural Networks., vol. 8, no. 1, pp. 32–42, Jan. 1997.

[10] Kittler,M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," IEEE Trans. Pattern Anal. Machine Intelligence. vol. 20, pp. 226–239, 1998.

[11] D. H. Wolpert, "Stacked generalization," Neural Network., vol. 5, no. 2, pp. 241–259, 1992.

[12] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier system," IEEE Transaction on pattern analy. Machine Intel.,vol. 16, no.1, pp. 66-75, 1994.

[13] R. Polikar, "Bootstrap-Inspired techniques in computation intelligence," IEEE Signal Process. Mag., vol. 24, no.4, pp. 59-72, Jul. 2007.

[14] L. I. Kuncheva, "Classifier ensembles for detecting concept change in streaming data: Overview and perspectives," in Proc. Eur. Conf. Artif.Intell., 2008, pp. 5–10.

[15] M. Muhlbaier, A. Topalis, and R. Polikar, "Learn++.NC: Combining ensemble of classifiers with dynamically weighted consult-and-vote for efficient incremental learning of new classes," IEEE Trans. Neural Netw.,vol. 20, no. 1, pp. 152–168, Jan. 2009.

[16] C. Giraud-Carrier, "A Note on the Utility of Incremental Learning," Artificial Intelligence Comm., vol. 13, no. 4, pp. 215-223, 2000.

[17] Sylvain Roy, "Nearest Neighbor With Generalization," University of Canterbury, Christchurch, New Zealand, 2002.

[18] Cleary, John G., and Leonard E. Trigg. "K*: An Instance-based Learner Using an Entropic Distance Measure," ICML, 1995.

[19] Witten I. H, Frank E, Hall MA, "Data Mining Practical Machine Learning Tools and Techniques," Third Edition, Elsevier, 2011.

[20] Haibo He, Yuan Cao, "SSC: A Classifier Combination Method Based on Signal Strength", IEEE Transaction on Neural Networks and Learning Systems, Vol. 23, No., July 2012.