METHODS FOR INCREMENTAL LEARNING: A SURVEY

International Journal of Data Mining & Knowledge Management Process (IJDKP)

METHODS FOR INCREMENTAL LEARNING: A SURVEY

Ms. R. R. Ade, Assistant Professor, GHRIET, Pune
rosh513@gmail.com
Dr. P. R. Deshmukh, Professor, SCOE&T, Amravati
pr_deshmukh@yahoo.com

ABSTRACT

Incremental learning is a machine learning paradigm where the learning process takes place whenever new example/s emerge and adjusts what has been learned according to the new example/s. The most prominent difference of incremental learning from traditional machine learning is that it does not assume the availability of a sufficient training set before the learning process, but the training examples appear over time. In this paper we discuss the methods of incremental learning which are currently available. This paper gives the overview of the current research in the incremental learning which will be beneficial to the research scalars.

KEYWORDS

Incremental learning, Adaptive Classification, supervised, unsupervised, machine learning, mapping function, Ensemble Methods

1. INTRODUCTION

When we observe human learning, we clearly see that it is incremental. People learn concept description from facts and incrementally refine those descriptions when new facts and observations become available. Newly gained information is used to refine knowledge structure and models, and rarely causes reformulation of all the knowledge the person has about the subject at hand. There are two major reasons why humans must learn incrementally 1) sequential flow of information and 2) limited memory and processing power.

Incremental learning is an important capability for brain-like intelligence as biological systems are able to continuously learn through their lifetimes and accumulate knowledge over time. Key objectives of machine learning research are: transforming previously learned knowledge to the currently received data to facilitate learning from new data, accumulating experience over time to support decision making process and achieving global generalization through learning to accomplish goals.

During the incremental learning situation, raw data that come from the environment with which the intelligent system interacts become incrementally available over an indefinitely long leaning lifetime. Therefore, the leaning process is fundamentally different from that of traditional static

DOI : 10.5121/ijdkp.2013.3408

learning process, where representative data distribution is available during the training time to develop the decision boundaries.

Concept drifting is important for understanding the robustness and leaning capability during the incremental learning. For example, In case of scene analysis, new objects may appear in the visual field during the learning period. Intelligent system should have the capability to automatically modify its knowledge based to learn new data distributions.

2. INCREMENTAL LEARNING

Incremental learning algorithm can be defined as one that meets the criteria

1. It will be able to learn and update with every new data-labeled or unlabeled

2. It will preserve previously acquired knowledge

3. It should not require access to the original data
.
4. It will generate new class or cluster when required. It will divide or merge clusters as needed

5. It will be dynamic in nature with the changing environment.

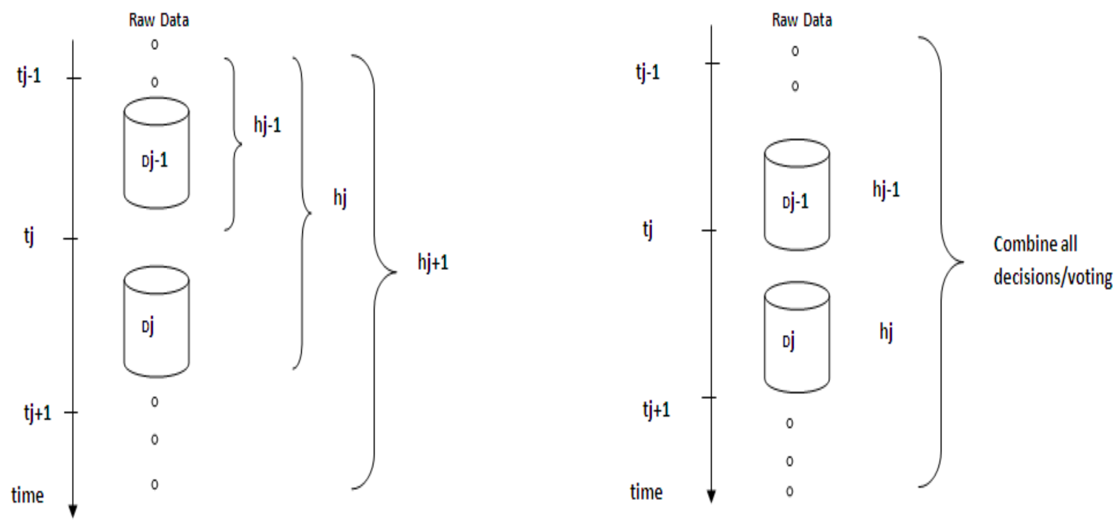

Fig: Two traditional approaches of incremental learning 1. Data accumulation methodology 2. Ensemble learning methodology

In the first method, when new chunk of data Dj is received, it will discards hj-1 and develops a new hypothesis hj, based on all the available data accumulated so far.

In the ensemble learning, when a new chunk of data Dj is received, either a single new hypothesis or a set of new hypothesis is developed based on the new data. Finally a voting mechanism can be used to combine all the decisions from different hypothesis to get the final prediction

Foundation of the Problem:

Let Dj-1 represents the data chunk received between time tj-1 and tj, and hypothesis hj-1 developed on Dj-1.

The system will adaptively learn information when new data chunk Dj is received. There are two categories, data accumulation method and ensemble learning method.

In the first method, when the new data chunk Dj is received, one simply discards hj-1 and develops a new hypothesis hj, based on all data accumulated so far.

In ensemble learning method, whenever a new chunk of data is available, either a new hypothesis hj or a set of hypothesis H:h1, i1,2,…,M, are developed based on the new data. Then the voting mechanism is used to combine all he decisions from different hypothesis to reach the final prediction. The major advantage of this approach that we do not require storage to the previously seen data, knowledge has been stored in the series of hypotheses developed along the learning life.

Knowledge at time t:

Dt is a chunk of data with n instances (i=1,…,n)

(xi,yi) is an instance in the m-dimensional feature space X

Yi $\in$Y ={1,…,K} classes

Distribution function Df

A hypothesis ht, developed by the data based on Dt with Pt

New input will be available at time (t+1)

Leaning algorithm:

1. Find the relationship between Dt and Dt+1

2.Update the initial distribution function Dt+1

3.Apply hypothesis ht to Dt+1 and calculate the pseudo-error of ht

4.Refine the distribution function for Dt+1

5. A hypothesis is developed by the data based on Dt+1 with Pt+1

6. Repeat the procedure when the next chunk of new data set is received.

Output: The final Hypothesis.

The incremental learning framework discussed in[22] focuses on two important issues, how to adaptively pass the previously learned knowledge to the presently received data to benefit leaning from the new raw data, and how to accumulate experience and knowledge over time to support future decision making processes. Mapping function is the key component of ADAIN that can effectively transform the knowledge from the current data chunk in to the leaning process of the future data chunks. Three approaches of mapping functions are namely, Mapping function based on Euclidean distance, based on regression learning model, based on online value system.

3. UNSUPERVISED INCREMENAL LEARNING

In [36] CF1 and CF2, unsupervised incremental learning algorithms based on the LF algorithm, for adapting concepts in a changing environment. It was argued that depending on the problem area, the use of unsupervised CFs such as CF1 and/or CF2 may offer significant advantages compared to alternative incremental supervised learning methods. The experimental results suggest that CF1 and CF2 are both stable and elastic.

The algorithm ARTMAP is based on generating new decision cluster in response to new patterns that are sufficiently different from previously seen instances. Each cluster learns a different hyper-rectangle shaped portion of the feature space in an unsupervised mode, which are then mapped to target classes. Since previously generated clusters are always retained, ARTMAP does not suffer from catastrophic forgetting. ARTMAP does not require access to the previously seen data, and it can accommodate new classes.

SVM algorithm based on clustering does not need to specify the number of classification compared with traditional K-means algorithm [34]. The said algorithm is appropriate for unbalanced data since centers of clusters could reflect the distribution of data well, if clustering radius is set properly.

4. SUPERVISED INCREMENTAL LEARNING

Learn++ [1]is based on the following intuition: Each new classifier added to the ensemble is trained using a set of examples drawn according to a distribution, which ensures that examples that are misclassified by the current ensemble have a high probability of being sampled. In an incremental learning setting, the examples that have a high probability of error are precisely those that are unknown or that have not yet been used to train the classifier.

In Learn++.NC[32] asks individual classifiers to cross reference their predictions with respect to classes on which they were trained. Looking at the decisions of each classifier, each classifier decides whether its decision is in line with the classes others are predicting and the classes on which it was trained. If not, the classifier reduces its vote, or possibly refrains from voting all together. It uses a new voting procedure DW-CAV, a novel voting mechanism for combining classifiers, where classifiers examine each other’s decisions, cross reference those decisions with the list of class labels on which it was trained and dynamically adjust their voting weights for each instance.

Learn++.NC [32] was later developed for learning New Classes (NC), with new data from existing classes assumed to remain stationary. Learn++.NC employs a dynamically weighted consult-and-vote mechanism to determine which classifiers should or should not vote for any given instance based on the (dis)agreement among classifiers trained on different classes. In Learn++.MF, ensemble members are trained on different subsets of the features, so that Missing Features (MF) can be accommodated by combining ensemble members trained on the currently available features. While all former Learn++ algorithms do some form of incremental learning, none of them is capable of learning from a nonstationary environment, and Learn++.NSE is developed specifically to fill this gap.

The use of Genetic Algorithm for incremental learning [33]. Each classifier agent may have a certain solution based on the attributes, classes or data are sensed from the environment or other agents, the GA is then used to learn new changes and evolve into a reinforced solution. As long as the learning process continues, this procedure can be repeated for incremental leaning.

SVM are found to be effective in large number of classification and regression problem [5,6,7,8,9]. The incremental learning algorithm[35] has two parts in order to tackle different types of incremental learning cases, online incremental learning and batch incremental learning.

5. DISCUSSION AND CONCLUSION

In this paper the different methods of incremental learning has been discussed. There are different applications of incremental learning methods, as it can be online or batch incremental learning,
The incremental leaning has wide range of applications across different domains. For instance, incremental learning from video stream, incremental leaning for spam e- mail classification, Some of the existing techniques for spam email classifications are SVM based method, neural network method, the practical swarm optimization method, the Bayesian Belief network and many others. For the above two applications the algorithms based on concept drift, ensemble methods can be used.

The paper provide useful suggestions of using supervised, unsupervised, semisupervised approach to solve practical application problems

REFERENCES

[1] Robi Polikar, Lalita Udpa, Satish S. Udpa, and Vasant Honavar, “Learn++: An Incremental Learning Algorithm for Supervised Neural Networks”, IEEE transactions on systems, Man and Cybernetics-part C: Applications and review, vol. 31 no. 4, November 2001

[2] 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.

[3] L. J. Cao, “ A SVM with Adaptive Parameters in Financial Time series Forecasting” IEEE Transactions on Neural Networks, Vol 14, No 06, pp.788, NOVEMBER 2003.

[4] Amura, Shinichi; Higuchi, Seihaku, Tanaka, Kokichi, Department of Information and Computer Sciences, Faculty of Engineering Science, Osaka University, Toyonaka, Japan 560, “Pattern Classification Based on Fuzzy Relations”, Systems, Man and Cybernetics, IEEE Transactions , pp. 61-66, 08 February 2010.

[5] R. Roscher, W. Forestner, B. Waske, I2VM: Incremental import vector ,acjomes, journal of Image and Vision Computing, Elsevier, 2012.

[6] N. A. Syed, H. Liu, and K. K. Sung, “Incremental Learning with Support Vector Machines”, International Joint Conference on Artificial Intelligence (IJCAI), Stockholm, Sweden, 1999.

[7] Alistair Shilton, M. Palaniswanmi, “Incremental learning of SVM”, IEEE transactions on Neural networks, vol. 16.pp.456-61, January 2005

[8] Alistair Shilton, "Incremental Training of support vector machines", IEEE Transactions on Neural Network, Vol 16, No. 01, January 2005.
[9] Xiao R., "An Incremental SVM learning Algorithm alpha-ISVM", Journal of Software, 12, pp. 1818-1824, 2001.
[10] S. Ozawa, S. Toh, S. Abe, S. Pang, N. Kosabov, Incremental learning for online face recognition, Proc. of IEEE Conference on Neural Networks, Vol. 5, 2005, pp 3174-3179
[11] Jianpei Zhang, Zhongwei Li and Jing Yang, "A divisional incremental training algorithm of support vector machine", Proceeding of the IEEE, International Conference on Mechatronics & Automation Niagara Falls, Canada, July, 2005.
[12] Xiaodan Wang, Chunying Zheng, Chongming Wu, Wei Wang,"A New Algorithm for SVM Incremental Learning",. ICSP Procedding, 2006.
[13] Chongming Wu, Xiaodan Wang, Dongying Bai, Hongda Zhang, "Fast SVM Incremental Learning Based on the Convex Hulls Algorithm", International Conference on Computational Intelligence and Security. 2008
[14] Ying-Chun Zhang, Feng-Feng Zhu "A New Incremental Learning Support Vector Machine", International Conference on Artificial Intelligence and Computational Intelligence, 978-0-7695-3816-7/09, 2009.
[15] X. Su, Y. an, R. Qin, A fast incremental clustering algorithm, Proc. Of International Symposium on Information Processing, 2009, pp 887-892.
[16] S. Ozawa, S. Pang, N. Kasabov, Incremental learning of chunk data for online patteren classification systems, IEEE Transactions on Neural Networks, Vo.19(6), 2008, pp 1061-1074.
[17] Yuping Qin,Qiangkui Leng,Xiangna Meng,Qian Luo, " A New Incremental Learning Algorithm Based on Hyper-Sphere SVM", Seventh International Conference on Fuzzy Systems and Knowledge Discovery ,2010.
[18] N. M. Norwawi, s. F. Abdusalam, "Classification of students performance in computer programming course according to learning style", 2nd. Conference on Data Mining and Optimization, Selangor, Malaysia., pp. . 27-28, October 2009.
[19] Durgesh K. Srivastava, Lekha Bhambhu " Data Classification using Support Vector Machine". Journal of Theoretical and Applied Information Technology, 2005.
[20] X. Yang, B. Yuan, W. Liu, Dynamic Weighting Ensembles for incremental learning. Proc. Of IEEE Conference in Pattern Recognition, 2009, pp 1-5.
[21] R. Elwell, R. Polikar, Incremental learning of Concept drift in nonstationary environments, IEEE Transactions on Neural Networks, vo,22(10), 2011 pp 1517-1531.
[22] Haibo He, Kang Li, Xin Xu, "Incremental learning from stream data", IEEE Transactions on Neural Networks, Vol. 22(12), 2011, pp 1901-1914.
[23] Gregory Dotzler, Michael D. Muhlbaier, and Robi Polikar, "Incremental Learning of New Classer in Unbalanced Datasets: Learn++.UDNC, MCS:2010, LNCS 5997,pp.33-42, 2010, Springer-Verlag Berlin Heidelberg 2010.
[24] M. Muhlbaier, A. Topalis, and R. Polikar, "Learn++.MT : A new approach to incremental leaning," 5th Int. Workshop on Multiple Classifier Systems, Springer INS, 2004, vol.3077, pp-52-61.
[25] S. U. Guan and F. Zhu, "An incremental approach to genetic-algorithmsbased classification," IEEE Trans. Syst., Man, Cybern., Part B: Cybern., vol. 35, no. 2, pp. 227–239, Apr. 2005.
[26] A. Sharma, "A note on batch and incremental learnability," J. Comput.Syst. Sci., vol. 56, no. 3, pp. 272–276, Jun. 1998.
[27] G. Gasso, A. Pappaioannou, M. Spivak, and L. Bottou, "Batch and online learning algorithms for nonconvex Neyman-Pearson classification," ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, pp. 28–47, Apr. 2011
[28] H. He and E. A. Garcia, "Learning from imbalanced data," IEEE Trans.Knowl. Data Eng., vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
[29] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in Proc. Adv. Neural Inf. Process.Syst. 9, 1996, pp. 155–161
[30] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," Stat. Comput., vol. 14, no. 3, pp. 199–222, 2003.

[31] P. J. Werbos, “Intelligence in the brain: A theory of how it works and how to build it,”Neural Network., vol. 22, no. 3, pp. 200–212, Apr. 2009

[32] M. Muhlbaier, A. Topalis, and R. Polikar, “Learn++.NC : Combining Ensemble Classifiers With Dynamically Weighted Consult-and-Vote for Efficient Incremental Leaning of New Classes” IEEE Transactions on NN vol. 20, No. 1, January 2009.

[33] Sheng-Uei Gaun and Fangming Zh, “An incremental approach to genetic-algorithm-based classification”, IEEE Transactions on systems, man and cybernetics-Part B: Cybernetics vol 35, No. 2, April 2005.

[34] Zhenyu Wu, Jan Sun, Lin Feng, Bo Jin, “ A policy of Clusster Analyzing Applied to Incremental SVM Learning with Temporal Information”, Journal of Convergence Information Technology, Vol. 6, No. 7, July 2011.

[35] Yang Hai, Wei He, Lei Fan, “An incremental learning algorithm for SVM based on Voting Principle”,International Journal of Information Processing and Management, Vol 2, No. 2, April 2011.

[36] David Hadasa, Galit Yovelb, Nathan Intratora, “ Using unsupervised incremental learning cope with gradual concept drift” , Connection Science, 23: 1, 65 — 83,13 May, 2011