Decision-Making in Multi-Perspective Environment with Incremental Learning

Prachi Joshi and Parag Kulkarni

Abstract—**In many applications that require decision making, dependency is on the data obtained from various sources. Each source of information provides a different perspective to the problem domain. With the existing data that is available and new perspective that evolves at later stages, there is need to learn incrementally. In this paper we recognize the similarity between the new perspective and the existing knowledge based on their pattern. The proposed work, referred as 'Multi-Perspective based Incremental Learning' [MPIL], effectively modifies the knowledge with available perspectives, understands the patterns of the data to determine the update and at the same time maintains them for reuse. With application to education sector, experiments show that the proposed approach exhibit better decision-making capacity in a multi-perspective environment.**

Index Terms—**Clustering, decision-making, incremental, pattern recognition.**

I. Introduction

Considering the large availability of data and most of it been unlabeled, it is essential to learn and update the clusters at every step. The formulation of the clusters or the classes needs to take place with accuracy along with knowldege amassing. The cluster boundaries, the way it is formed, the impact of addition of new data are parameters that play a vital role for decision making. There are many approaches for clustering, but most of them are restricted with capacity to handle new data[1].

Learning and using the knowledge is a key factor that has to be focused. Learning where entire clustering is to be redone with every new unlabeled data is quiet expensive in terms of space and time. Exploitation of the existing data and knowledge updation along with the new data to come up with better decisions is required at next level of learning.

Supervised learning methods are bound by the availability of the training sets and unsupervised as said with the unlabeled. There is a need to develop learning that is active at all the time and helps in gaining the knowledge that can be used further. The data availability cannot be at once and it can be evolved at any point of time.

To achieve this next level of learning, it is necessary that incremental learning takes place. The incremental algorithms handle the new data by assigning them to clusters/classes in incremental way. It is not just classifying or assigning class but to evolve the knowledge base. Each decision of the class assignment may affect the pervious decisions or may put forth a need to generate new class.Current work in incremental leraning is directed towards forecasting and decision making in various applications starting from the

The proposed approach of MPIL performs the task of incremental clustering in a multi-perspective environment.

It is based on the concept that for a partical problem there are different views. The percept or the view is mapped to be a pattern. A novice method for incremental update with closeness value is introduced for the pattern analysis. The learning process helps in identifying new domains that will lead to effective estimation for incremental learning.

II. Related Work

In recent years, a number of clustering approaches have been proposed. Methods that perform clustering are dependent on availability of entire data set. The statistical methods like k-means and others [1] [2] rely on closed data set that is available at once. For new data categorization, though the methods can have re-computing but they face a major challenge of time complexity [3], [4], [5].

Other hierarchical techniques [6] tend to consider the data points that are local neighbors and the overall size of the clusters is unseen.

Methods with supervised techniques lack the ability to learn from the unlabeled data and hence learning in semi-supervised way is gaining interest in terms for theory and practise. Accuracy is also maintained taking in account the labeled and unlabeled data [7], [8].

In order to have better decisions and learning from new data, incremental clustering [9], [10] techniques have been put forth to deal with the continuous input. Selective incremental update of clusters is what now been looked at. It needs to keep intact the formed clusters while incrementally accommodating the new data. Formation of the clusters with respect to the underlying pattern [11] of the data is now considered during the process of learning incrementally [12].

Recently the incremental learning methods are focusing on the multi-sensor data with approach of ensemble based learning [13], [14]. Various dynamic weighted schemes are been employed to carry out the incrmental learning with ensemble based approach where the previous knowledge is also preserved [15].

Boosting methods to learn incrementally also have taken up researchers to work in the areas of visual tracking and object detection.[16]

Manuscript received March 4, 2012; revised April 3, 2012.

P. Joshi is Research Scholar at College of Engineering Pune and is working with MIT College of Engineering, Pune (e-mail: prachi.joshi@mit.edu.in).

P. Kulkarni was Chief Scientist with Capsilon India Pvt. Ltd. and currently is the Chief Scientist and founder of Eklat Research, Pune (e-mail: paragakulkarni@yahoo.com).

Methods employing online incremental learning have to take into account the new knowledge while the previous one is retained. [17]. Some incremental neural network based algorithms take advantage of learning with new informationlike the Self Organising Maps and others.[18]

To get the best in terms of results and accuracy, semi-supervised incremental learning is now been looked at. With the existing semi-supervised methods, new class evolution is considered. Gaussian distribution and Neural Network based methods are now being used for incremental learning [19], [20].

The need for incremental learning today is with respect to the applications in terms of decision making and forecasting. Evolving a new domain of class during the process of learning has gained importance with the learning to be selective at the same time. Selective nature of the learning and the approach to be adaptive is equally essential in the entire process. Reducing the dependency on the traning data sets and learning with new data sets is also a required factor.

While considering learning in multi-perspective environment, the key ides is to exploit and use all the available perspectives[21] for decision-making. With every new perspective available

We propose a MPIL approach that considers different perspectives for a problem domain. The perspectives are used for learning with incremental update to evolve with better decisions. The approach is based on identification of series with similar behavior and in accordance updation of cluster takes place. Learning to maintain the clusters effectively and taking decisions with every perspective available is considered during the process. It does not require the scanning of the entire data and it processes adequately the new perspective without the need to reconsider the whole system. This update helps in decision making as well as building a strong knowldege base.

III. Multi-Perspective Environment

The proposed work considers sources with different perspectives and hence is applied in a multi-perspective environment.

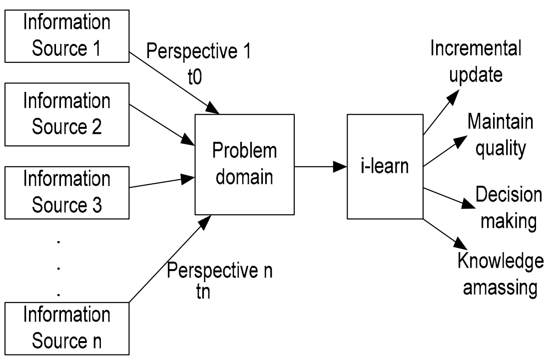

Fig. 1. Multi-perspective environment.

As shown in Figure 1, for a specific problem domain, each and every information source will have its own perspective. It is not unusual that the data obtained may provide complementary information.

There is always a need to consider the information available from different perspectives, as the information or the knowledge gained from a single perspective may not be sufficient or may be biased.

It is necessary to combine all the available perspectives and develop a learning that will adapt with each new available perspective.

With proper fusion of the data, it can lead to improving the accuracy or helping to build better systems with better decisions. The 'i-learn' module is the one that carries out the task of incremental learning in this multi perspective environment to get the decisions. It is the i-learn module that is responsible for learning incrementally with knowledge update.

When operating in a multi-perspective environment, it is crucial that the knowledge acquired from all the sources of information be utilized to full extent. The knowledge should be selectively updated to avoid any loss of some critial information that would impact the decisions.

IV. Proposed Work

As discussed, the basic idea behind the work is to have incremental learning in a multi-perspective environment.

For the same, the pattern of the data coming from each source is analyzed. In MPIL we have considered semi-supervised learning as the perspectives available could be already labeled. MPIL algorithm handles the issue of incremental clustering based on the nature of the data. The decision related to learning from a newly available perspective and updation is based on its 'closeness value' with the existing clusters.

A. Proximity Value: (P Value)

Here we discuss method for calculating proximity (a new distance measure) between the patterns that will be applied to learn from multiple perspectives. The proximity value is explained as follows [22], [23]: -

Consider two patterns P_1 and P_2. $P_i(j)$ is point j in ith pattern. Sumall(j) is the total of parameters in the pattern.

$$Sumall\ (j) = \sum_{j=1}^{n} P1(j) + P2(j) \tag{1}$$

The outcome of pattern is given by the equation.

$$Outcome = \sum_{x=1}^{n} P1(x) / \sum_{x=1}^{n} Sumall\ (x) \tag{2}$$

where the expected value now will be

$$P\exp(\ x) = Outcome\ / Sumall \tag{3}$$

The error value calculation is the difference between the expected value and the actual.

$$error = \frac{P\exp - Pactual}{\sqrt{Sumall\ * P\exp * \ (1 - P\exp)}} \tag{3}$$

Finally, the P value is calculated as

$$P = \sqrt{\sum_{x=1}^{n} (error\ (x))^2 * \sqrt{sumall\ (x)} \Big/ \sum_{x=1}^{n} \sqrt{sumall\ (x)}} \tag{4}$$

It is found that lower the P value, the pattern are perfectly matching, else there is a difference. This value has a significant impact in learning method.

B. Algorithm

Any incremental learning method in a semi-supervised environment works on the following lines. The basic steps carried out by the i-learn module are as follows:

1) Read the input data set
2) Find P between the patterns.
3) Using P value, take decisions
    a. to update the existing class/cluster or
    b. to evolve a new class.
4) Determine the P values inter/intra between the classes and clusters
5) Update the knowledge base with each new decisions
6) Continue step 1 till the end of the data.
7) Perform incremental update for every new training data or unlabeled data.

C. Updating the Class/Cluster

The cluster feature representation is very important in case of any learning. The revisions of the feature set needs to happen with every learn step and it is to be used with every new learning. The following steps are carried out in incremental update of the knowledge base.

When a new perspective is available, incremental update of the clusters takes place. Consider that we have Cluster/Class C comprising of data pattern PS1,PS2,…PSn. Let the patterns be from pattern 1 to pattern n-

$$PS_1 = para_{11}, para_{12} \cdots para_{1m}$$

$$PS_2 = para_{21}, para_{22} \cdots para_{2m}$$

and so on till

$$PS_n = para_{n1}, para_{n2} \cdots para_{nm}$$

Now if the entire details of the parameters can be saved with the feature set, the incremental learing or the revisions can happen with the P value. Let us say that the parameters stored are-

$$Para\_all = \sum_{x=1,}^{x=n} \sum_{y=1}^{y=m} para_{nm} \tag{5}$$

Let the new pattern series be

$$PS_{new} = para_{new1}, para_{new2} \cdots para_{newm}$$

The cluster/ class stores attribute/para-details which is the sum of all the parameters in the pattern. Let us say that it is stored as

$$C = att_1\_tot, att_2\_tot, \ldots att_n\_tot$$

Now, the total is calculated as:

$$total\_all = \sum_{x=1,}^{x=m} para_x \tag{6}$$

The C or the class feature is to be updated with the new pattern, that too incrementally. This is done in the following manner-

$$C_{new} = att_{new1}\_tot, att_{new2}\_tot, \ldots att_{newn}\_tot$$

Now, the attnew_total can be obtained by

$$att_{newk}\_tot = \frac{att_k\_tot * total\_all(C) + para_{new}k}{total\_all(C) + total\_all(para_{new}k)}$$

With this update, the Class feature is maintained and is used in each consecutive learning method. So, if the total of the parameters is maintained, then it becomes relatively easy to update the knowledge base with new data, where the cluster feature gets modified effectively.

V. Experiments

A. Data Sets

The proposed method of incremental learning was tested on the data sets from the UCI repository [23] along with k-means and COBWEB for the data sets of wine and iris.

TABLE I: Purity Results

Algorithm	Wine	Iris
k-means	0.889	0.83
COBWEB	0.99	0.783
MPIL	0.982	0.896

The proposed of incremental learning shows at par results when applied to the two data sets. K-means output is the average of 10 runs as it is based on the random centroid selection method.

The proposed approach works in semi-supervised environment. New class is evolved with respect to the proximity – P measure based on the pattern analysis and incremental update takes place.

B. In Multi-Perspective Environment

We present two different examples to show how the method of MPIL has proved to be effective for decision making.

To operate in a multi-perspective environment, an application to school system was considered for decision making.

Reviews from 150 students and their parents with regard to overall school system were obtained. The parameters that were considered in the learning with respect to the class of the student were: teaching, sports, competitive exams, co-curricular activities, examination schemes and infrastructure. Here with multi-perspective environment, every parent/ guardian will be having a different view on the entire operation of the system.

Fixed patterns related to overall school system were pre-determined. Since the working of the proposed method is in semi-supervised environment, the reviews were mix of labeled and unlabeled data. The categorization of the system initially was among 2 classes- best and satisfactory. With the incremental method of MPIL, there was a need to evolve with a new class. The P-factor calculation between the existing classes and the class feature made it evolve a new domain of class. The analysis of the new class that was evolved show cased that, a considerable amount of percentage was grading the system into a just satisfactory zone/ need improvement. With learning from each perspective, the feature set was updated and hence learning method itself was getting evolved with every new perspective.

The percentagewise gradation helped the school system understand the parameters affecting their system. To have an improvement to the same, reviews were conducted again at a gap of one month to see the feedback. Now the learning system of MPIL took into consideration 3 classes, learning on top of earlier one. The results were recorded and compared by the school administration with the previous ones.

Following graph shows that with the proposed approach, the school administration to could come up and provide better facilities and improve the overall rating of the school.

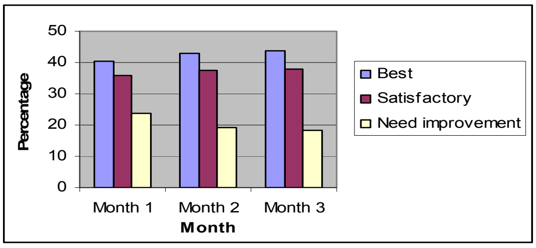

Fig. 2. Month-wise graph generated with MPIL.

In another approach of learning in multi-perspective environment in the school system, decision-making in terms of new class identification was done.

There was need to have MPIL as selection of student for quiz in science, and mathematics was required. From fifteen teachers in the school perspective/views were obtained for 5 candidates. Using the i-learn module, where predefined criteria of labeled data was existing, the learning took place. The parameters used for the learning were- Ability to solve problems, scientific approach to methods, response to question, analytical thinking, communication level and grades in mathematics and science at school level. With MPIL, it was noticed that as per the perspectives available, for one student he was best suited for some other field that was identified. The following graph shows the domains that were identified for the students.

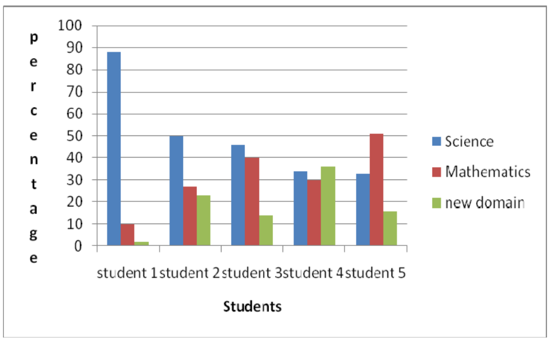

Fig. 3. Percentage-wise student selected for the domains, with new domain identification using MPIL.

It is observed that with the MPIL method, a new domainwas evolved. Based on the patterns, a new domain for the students where their points would have overruled the probability of getting selected for a science quiz or mathematics quiz was generated. It putforth that there was a need to generate or assign some new competative quiz for the student other than existing ones.

The overall system of MPIL helped the school faculties to come up with new areas and identify them thus helping them with better decisions for the same.

VI. Future Work

A novel approach for incremental learning in a multi-perspective environmtne is put forth. The main objective of MPIL is to have learning with new data observing the pattern and giving better decisions. Through experiments we have validated the performance of MPIL, and results show that it has incremental learning abilities that are worth to notice.

The method is governed by the amount of label data that is available in the learning process and it is obsedver that when the percentage of labeled data is considerably high, the accuracy in the results was better.Though the data sets used for evaluation are more biased towards clusters/classes of smaller size and still the approach needs to be investigated with larger data sets, as results are dependent on the contribution of labeled data.

Combining existing methods of supervised learning and other distance measures for pattern analysis for better accuracy and classification will be considered ahead in the research work.

Fuzzy set approach will be evaluated for the data sets. In terms of the data that lie on the boundry, there is a need to take appropriate decisions in terms of classification and it is equally essential to consider the impact of the decisions on the earlier decisions that are taken.

The scalability of the approach in terms of new dimensions added to the data will also be investigated further as a part of research.

References

[1] Y. Lui, J. Cai, J. Yin, and A. Fu, " Clustering text data streams," *Journal of Computer Science and Technology,* Jan 2008, pp 112-128.

[2] F. Camastra and A. Verri, "A novel kernel method for clustering," *IEEE Trans. on Pattern Anal. Mach. Intell*., 2005, pp 22-32.

[3] H. Becker, M. Namman, and L. Gravano, "Learning similarity metrics for event identification in social media" in. *Proc. of ACM Conf. on Web Search and Data Mining*, 2010, pp 291-300.

[4] A. Fahim, G. Saake, A. Salem, F. Torky, and M. Ramadan, "K-means for spherical clusters with large variance in sizes," in *Journal of World Academy of Science, Engineering and Technology*, 2008.

[5] P. Kulkarni and M. Kulkarni, "Advance Forecasting Methods for Resource Management and Medical Decision-making" in. *Proc. of National Conference on Logistics Management: Future Trends*, 2002.

[6] A. Geva, "Hierarchical Unsupervised Fuzzy Clustering" *IEEE Trans. on Fuzzy systems*. 1999.

[7] O. Chapelle, B. Schölkopf, and A. Zien, "Semi-Supervised Learning," MIT Press, 2006.

[8] X. Zhu, "Semi-Supervised Learning Literature Survey," Technical Report 1530, University of Wisconsin, Madison, July 2008.

[9] Y. Liu, Y. Ouyang, H. Sheng, and Z. Xiaong, "An incremental algorithm for clustering search results" in *Proc .of IEEE Conf. Signal* Image *Technology and Internet Based systems,* 2008, pp: 112 -117.

[10] X. Su, Y. Lan, R. Wan, and Y. Qin, "A fast incremental clustering algorithm".in *Proc. of Symposium on Information Processing*, 2009, pp: 175-178.

[11] C. Hsu and Y. Huang, "Incremental clustering of mixed data based on distance hierarchy" *ACM Transactions on Expert systems with Applications*, 2008.

[12] S. Ozawa, T. Tabuchi, S. Nakasaka, and A. Roy, "An autonomous incremental learning algorithm for radial basis function networks" in *Journal of Intell. Systems and Applications*, pp 179-189.2010.

[13] D. Parikh and R. Polikar, "A multiple classifier approach for multi-sensor data fusion," in *Proc. of IEEE International Conference on Information Fusion*, 2005, pp 453-460.

[14] D. Parikh, N. Stepenosky, A. Topalis, D. Green, J. Kounios, C. Clark, and R. Polikar, "Ensemble based data fusion approach for early diagnosis of Alzheimer's disease" *IEEE Conf. on Engineering in Medicine and Biology Society*, pp 83-95. 2008.

[15] X. Yang, B. Youan, and W. Lui, "Dynamic weighting ensembles for incremental learning", *Proc. Of IEEE Chinese Conference on pattern* recognition, Dec 2009, pp 1-5.

[16] H. Grabner and H. Bischof, "On-line boosting and vision", *Proc. of IEEE Conf. in ComputerVvision and Pattern Recognition*, 2006.

[17] F. Shen and O. Hasegawa, "An Incremental Network for On-line Unsupervised Classification and Topology Learning," in *Elsevier* Journal *of Neural Networks*, Vol.19, No.1, pp. 90-106, Jan 2006.

[18] T. Kohonen, editor, "Self-Organizing Maps," Springer Series in Information Sciences, Springer, Heidelberg, 1997.

[19] A. Bouchachia, M. Processenger, and H. Duman, "Semi-Supervised incremental learning," in *Proc. IEEE International Conference on Fuzzy Systems*, 2010, pages 1-6.

[20] C. Yuan and H. Haibo, "Learning from testing data: A new view of incremental semi-supervised learning," in *Proc. IEEE International* Joint Conference in Neural Networks, 2008, pp 2872-2878.

[21] C. Tekinay, M. Seck, M. Fumaralo, and A. Verbraek "A context based multi-perspective modeling and simulation framework," in *Proc. of IEEE Simulation Conf, 2010, p 479-489.*

[22] P. Joshi and P. Kulkarni, "A novel approach for clustering based on pattern analysis," *International Journal of Computer Applications*, 2011, pp 1-4.

[23] P. Joshi and P. Kulkarni, "Multi-perspective based incremental learning:" in *International Conf. on Computer and Computational Intelligence*, 2011 pp.

[24] A. Frank and A. Asuncion, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml], 2010.

Parag A. Kulkarni is Ph.D. from IIT, Kharagpur and holds Management degree ffrom IIM Calcutta. He is also conferred DSc - Higher Doctorate by Swiss University for his Research, Professional and Business contributions. He is working in IT industry for more than 13 years. He is on research panel and Ph.D. guide for University of Pune, BITS and Symbiosis deemed University. He has conducted tutorials at various international conferences and was a keynote speaker for three international conferences. He has also worked as a referee for International Journal for Parallel and Distributed Computing, IASTED conferences. He is also member of IASTED technical committee of Parallel and Distributed Computing. He is founder of Eklat Research, Pune and is working as Chief Scientist. He is also Honorary Professor at AISSM Engineering College, Pune and on board of studies for a couple of Institutes. He has worked as Senior Manager (R&D) at Scientific Applications Center, Siemens information systems Ltd., Pune and Research Head for Capsilon Pune. His areas of interest include image processing, security systems, decision systems, expert systems, classification techniques, load balancing and distributed computing.

Prachi Joshi is working on her Ph.D. in the area of Machine Learning under the University of Pune. She has completed her M.Tech (Computer Engineering) and B.E. (Computer Engineering) from College of Engineering, Pune, India. She is working in education sector since 2002.