A Survey of Online Learning Algorithm

Bharati Patel
Faculty
Department of Computer Science
NIT Raipur(C.G.),India

Abstract : Online learning is the powerful research area which is found since so many years in the research history of machine learning and it is considered as a part of machine learning or as a subset of machine learning. Research area concludes it's limited benefits in the area of real-time applicability because of its high requirement of processing environment. In this survey the various techniques and methods according to which the online learning algorithms have been able to successfully producing the results over the World Wide Web (WWW) is studied. It is very demandable and considered to be a challenging research area because of the problem of these parameters of bandwidth & latency where the non-linear classifiers of machine learning are required to be used in combination of weights(collection of new instances) that are to be learned simultaneously for the processing of an online given task.

***IndexTerms* - Component Online learning, relative loss bounds, bandwidth, latency, multi-processing, multimode.**

I. INTRODUCTION

The past decade has witnessed the emergence of kernel based online learning or simply online learning paradigm which is just a subset of machine learning with a combination of ubiquitously computational load balancing approach over a network [1][5]. This online algorithms tend to work in closed loops or rounds, where each of this round generates several new instances of confusion matrices for the consolidate sets known by the name as support set. Now, since all confusion matrices and prediction states aren't sufficient to be stored in the support sets which lead to the problem of relative memory loss bounds. The learned data in an online algorithm is known as hypothesis; thus due to this relative loss in memory bound for the enough support sets to exist required for decision making for online computational jobs is hindered. Therefore, there are several functions are deployed in an online leaning approach, wherein the classification of the weighted sum of combination are stored in support sets. Hence, making the system to update itself with the variation of the weighted sum and prioritize or predict accordingly with the modified online hypothesis. But since, if this continues so on then the support required to keep the online hypothesis will grow unbounded and leading to memory explosions. Thus, far this has been a major problem in online learning algorithms and is required other several novel methods for its resurrection and effective applicability. Since, such algorithm is associated with autonomous web agents, big data analytics and real-time data analysis. Therefore a solution to this problem will give boost to this research scenario. To accomplish the same the researchers in the past have used a method which relies on the term known as budget. The budget is the limit of the support set which uses heuristic method to decide which instance has to be kept and removed [2][3][6]. Others strategy include the use of NORMA & SILK which are very similar on such an approach [5][7]. In a study, they have used Forgetron algorithm, which relies on a hybrid approach of memory budget and relative mistake bounds [8]. Latter in the studies a stochastic algorithm is presented which gives similar performance range on an average basis [9][11]. The Gaussian approach is another approach used for the same where instances are discarded and spanned over the mistakes in support sets [12]. Due to its effectively of avoiding hinge loss its gives sparser solutions but not as much as compared with the Langford. Here the few parameters are proposed to induce scarcity in online learning budgets [13].

II. LITERATURE REVIEW

2.1Taxonomy of Online Parallelization Strategies

2.1.1 Instance based sharding- Instance based sharding approach gives the linear sequence of prediction by the single data weigth is called instance based sharding. [1]

2.1.2 Feature based sharding- Feature based sharding approach gives the collection of indivisual predicted data by the prediction based system.[3]

2.1.3 Multinode feature based sharding- multimode feature sharding approach gives the prediction based on multiple nodes and all the nodes are connected to a master node so the whole process will give the combined result which is coming from all the nodes.[4]

2.1.4 Multicore fearture based sharding- multicore feature sharding approach gives the prediction based on a linear sequence of predicted values from each core.[1-5]

2.1.5 Multinode instance based sharding- multimode instance based sharding provides the facility to each node is geeting the abd there is no problem og bottleneck.[8]

2.1.6 Multicore instance based sharding- multicore instance based sharding provides the facility of each data one by one collected by the system and getting updated by thr partially predicted system.[13]

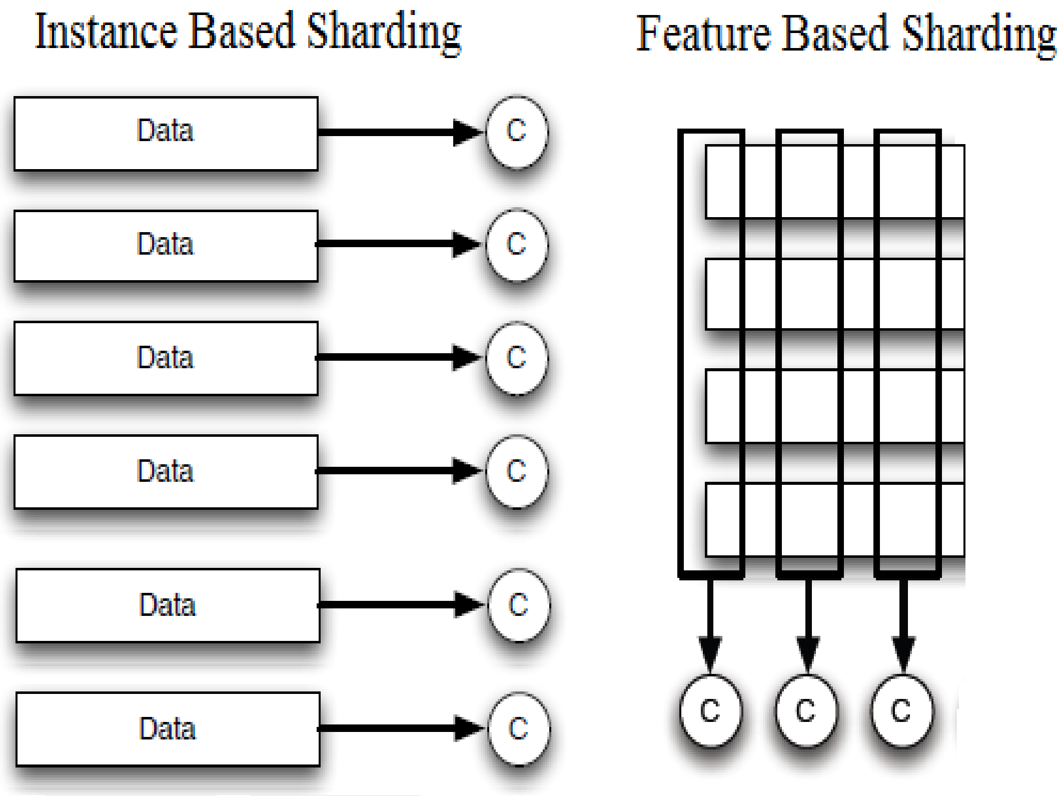

Fig. 1: Data Splitting for Online Learning Depending on Two Approach Namely Instance Based Sharding on the left side of the image and feature based sharding on the right side of the image and these are

3.1 Multicore Feature Sharding

In this type of multi-core sharding approach there were multi-core processors are used which consists of multiple (Central processing Units) CPUs to operate the computational job processing in a shared memory space. Since, it is a multi-core parallelization thus it doesn'taddress the primary bandwidth at its bottleneck, thereby increasing its usefulness to some extent but limited to only certain datasets and learning algorithms due to its substantial computation per raw cycle of its sharding. In the current scenario, this is implemented as the paring of features sets from the instances of online leaning and giving a linear prediction to the support set.[10]

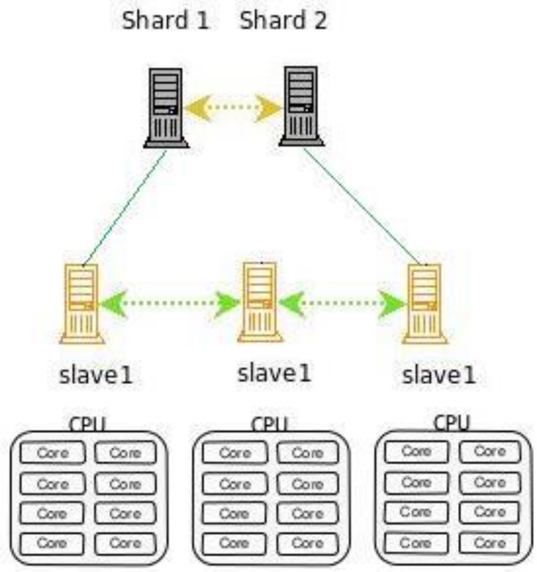

Fig: 2: Illustration of Multi-Core Feature Sharding

3.2 Multinode Feature Sharding

The strategy that lays her is to employ sharding across the nodes to update its parameters online like as single node learning algorithm. Hence, ignoring the overheads caused in creating & distributing of shardes as was in the previous case. The datasets are minimized per instances by rearranging itself and thereby resulting in a fully decoupled computation which doesn't require other computation to be completed before the next one. Consequently, we have n predictor for n sharding nodes. The steps in this type of process involve:

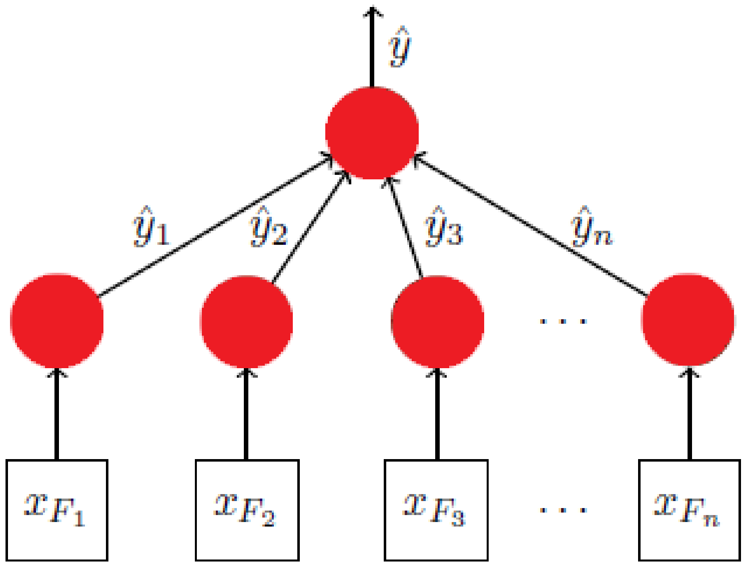

Fig. 3: Illustration of Multi-Node Sharding.

3.2.1 Each node computes and transmits prediction states to a master node in each of its received instances and simultaneously updating each of its parameters.

3.2.2 The master node treats this received prediction states as a feature sets. Thus, making the master node to semantically learn in order to predict the label, corresponding to each of its instances and that of the other ones. The main distinction between multi-core and multi-node parallelization strategy is the latency, as the latency between many nodes is obviously larger than that of between the cores. Such that giving larger time to process between nodes for any per-instances based operations than that was used for blocking operations in multi-core scenario. This implied towards delayed updates and thereby degrading the performance. [9][8][10]

IV. PROBLEM IDENTIFICATION

The problematic features that restrict the usage of online learning are as follows:

4.1 The limits of bandwidth.

4.2 The limits of latency.

4.3 Relative Memory Loss Bound.

4.4 Memory Explosions.

4.5 The substantial part of the degraded performance is the slowdown posed by all in combination.

4.6 Sparsely of the data sets there occurs latency to access RAM or cache for many

4.7 cycles and ultimately causing substantially slowdown the computation & learning.

4.8 Large data instances are required to be loaded for a complete online hypothesis to function.

V. CONCLUSION

In this paper I found that there is so many future work on the basis of online machine learning like sentimental analysis from the web based data from the several users. It gives the benefit of time saving because there is less time to train the data and specifically to train the new data set it requires lots of time therefore pre-processing time will be less.

REFERENCES

[1]. Y. Freund and R. E. Schapire. Large margin classification using the Perceptron algorithm. Machine Learning. 1999; 277–296p.
[2]. K. Crammer and Y. Singer. Ultraconservative online algorithms for multiclass problems. Journal of Machine Learning Research. 2003; 3: 951–991p.
[3]. K. Crammer, J. Kandola, and Y. Singer. Online classification on a budget. In Advances in NeuralInformation Processing Systems. 2003;16.
[4]. K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer. Online passive-aggressivealgorithms. Machine Learning Research. 2006; 7: 551–585p.
[5]. J. Kivinen, A. Smola, and R. Williamson. Online learning with kernels. IEEE Trans. on Signal Processing. 2004; 52(8): 2165–2176p.
[6]. J. Weston, A. Bordes, and L. Bottou. Online (and offline) on an even tighter budget. In Robert G.Cowell and ZoubinGhahramani, editors. Proc. of AISTATS. 2005; 413–420p.
[7]. L. Cheng, S. V. N. Vishwanathan, D. Schuurmans, S. Wang, and T. Caelli. Implicit online learning with kernels. In Advances in Neural Information Processing Systems. 2007; 249–256p.

[8]. O. Dekel, S. Shalev-Shwartz, and Y. Singer. The Forgetron: A kernel-based perceptron on a budget. SIAM Journal on Computing. 2007; 37(5): 1342–1372p.

[9]. N. Cesa-Bianchi, A. Conconi, and C. Gentile. On the generalization ability of on-line learningalgorithms. IEEE Trans. on Information Theory. 2004; 50
(9): 2050–2057p.

[10]. N. Cesa-Bianchi, A. Conconi, and C. Gentile. A second-order perceptron algorithm. SIAM Journal on Computing. 2005; 34(3):640–668p.

[11]. N. Cesa-Bianchi, A. Conconi, and C. Gentile. Tracking the best hyperplane with a simple budget Perceptron. In Proc. of the 19th Conference on Learning Theory. 2006; 483–498p.

[12]. L. Csat´o and M. Opper. Sparse on-line gaussian processes. Neural Computation. 2002; 14: 641–668p.

[13]. Thanh-Nghi Doan, Thanh-Nghi Do, Francois Poulet, "Parallel incremental SVM for classifying million images with very high-dimensional signatures into thousand classes", Neural Networks (IJCNN) The 2013 International Joint Conference on, pp. 1-8, 2013, ISSN 2161-4407.