Instance-based vs Batch-based Incremental Learning Approach for Students Classification

Roshani Ade
Research Scholar
Department of Computer Science and Engineering
Sipna College of Engineering and Technology, SGBAU,Amravati

P. R. Deshmukh Ph.D.
Professor
Department of Computer Science and Engineering
Sipna College of Engineering and Technology, SGBAU,Amravati

ABSTRACT

It is difficult to find the hidden information in the educational database system, because of the rapid increase of the student's data. The hidden information from the educational databases can be used for the various predictions like students' performance, offering different career choices to students, prediction of student's enrollment into various courses and many more. This data can be learned incrementally by using instance based or batch based approach. The instance based method is just like an online learning, the system will handle each instance incrementally, the algorithm itself is an updatable, and the knowledge will be updated by every instance in time. In the batch based approach, instances are coming in the batches and will be operated in a bulk, so the processing time requires for it is less as compared to instance based approach and learning new concept is possible when the data is available in a batches. The paper proposes three approaches of incremental learning and compares for handling students data and compares the results of the same.

General Terms

Data Mining, Machine Learning.

Keywords

Incremental learning, education system, ensemble, voting scheme..

1. INTRODUCTION

The data in the education system may be generated online or batch wise. The student's data in the online education system or distance based education system is not only dynamic but also generates real time, so there is a need to handle the data instance wise. In batch wise data or yearly data of a students, there is a need of batch based incremental learning system, which can handle the student's data batch wise. The idea is that, when the new data introduces with the system, there should not be need to train the system from scratch, the system should update itself, without forgetting the previously acquired knowledge and without referring to the previously leaned data, it should learn new knowledge from the new data[1].

In the literature, there are many classifiers which can handle the instance based data incrementally, some of them are naïve bayes updatable, K star algorithm, Nearest Neighborhood, Winnow, and Regression based learner and locally weighted learning algorithms. All these algorithms are available in weka. For the instance based learning, these algorithm can be combined together to get the good classification result. The strategy used in [2] can be used for ensembling of these algorithm for different applications. Mainly there are three types of approaches or various ways of using supervised algorithms for instance based or batch based incremental learning. Following are the approaches studied in the literature.

1. Instance- based incremental learning
2. Ensembling of instance based algorithms
3. Batch based incremental learning algorithms, only one base classifier will be used.

In the first approach, the algorithm which itself can handle the incremental data, called as self updatable algorithms, in which the algorithm update itself instance wise. The advantage is that, there will be only single instance in the memory as algorithm will handle the data instance wise [3]-[4]. In the second approach, the ensembling of these updatable algorithms can be done to get the better results [5]-[6]. In this approach also, there will be single instance in memory but the advantage over the first approach is that, it will take the voting [7]-[8] of several instance based algorithm before predicting the final class of the instance. In the third approach, any supervised classifier can be used as a base classifier [9]-[14]. There are mainly three ways to apply the base classifier, to the subset of dataset.

1. Every time some random samples can be taken from the dataset and the base classifier will be applied, this can be used when the data set is small.
2. The dataset is divided in to batches, and for every batch, classifier will be applied, this method generally used, when there is a need to handle huge amount of data.
3. For every batch, different features of the dataset will be chosen and the data will be classified by using base classifier.

In the second and the third approach, because of the multiple opinions, ensembling strategy is used to combine the final decision. The output of the every classifier, called as hypothesis hereafter, can be combined by using different voting rules available in the literature[15], median rule, max rule, min rule, geometric average rule, arithmetic average method, majority voting method, weighted arithmetic average rule weighted majority voting, can be used. In this paper, all the above three approaches are used for handling students data. This paper is organized as follows, section II introduces the concept of incremental learning. The instance based and batch based incremental learning approach for students classification are proposed in the section III. Analysis of incremental learning approaches with experiments and comparisons is given in section IV. Section V gives the conclusion and the future scope of the study.

2. INCREMENTAL LEARNING

A practical approach for learning from new data is nothing but, discarding old classifier and retraining the new classifier, with all the data. This type of approach is having the problem of cataclysmic forgetting. It is not a desirable approach as retraining is involved which is financially costly, and requires more time. For the instance based and batch based approach, supervised algorithms are used for classification purpose. In the literature, many of the incremental learning algorithms are used for online learning applications [16], which are itself updatable.

In the batch incremental learning concept, most of the algorithms uses the adaptive principle of knowledge transformation, can be achieved using weight distribution function used in the ADABOOST [17]-[19]. The idea is that, initially the weights of all the samples in the dataset, are equal, means all samples are having equal importance for classification, when new samples arrives for the classification, the weight distribution function changes, accordingly, like examples which are hard to classify will have more weight as compared to the examples which are easy to learn, means hard examples which are difficult to classify will be given more weightage.

The difference between the ADABOOST and the incremental learning algorithm, is ADBOOST does not compute the compound hypothesis, whereas in the incremental learning concept compound hypothesis will be calculated to get the final knowledge accumulation.

3. INSTANCE BASED AND BATCH BASED APPROACH

As described in the introduction section, there are three approaches of incremental learning used in this paper. The incremental learning algorithms suitable for students data classification, which can handle the numeric attributes of the students dataset are namely Naïve bayes updatable, IB1, IBk, NNge, and Kstar. These algorithms can be ensembled using different voting rules to get the good classification result. At last, the algorithm for batch incremental learning is explained. The working of all the algorithms are given below

3.1 Naïve Bayes Updatable

It is an incremental form of Bayesian networks, as it assumes that each feature is not dependent on the remaining features. The naïve Bayes algorithm usually used for a batch learning, because when algorithm handles each training sample separately, it could not perform its operations well, described in[20]-[24]. As per the characteristics of the incremental learning algorithm, the naïve Bayes algorithm can be trained by using one pass only as per the steps below:

1. Initialize count and total=0
   a. Go through all the training samples, one sample at a time
   b. b.Each training sample, t (x, y) will have its label associated with it.
   c. c.Increment the value of count, as it goes through the particular training sample.
2. The probability is calculated by dividing individual count by the set of training data samples of the similar class attribute.
3. Compute the previous probabilities p(y) as the portion of entirely training samples which are in class y.

3.2 IB1/IBk

This algorithm does not built the model, it generates prediction for a test instance at time. The particular case is classified by using majority voting of its neighbor with the case being assigned to a class most common amongst its neighborhood by using distance function (1).

$$\sqrt{\sum_{i=1}^{k}(xi-yi)^2} \quad (1)$$

IBk implements kNN. It uses normalized distances for all attributes on different scales have the same impact on the distance function. The number of neighbor's returns from it may be more than k, if there are ties in the neighbors. The neighbors are voted to form a final classification [25].

3.3 NNge

It is a non-nested generalized exemplars. This algorithm is based on IB1, IBk and kNN algorithms. In the instance based learning, the classification time required is more, as there will be one instance at a time in memory, so the generalized exemplars can be the solution to deal with this. Generalized exemplars are the one which is representation of more than one of the actual instances in the training set. In NNge, the generalization is formed by, a new examples are added in the database each time by joining it to its nearest neighbor of the same class [26]-[27].

Algorithm:

1. For each exemplar E in the training data, find the hyper rectangle Hk which is closest to Ej

2. If distance between Hk and Ej is zero then

a. If class of Hk and Ej are note equal, split Hk with E, else H'= extend Hk with Ej

3. If H' overlaps conflicting hyper rectangle the and Ej as a non-generalized exemplar

a. Else Hk= H'

Hyper rectangle H is given by equation (2)

$$D\ (E,H) = \sqrt{\sum_{i=1}^{n}\left(W\,\frac{d(\mathrm{Ei},\mathrm{Hi})}{\mathrm{Ei}^{max}-\mathrm{Ei}^{min}}\right)^2} \quad (2)$$

3.4 Kstar

It is a sample based learner, where the test sample case is decided by using the class label of training samples based on some kind of similarity function. It uses equation (3), called as entropy based distance function, based on the probability of transforming one instance in to another by randomly choosing between all possible transformations and turns out to be much better than Euclidean distance for classification [28].

$$K^{*}=(y_{i,},x)=lnP^{*}(y_{i,},x) \quad (3)$$

4. ENSEMBLE OF INCREMENTAL LEARNING ALGORITHM

The combinations of the above mentioned algorithms are tried to get the maximum accuracy for the student's classification. The hypothesis of the different classifiers are combined by using different voting rules.

Two algorithm namely Naïve bayes and Kstar are ensemble using majority voting rule to get the final hypothesis. The voting rules namely Geometric average rule, Arithmetic average rule, Majority Voting Rule, Min rule, Max rule are used for combining the hypothesis. The first three rules are given in equation (4) to (6), for min and max rule, minimum and maximum value of the final hypothesis will be chosen according to the voting rules (7) and (8) respectively.

$$x_t \rightarrow yj \text{ satisfy max}_{yj} \frac{1}{L}\prod_{i=i}^{L} Pi\left(\frac{y_j}{x_t}\right) \quad (4)$$

$$x_t \rightarrow yj \text{ satisfy } \max_{y_i} \prod_{i=i}^{L} Pi\left(\frac{y_j}{x_t}\right) \quad (5)$$

$x_t \rightarrow y_j$ satisfy $\max_{y_i} \sum_{i=1}^{L} \Delta\left(\frac{y_j}{x_t}\right), \Delta\, i\left(\frac{y_j}{x_t}\right) = 1; if$ ht $x_t \rightarrow y_j$ $0; otherwise$ (6)

5. BATCH BASED INCREMTNAL LEANRING ALGORITH

In the batch based incremental learning, the dataset is divided into batches, some k classifiers are trained on every batch, hypothesis of one batch is given to next batch for the purpose of knowledge transformation and finally composite hypothesis is calculated to get the final voting.

Algorithm:

Input: Dataset is divided into equal subsets, D1,….Dk

Base classifier CART is used

Ck- no. of classifiers/no. of iterations

Learning Procedure:

Do for k=1 to k

Initialize weight with equal distribution, as nothing is learned yet, otherwise

Do for t=1 to Ck

1. Set Pt=W. $/ \sum_{i=1}^{m} wt(i)$. So that Pt is a weight distsribution
2. Select Training set, TR and testing set,TE from the first subset Dt
3. Train training set, TE of Dt with CART and get the hypothesis ht
4. Use the hypothesis ht to calculate the error for testing data TE.
5. Normalize the error, using $\beta = \frac{\alpha}{(1-\alpha)}$
6. Call the weighted majority voting rule, eq (6) to get the compound hypothesis, calculate the composite error Et.

$$x_t \rightarrow y_j \text{ satisfy } \max_{y_j}\{\min Pi(\frac{y_j}{x_t})\}$$

$$x_t \rightarrow y_j \text{ satisfy } \max_{y_j}\{\max Pi(\frac{y_j}{x_t})\}$$

7. If Et>0.5 discard the composite hypothesis and go to step 2.
8. Normalize the composite error
9. Update the weight, so that examples hard to classify will get more importance.

Output: Final hypothesis is calculated by using majority voting rule given in equation (6).

6. ANALYSIS OF INCREMENTAL LEARNING APPROACHES

There are total three experiments done with the approaches discussed in section I. The first dataset of students is created by conducting test on student, we named it as Indian Students dataset and other is Turkey student's dataset. The Indian Students dataset attribute description is given in Table I. These three experiments have been done on the datasets mentioned in Table II. Table III gives the performance of an instance based classifiers on both the students' dataset.

TABLE I Attribute of the dataset.

Name of the Dataset	No. of Instances	No. of Attributes	No. of Classes
Students Data	250	10	7
Turkey Students Data	2743	34	3

TABLE II: Dataset Description

A Self Awareness	B Empathy	C Self-Motivation	D Emotional Stability	E Managing Relations	F Integrity	G Self Development	H Value Orientation	I Commitment	J Altruistic Behavior
11and Above	15 and above	18 and above	11 and above	12 and above	8 and above	6 and above	6 and above	6 and above	6 and above
4 to 10	7 to 14	9 to 17	4 to 10	5 to 11	4 to 7	2 to 5	2 to 5	2 to 5	2 to 5
3 and below	6 and below	8 and below	3 and below	4 and below	3 and below	1 and below	1 and below	1 and below	1 and below

in the first experiment, the performance of an instance based classifiers are calculated for both the datasets. It has been found that the Naïve Bayes updatable classifier performs well on students data as compared to other instance based learning algorithms. The results are shown in TABLE III. In the second experimentations all the permutations and combinations of the algorithms have been done with different voting schemes and it is observed that the ensemble of Naïve Bayes and Kstar gives highest performance with majority voting rule. The results of the same are shown in TABLE IV and V.

In the third experiment, the student's data fed to the system batch wise. For the simulation, the dataset is divided into batches, for every batch the data taken for training and testing purpose.

TABLE III: Performance of instance based classifiers on student's dataset

Name of the classifier	Classification % (Students data)	Classification % (Turkey Dataset)
Naïve Bayes	89.6	86.65
IB1	87.6	88.18
IBk	89.2	88.18
Kstar	89.2	99.52
NNge	88.8	99.92

TABLE IV: Result of Ensemble of two incremental learning classifiers for Students dataset

Ensemble of classifiers	Classification rate by using different voting rules				
	Geometric average rule	Arithmetic average rule	Majority Voting Rule	Min Probability	Max Probability
NB+IB1	87.6	87.6	90	87.6	87.6
NB+IBk	88	87.6	91.6	88	88
NB+Kstar	90	89.6	92	89.6	90
NB+NNge	88.4	88.8	90	88.8	88.4
IB1+IBk	87.6	87.6	90	87.6	87.6
IB+Kstar	87.6	87.6	90	87.6	87.6
KB1+NNge	87.6	84	89.6	84	87.6
IBk+Kstar	89.6	89.2	89.6	89.2	89.2
IBk+NNge	88.8	88.8	89.6	88.8	88.8
Kstar+NNge	88.8	88.8	90	88.8	88.8

TABLE V: Result of Ensemble of two incremental learning classifiers for Turkey Students dataset

Ensemble of classifiers	Classification rate by using different voting rules				
	Geometric average rule	Arithmetic average rule	Majority Voting Rule	Min Probability	Max Probability
NB+IB1	88.18	88.18	86.87	88.18	88.18
NB+IBk	87.71	88.73	86.87	89.39	87.56
NB+Kstar	97.01	98.14	92.78	97.48	96.93
NB+NNge	99.92	99.92	92	99.92	99.92
IB1+IBk	88.18	88.18	88.18	88.18	88.18
IB+Kstar	88.18	88.18	93.4	88.18	88.18
KB1+NNge	94.34	88.11	93.54	88.11	94.34
IBk+Kstar	96	96.6	93.4	95.88	96.2
IBk+NNge	99.92	99.92	99.92	99.92	99.92
Kstar+NNge	99.92	99.92	99.7	99.92	99.92

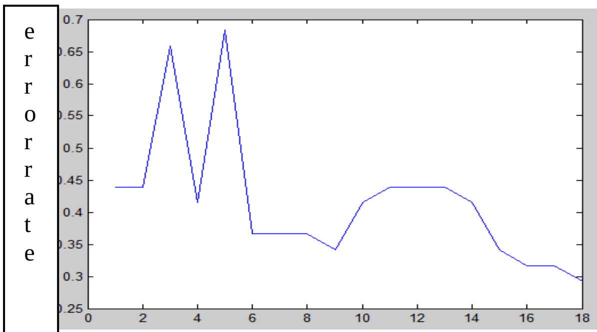

Fig. 1 Performance of a batch based incremental learning algorithm

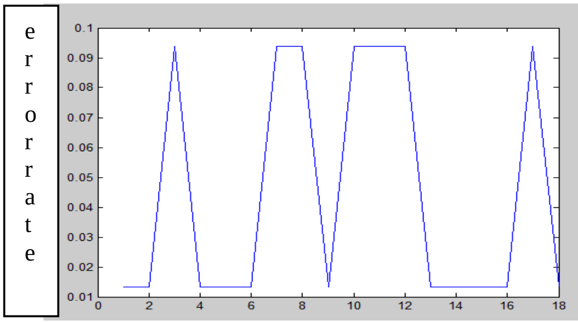

Fig. 2 Performance of a batch based incremental learning for students dataset

7. CONCLUSION AND FUTURE SCOPE

This paper aims to study and experiment the instance-based and batch based approaches for incremental learning. When the amount of students data in the database grows and so the knowledge taken out from these data need to be updated continuously, the proposed approaches are used to handle the students data instance-wise or batch-wise. All the experimental results shows that, all the three approaches of incremental learning are applicable for the student's classification problem. Use of these approaches of incremental learning algorithm can be used for any kind of stream data. These approaches can be used depending on the application.

There is a scope of inventing new weight distribution function which can be used for selecting samples for batch learning. The researchers can work on new class detection in the incremental data.

8. REFERENCES

[1] Robi Polikar, Lalita Udpa, Satish S. Udpa, Vasant Honavar, “Learn++: An incremental learning algorithm for supervised neural networks”, IEEE Transactions on systems, Man, and Cybernetics-PART C: Applications and Reviews, Vol. 31, No. 4, November 2001.

[2] S. B. Kotsiants, K. Patriacheas, M. Xenos, “A combinational incremental ensemble of classifiers as a technique for predicting students performance in distance education”, Knowledge based systems, pp. 529-535, vol. 23, 2010.

[3] N. Oza, “Online ensemble learning,” Ph.D. dissertation, Dept. Comput. Sci., Univ. California, Berkeley, 2001

[4] Seichi Ozawa, Shaoning Pang, and Nikola Kasabov, “Incremental Learning of Chunk Data for Online Pattern Classification Systems”, IEEE Transaction on Neural Networks , pp. 1045-9227 , 2008

[5] M. Muhlbaier and R. Polikar, “An Ensemble Approach for Incremental Learning in Nonstationary Environments,” Proc. Seventh Int’l Conf. Multiple Classifier Systems, pp. 490-500, 2007.

[6] S. Kotsiants, “ An icremental ensemble of classifiers”, Artof Intell Rev, pp. 249-266, Vol. 36, 2011.

[7] Kittler, Josef, et al. "On combining classifiers." IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 226-239, 1998.

[8] N. Littlestone and M.Warmuth, “Weighted majority algorithm,” Inform. Comput., pp. 212–261, vol. 108, 1994.

[9] B.V. Dasarathy and B. V. Sheela, “Composite classifier system design: concept and methodology,” proceeding of the IEEE, vol. 67, no.5, pp. 708-713, 1979.

[10] K. Woods, W. P. J. Kegelmeyer, and K. Bonwyer, “Combination of multiple classifiers using local accuracy estimates,”IEEE Trans. Pattern annal. Mach. Intell.,vol.19, no. 4,pp. 405-410,Apr.1997.

[11] L. I. Kuncheva, “Classifier ensembles for changing environments,” in Multiple Classifier Systems, vol. 3077. New York: Springer-Verlag, 2004, pp. 1–15.

[12] 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.

[13] H. He and S. Chen, “IMORL: Incremental Multiple-Object Recognition and Localization,” IEEE Trans. Neural Networks, vol. 19, no. 10, pp. 1727-1738, Oct. 2008

[14] Yang Hai, college of of Science, Minzu university of China, Wei He, College of Science, Minzu university of China, " Incremental learing algorithm for SVM based on voting principle", International journal of Information Processing and Management, Vol. 2, Number 2, April 2011.

[15] Haibo He, Yuan Cao, “SSC: A Classsifier combinaion method based on signal strength”, IEEE Transactions on Neural Networks and Learning systems, Vol 23, No. 7, July 2012.

[16] Jesse Read, Albert Bifet, Bernhard Pfahringer, Geoff Holmes, “Batch Incremental versus Instance-incremental learning in Dynamic and evolving data,

[17] Y. Freund and R. E. Schapire, “Decision-theoretic gneralization of online learning and an appliation to boosting,” Journal of Computer and Systme Sciences, vol. 55, no.1, pp. 119-139, 1997.

[18] R. R. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, “Boosting the Margin: A new explanation for the effectivenesss of voting methods,” Annals of Statistics, vol. 26, no. 5, pp. 1651-1686, 1998.

[19] N. C. Oza, “AveBoost2:Boosting for Noisy Data, “5th Int. Workshop on multiple classifier systems, in Lecture Notes in Computer Science, vol. 3077, pp. 31-40. F. Roli, J. Kittler and T. Windeatt, Eds. Springer, 2004.

[20] StijnViaene, Richard A. Derrig, and Guido Dedene, “A Case Study of Applying BoostingNaive Bayes to Claim Fraud Diagnosis” ,Actions On Knowledge and Data Engineering, Vol. 16, No. 5, May 2004, 612-620.

[21] BojanMihaljevic,Pedro Larrañaga, Concha Bielza, “Augmented Semi-naive Bayes Classifier” ,IEEE Transactions on Systems, Man and Cypernetics-PartB: Cybernetics, Vol. 36, No. 5, Oct 2006, 1149-116.

[22] V. Robles,P. Larrañaga, J. M. Pria, E. Menasalvas, M. S. Perez, “ Interval Estimation Naïve Bayes”,Advanced Data Mining and Applications”, Vol 4632, 2007, pp 134-145

[23] Liangxiao Jiang,Dianhong Wang, ZhihuaCai, Xuesong Yan, “Survey of Improving Naive Bayes for Classification”,Advances in Artificial Intelligence,Vol 8109, 2013, pp 159-167

[24] Roshani Ade, Dr. P. R. Deshmukh, “Classification of students using psychometric tests with the help of incremental naïve baiyes algorithm, IJCA, Vol.89, No. 14, pp.27-31, March 2014.

[25] D. Aha and D. Kibler, “Instance-based learning algorithms. Machine Learning”, vol. 6, pp. 37–66, 1991

[26] Martin Brent, “Instance based learning:Nearest neighbor with generalization”,Hamilton, New Zealand, 1995.

[27] Sylvain Roy, “Nearest Neighbour with Generalization” , Christchurch, New Zealand.

[28] Cleary, John G., and Leonard E. Trigg, “K*: An instance based learner using an entropic distance measure.” ICML,pp. 108-114, 1995.

[29] C. Giraud-Carrier, “A Note on the Utility of Incremental Learning,” Artificial Intelligence Comm., vol. 13, no. 4, pp. 215-223, 2000. Roshani Ade, Dr. P. R. Deshmukh, "An incremental ensemble of classifiers as a technique for prediction of students career choice”, “IEEE conference on network and soft computing, pp 427-430,aug 2014.

[30] S. Lange and S. Zilles, “Formal Models of Incremental Learning and Their Analysis,” Proc. Int’l Joint Conf. Neural Networks, vol. 4, pp. 2691-2696, 2003.

[31] H. He and S. Chen, “IMORL: Incremental multiple-object recognition and localization,” IEEE Trans. Neural Netw., vol. 19, no. 10, pp. 1727–1738, Oct. 2008.