Robust Classification using Twosome Classifier

Roshani Raut

Abstract: The proficient knowledge revolution with the concept class detection is a significant challenge for the incremental learning classification, here the batch of data is flowing unremittingly. The algorithm recommended in this research learns continual raw data, transform preceding knowledge to the present data without stating to the longstanding data and able to proficiently allow new concept class noticed by the classification. A major aim was knowledge transformation and the accumulation of the same with concept class detection in the incremental data flow. After the analysis of non-incremental ML approaches for classification and incremental learning algorithms in the literature, this research introduces new incremental learning algorithm, which uses twosome classifiers, in which knowledge transformation and new class detection have been done efficiently. The performance of the system is validated using simulation results on available datasets. The class wise and batch wise accuracy is calculated. The projected technique is also used to detect concept class detection.

Keywords: Incremental Learning, Ensemble Learning, Voting Scheme, Concept Class, Classifiers

I. INTRODUCTION

Incremental learning has recently involved rising courtesy in academia as well as in industry. There are basically four important properties of incremental learning namely 1. Learning innovative information from fresh data 2. Not requiring admittance to the previous data 3. Not suffering from disastrous disremembering 4. Detecting concept class announced in the fresh data. The overall notion is that, when the fresh data familiarizes to the system, it is not required to train the system from beginning, the system will modernize itself, without overlooking the formerly acquired information, without mentioning to the earlier learned data and it should learn new information from the fresh data.

A applied way for learning from the fresh data is nothing but, neglecting the old leaning model and rebuilding the new learning model, for complete data perceived thus far. This methodology has the difficulty of cataclysmic overlooking. It is not a appropriate methodology as retraining is complicated, which is commercially expensive, and requires more time. In the supervised machine learning approach, the system is built using predefined input and the knowledge generated from this will be used to predict the unseen data.The Machine Learning (ML) approach using batch incremental learning yields predictive models with satisfactory performance.

The adaptive classification has been done using twosome classifiers which uses base classifiers namely CART and SVM. It uses the classifiers MLP, SVM and C4.5 for weight distribution. It is adaptive in the sense that successive base learner are pinched in favor of those instances misclassified by previous classifiers. The proposed framework attains every property of incremental learning efficiently for any kind of classification.

This paper is organized as follows, section 2 explains background of this research which includes incremental learning algorithms and its strategies in the literature, incremental learning using ensemble of classifiers and the concept class in incremental learning system. Section 3 expounds the proposed algorithm for incremental learning, section 4 gives the experimentations and outcomes of the proposed algorithm for real world data. Finally section 5concludes the work and gives its forthcoming enhancement.

II. BACKGROUND

In this research, incremental learning using supervised learning methods has been considered.

A. Incremental Learning

Incremental learning approaches in the literature can be divided in to three types, first is example centered incremental learning, second is ensembling of example based algorithms and third is batch based incremental learning [1]-[5]. In instance based learning also called as online learning [6]-[7], where the algorithm update itself instance wise. There will be simply one instance in the remembrance as algorithm handles the data example wise [8]. In the further method, number of instance based algorithms can be combined to reach to the final conclusion [9]. In this there are basically three methods to use base classifier in the framework, according to the size of the dataset. (a) At each iteration, some random samples are selected from the original dataset and the same classifier can be used for each chunk of data and then the voting can be applied. This approach is used when the data set is minor. (b) The data set is separated into equal number of batches and base classifier will be applied to the all the batches and the output will be combining using voting algorithms. This approach is used when data set is huge. (c) For each batch of data, different features are used in each chunk, then the base classifiers will be used to classify this data and then the voting will be applied. In the group based incremental learning, the dataset is separated into separate bunches, some n classifiers are trained for every bunch, hypothesis of a bunch is given to the next set in the view of information revolution and finally the amalgamated hypothesis is considered to acquire the final hypothesis[5]-[10]. In the proposed approach, batch based incremental learning has been used.

Revised Manuscript Received on January 10, 2020.
***** Correspondence Author
Dr. Roshani Raut*, Department of Computer Engineering, Vishwakarma Institute of Information Technology, SPPU, Pune, India. E-mail: rosh513@gmail.com

B. Ensemble of Classifiers

In ensembling, the numerous hypothesis, which chains the ultimate resolution, are coupled for ultimate decision. The precision of the system can be model can be enhanced using ensemble learning scheme and a strong model can be constructed, using ensemble learning conception instead single hypothesis model. The experimentations in the ensemble learning is intensifying rapidly, with many inspired philosophies, the work includes the combined classifier systems [11]-[13], experts mixture [14], stacked generalization approach, combining of multiple classifiers and capacity of voting . The concept behind the ensemble learning is nothing but, the process of consulting several authorities, before reaching to the final decision. In this research, operating the training dataset construction strategy has been used for forming an ensemble. In which, multiple datasets are created by resampling the original data according to some sampling distribution.

C. Concept Class in Incremental Learning

Incremental learning algorithm has the ability to learn from newly arriving data. It constructs series of hypothesis for the training samples coming in sequences, in which the hypothesis represents the knowledge represented by the data seen thus far and it continuously hinge on previous hypothesis and the fresh training data [15]-[16]. Incremental learning algorithm learns the novel information and it upholds the earlier learnt information without retrieving the previously perceived data. The concept class detection is an significant features of incremental learning. All those concepts have been used in the proposed algorithm.

III. METHODS

The suggested incremental learning algorithm uses the pair of classifiers concept, in which primary classifier was used as a base classifier and the another classifier was used for weight distribution function, which was used to transform the knowledge from previous data chunk to the existing data chunk, without discussing to the original data and also detects concept class efficiently.

Algorithm:

Input: D_t , contains n instances, (x_i, y_i), $i = 1..n$.
x_i – Instances in n dimensionalfeaturespace.
$y_i - Y = \{1, \dots, C\}$
$DF_t = [W_1^t, W_2^t, \dots, W_n^t]$, $\sum_i DF_t = 1$

Classification:
Step 1: Use base classifier (CART or SVM) with weight distributionDF_t, which generates the hypothesis h_t

Step 2: Weights are restructured (called as DF'$_t$) using additional classifier in the couple (MLP or SVM or C4.5) which proceeds the participation as D_t and DF_t, when the model is steadied, updated weight will be reused as a weight distribution for $D_{t+1}$with DF_t'weight distribution function.

Step 3: The inaccuracy will be computed by applying h_t on D_{t+1} with DF_t'weight distribution function

Step 4: $\beta_t = \varepsilon_t / (1- \varepsilon_t)$, $C_t = Y_t \epsilon \{C_1, \dots C_n\}$, C_t has been attained to retain the trace the classes used to train the particular classifier h_t
Step 5: Calculate confidence of the classifier $CC(i) = \frac{\sum_k \sum_{h_t(x_i)=T_c} W_t}{Z_c}$
Step 6:

$$Z_c = \sum_k \sum_{T_c \in C_t} W_t$$

Step 7: Step 8: The instance distribution weights are updated using

$$W_t(x_i) = \begin{cases} W_t\big(1 - CC(x_i)\big), t{:}\, T_c \notin C_t \\ W_t\ t{:}\, T_c \epsilon C_t \end{cases}$$

Where, C_t – Class labels on which h_t have been trained.
$CC(x_i)$– Reliance of classifiers qualified on class T_c for selecting class T_c for
examplex_i.Normalize weight dissemination. Misplaced instances weight will be unchanged.
Recap the step3 to Step7 for each data portion
Step 8: Get the concluding hypothesis using weighted majority voting rule.

$x_t \rightarrow y_j$ satisfy $\max_{y_j} \sum_{i=1}^{L} \omega_i \Delta_i(y_j|x_t)$,

Where, ω_i is a weight coefficient for hypothesis

$h_i{:}\omega_i \geq 0$, and $\sum_{i=1}^{L} \omega_i = 1$

In this, dataD_1,D_2, ...,D_{t-1},D_t are nothing but the data coming at time t_1, t_2, $t_{t-1}$and t respectively. DF_1, DF_2,…, DF_{t-1} and DF_t are nothing but the weight distribution function associated with it, h_t, h_2,…, h_{t-1}, h_t are nothing but the hypothesis generated from these respective datasets. All the hypothesis are collected by using majority voting rule. Initially, all the data have equal weights, which has a meaning that nothing has been learned yet. The algorithm has a probability distribution. called as weight distribution over the training set. The algorithm run for a number of iterations where in each iteration, a subset of a training set was drawn without replacement according to weight distribution function. The classifiers MLP or SVM or C4.5 was used as a weight dissemination function. The classifiers, CART or SVM was used as a primary classifier as it performs well, when used as a primary classifier, as compared to other classifiers [1]-[5].

IV. EXPERIMENTS AND RESULTS

In the proposed algorithm, CART and SVM have been used as a primary classifier and MLP, SVM, C4.5 were used as a weight dissemination function. The following versions of the algorithms have been formed to evaluate the proposed structure.

1) CART.MLP2) SVM.MLP 3) CART.SVM
4) SVM.SVM 5) CART.C4.56) SVM.C4.5

A. Batch wise Accuracy

For the model objective, the dataset [17] was separated into identical ten sets. Each set signifies the data approaching at time $\boldsymbol{t}, \boldsymbol{t}+\mathbf{1}, \boldsymbol{t}+\ \mathbf{2}$ and so on. One set out of ten subsets, was nominated for the testing sequentially and the rest nine sets were specified to the proposed algorithm shown in "Fig. 1". The same was repeated for all ten data chunks. For Learn++ and Learn++. NC, the 10 chunks are nothing but, the dataset initially divided into equal five chunks, and then from each chunk the samples were selected two times sequentially. In the He's methods the distance based approach was used for the mapping of data [4]. "Fig. 1 (a)" and "Fig. 1 (b)" shows the batchwise accuracy of Waveform and OCR Dataset respectively

B. Class wise Accuracy

For this experiment, the dataset was divided into ten chunks, and the sampling with replacement strategy was used. For this any data chunk out of ten was selected for testing purpose and the rest nine were specified to proposed algorithm. The overall accuracy and the class wise precision of Waveform dataset is shown in Table I. Class wise, highest accuracies are highlighted, which shows the percentage of correctly classified samples of a particular class. The last column of Table I shows the overall accuracy, which shows that CART.MLP dominates the other methods

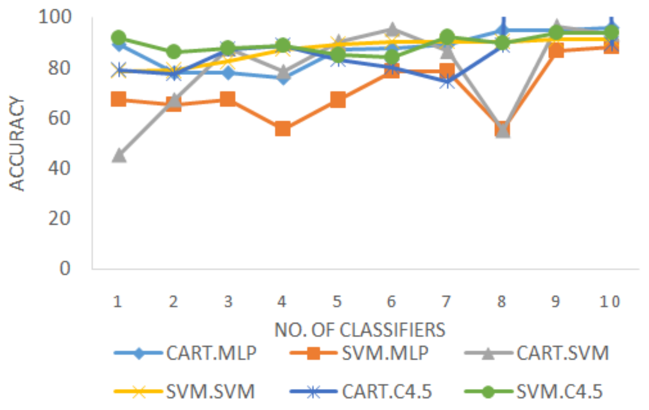

Fig. 1 "(a)" Batch wise accuracy of Waveform Dataset

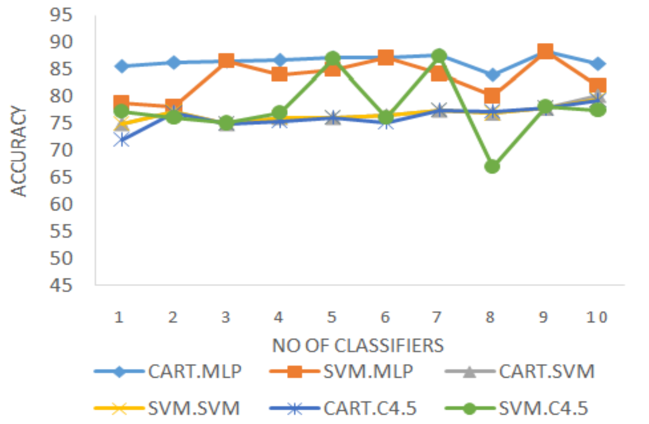

Fig. 1 "(b)" Batch wise accuracy of OCR Dataset

Table I: Classwise accuracy of Waveform Concept Class Detection

Methods	Class-wise Prediction Accuracy			
	Class 1	Class 2	Class 3	Overall Accuracy
Learn++[2]	0.8132	0.8512	0.8429	0.8356
Learn++.NC[1]	0.7937	0.856	0.8628	0.8372
He[4]	0.526	**0.9359**	**0.9444**	0.8
He[9]	0.5266	**0.9359**	**0.9444**	0.8002
Proposed Algorithm				
CART.MLP	**0.8398**	0.8784	0.8804	**0.866**
SVM.MLP	0.7843	0.8457	0.8508	0.8266
CART.SVM	0.5266	**0.9359**	**0.9444**	0.8002
SVM.SVM	0.7181	0.8258	0.8375	0.7932
CART.C4.5	0.7317	0.8252	0.8205	0.792
SVM.C4.5	0.7394	0.784	0.8	0.7742

Identifying a different class which originates with fresh data is one of the feature of incremental learning. The system should able to identify the fresh class proficiently when the new chunk of data is having the samples having new class, called as concept class. For the simulation of concept class identification the datasets are divided class wise. "Fig. 2" shows the concept class detection for both the dataset. In waveform dataset, the concept class has been detected with star mark, as the data of class three introduced in the chunk five. The single and multiple concept class has been detected in the OCR dataset in the "Fig. 2", with big square and the star mark respectively. In the chunk number three, the samples of class three were detected and in the fifth chunk samples of multiple classes i.e. class 4 and class 5 were detected. It has been observed Dataset Concept Class Detection.

Detecting a new class which derives with different data is individualproperty of incremental learning. The system should able to identify the different class efficiently when the new chunk of data is having the samples having new class, called as concept class.

For the simulation of concept class detection the datasets are divided class wise. "Fig. 2" shows the concept class detection for both the dataset. In waveform dataset, the concept class has been detected with star mark, as the data of class three introduced in the chunk five. The single and multiple concept class has been detected in the OCR dataset in the "Fig. 2", with big square and the star mark respectively. In the chunk number three, the samples of class three were detected and in the fifth chunk samples of multiple classes i.e. class 4 and class 5 were detected. It has been observed that the proposed method requires only one classifier to detect the new concept class.

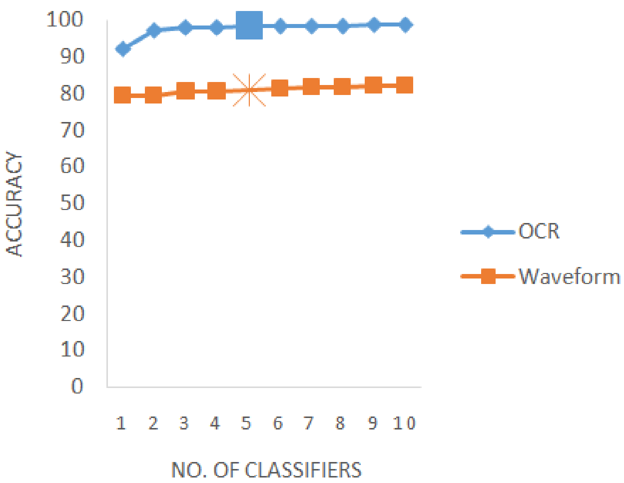

Fig. 2. Single and multiple concept class detection

V. CONCLUSION

It has been also perceived that, the proposed algorithm provides flexibility of using any classifier as a primary classifier as well for weight dissemination function. The system was validated with the number of input datasets and the analysis of the same has been done efficiently. The comparative analysis of the same proves the efficiency of the proposed method, compared to the methods in the literature for learning new information from fresh data and finding of new classes, presents with the new data.

Further research could also enhance focus on the concept drift. The imbalance data in incremental learning can also look as a challenging problem, and can be done with some improvements in the proposed algorithm

REFERENCES

1. RoozbenRazavi-Far, PieriBaraldi, and Enrico Zio, "Dynamic weighting ensembles for incremental learning and diagnosing new concept class faults in nuclear power system," IEEE Trans. On Nuclear science 59(2), 2520-2530, (2012)
2. RobiPolikar, LalitaUdpa, 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*, 31(4) ,(2001)
3. 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 Networks*, 21(1), 152–168, (2009)
4. H. He and S. Chen, "IMORL: Incremental multiple object recognition and localization," *IEEE Trans. Neural Netw.*, 19(10), 1727-1738, (2008)
5. Haibo He, "Self-Adaptive Systems for Machine Intelligence," First Edition, *Wiley*, 2011
6. Kotsiants S, Patriacheas K., Xenos M., "A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education," *ElsevierKnowledge based systems*, 23,529-535, (2010).
7. Kotsiants S, "An incremental ensemble of classifiers," *Springer science + business media B. V.*, Springer, 36, 249-266, (2011).
8. Gang Wang, Jian Ma, "A hybrid approach for enterprise credit risk assessment based on Support Vector Machine," *Expert Systems with Applications*, vol. 39, pp. 5352-5331, 2012.
9. ZekiErdem, RobiPolikar, FikretGurgen an Nejat Yumusak, "Ensemble of SVMs for Incremetnal Learning," *Springer-Verlag Berlin Heidelberg*, 246-256, (2005).
10. Ade, Roshani, and Prashant Deshmukh. "Efficient knowledge transformation for incremental learning and detection of new concept class in students classification system." Information Systems Design and Intelligent Applications. Springer India, 757-766, (2015).
11. Roshani Ade, "Students performance prediction using hybrid classifier technique in incremental learning", International Journal of Business Intelligence and Data Mining, Inderscience, Vol 15(2), May, (2019)
12. Ade Roshani, P. R. Deshmukh, "Classification of students by using an incremental ensemble of classifiers", *Reliability Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions) 2014 3rd International Conference on. IEEE*, 2014.
13. C. Ji and S. Ma, "Combination of weak classifiers," *IEEE Trans. Neural Networks.*, 8(1), 32–42, (1997).
14. R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Comput.* 3,9–87, (1991).
15. N. C. Oza, . F. Roli, J. Kittler and T. Windeatt, Eds, "AveBoost2:Boosting for Noisy Data," *Multiple classifier systems, in Lecture Notes in Computer Science*, Springer, 3077, 31-40, (2004).
16. Ade, Roshani, and P. R. Deshmukh. "Instance-based vs Batch-based Incremental Learning Approach for Students Classification." International Journal of Computer Applications 106(3), 37-41,(2014).
17. Gunduz, G. &Fokoue, E. (2013). UCI Machine Learning Repository, Available http://archive.ics.uci.edu/ml/datasets/ Irvine, CA: University of California, School of Information and Computer Science.

AUTHORS PROFILE

Dr. Roshani Raut received her BE degree in Computer Science and Engineering in 2002 and ME from Pune University in 2008. She has received Ph.D. Degree in Computer Science and Engineering from Amravati University in Feb-2016.. Currently, she is working as an Associate Professor in the department of Computer Engineering of Vishwakarma Institute of Information Technology, Pune. Her research interest is in field of Data Mining, Machine Learning, Deep Learning, Incremental learning etc. She is a member of CSI and ISTE. She has also availed the research grant from BCUD, Pune University.She has several research publications and patents on her name.