

Random Forest Based Approach for Concept Drift Handling

Aleksei V. Zhukov^{1,2}
 Denis N. Sidorov^{1,2} and Aoife M. Foley³

¹ Energy Systems Institute SB RAS, Irkutsk, Russia
 dsidorov@isem.irk.ru,

² Irkutsk State University, Irkutsk, Russia
 zhukovalex13@gmail.com

³ Queens University Belfast, Belfast, UK
 a.foley@qub.ac.uk

Abstract. Concept drift has potential in smart grid analysis because the socio-economic behaviour of consumers is not governed by the laws of physics. Likewise there are also applications in wind power forecasting. In this paper we present decision tree ensemble classification method based on the Random Forest algorithm for concept drift. The weighted majority voting ensemble aggregation rule is employed based on the ideas of Accuracy Weighted Ensemble (AWE) method. Base learner weight in our case is computed for each sample evaluation using base learners accuracy and intrinsic proximity measure of Random Forest. Our algorithm exploits both temporal weighting of samples and ensemble pruning as a forgetting strategy. We present results of empirical comparison of our method with original random forest with incorporated “replace-the-looser” forgetting and other state-of-the-art concept-drift classifiers like AWE2.

Keywords: machine learning, decision tree, concept drift, ensemble learning, classification, random forest.

1 Introduction

Ensemble methods of classification (or, briefly, ensembles) employ various learning algorithms to obtain better predictive accuracy comparing with individual classifiers. Ensembles are much used in research and most of the research is devoted to stationary environments where the complete datasets are available for learning classifiers and transfer functions of dynamical systems are not changing as time goes on. For real world applications (e.g. in power engineering [1], [2]) learning algorithms are supposed to work in dynamic environments with data continuously generated in the form of a stream on not necessarily equally spaced time intervals. Data stream processing commonly relies on single scans of the training data and implies restrictions on memory and time. Changes caused by dynamic environments (e.g. consumer behaviour in future smart grids) can be

categorised into sudden or gradual concept drift subject to appearance of novel classes in a stream and the rate of changing definitions of classes.

One of the most generally effective ensemble classifier is a Random Forest. This algorithm employs bagging [13] and principles of the random subspace method [14] to build a highly decorrelated ensemble of decision trees [16]. There are some attempts [10] to adapt Random Forest to handling concept-drift, but this approach is not fully discovered.

The objective of this paper is to propose a novel approach to classification to adapt to concept drifts. This novel classification method can be applied to consumer behaviour in future smart grids to predict and classify the random behaviour of humans when they interact with smart appliances in the home and charging and discharging of electric vehicles. Likewise in these methods can be used to forecast windpower and indeed solar power.

We propose to compute the base learner weight for samples evaluation using base learners accuracy and intrinsic proximity measure of random forest.

It is to be noted that concept drifting is related to non-stationary dynamical systems modelling [4], where the completely supervised learning is employed using a special set of input test signals. For more details concerning integral dynamical models theory and applications refer to monograph [5] and its bibliography.

This paper is organised into five sections. In Section 2, a brief review of ensemble streaming classifiers based on random forest is given. Section 3 delivers detailed presentation of proposed algorithm. In Section 4 experiments on both machine learning and tracking tasks are provided. Finally, the paper concludes with Section 5 where the main results are discussed.

2 Related Work

The use of drifting concepts for huge datasets analysis is not unfamiliar to the machine learning and systems identification communities [6]. In this work we restrict ourselves to considering decision tree ensemble classification methods only.

Let us briefly discuss methods most related to our proposal and employed in the experiments. For a more detailed overview of the results in this area, including online incremental ensembles, readers may refer to monograph [7] and review [8]. Bayesian logistic regression was used in [9] to handle drifting concepts in terms of dynamical programming. Concept drifting handling is close to methodology of on-line random forest algorithm [10] where ideas from on-line bagging, extremely randomised forests and on-line decision tree growing procedure are employed. Accuracy Weighted Ensemble (AWE) approach was proposed in [6]. The main idea is to train a new classifier on each incoming dataset and use it to evaluate all the existing classifiers in the ensemble. We incorporate this idea in our approach.

A Massive Online Analysis (MOA) framework [11] is the one of the popular benchmarks for testing online classification, where clusterization and regression

algorithms are written in Java. Java software contains the state of the art classifiers for concept drift handling such as SEA [20] and Online Bagging [22].

Analysis of ensemble methods with decision tree base learners is interesting topic to be addressed in this paper. The recursive nature of decision trees makes on-line learning a difficult task due to the hard splitting rule, errors cannot be corrected further down the tree.

3 Proximity Driven Streaming Random Forest

In paper [12] the authors clearly demonstrated that a classifier ensemble can outperform a single classifier in the presence of concept drifts when the base classifiers of the ensemble are adherence weighted to the error similar to current testing samples. We propose another approach which exploits Random Forest properties.

To produce a novel algorithm capable of handling concept drift the following questions need to be answered:

- How to adapt original Random Forest for data streaming?
- How to define sample similarity metric?
- How to choose the base classifier weighting function?
- How to choose forgetting strategy?

These questions are considered in the following subsections.

3.1 Streaming classifier based on Random Forest

Methodologically ensemble approaches allow concept-drift to be handled in the following ways: base classifier adaptation, changing in training dataset (such as Bootstrap [13] or RSM [14]), ensemble aggregation rule changing or changing in structure of an ensemble (pruning or growing). In this paper we propose Proximity Driven Streaming Random Forest (PDSRF) which exploit combinations of these approaches. Besides some methods are already incorporated to the original Random Forest. Contrary to conventional algorithms we use weighted majority voting as an aggregation rule of ensemble. This allows us to adapt the entire classifier by changing the weights of the base learners. In order to obtain the classifiers weight estimation we should store samples. For this purpose we use a sliding windows approach which is used in the periodically updated Random Forest [15]. The length of this window is fixed and can be estimated by cross-validation. Random Forest [16] uses unpruned CART [17] trees and for N instances and $Mtry$ attributes chosen at each decision tree node average computational cost of ensemble building is $O(TMtryNlog^2N)$, where T is a number of trees. It can be unsufficient for online applications. To reduce the complexity we use the randomization approach proposed in Extremely Randomized Trees [18]. In our implementation the split set consists of randomly generated splits and the best one is chosen by minimization of the Gini-index measure. So that $O(TMtryNlogN)$ cost complexity can be achieved.

3.2 Sample similarity metric

We employ the assumption that the base classifiers make similar errors on similar samples even under concept-drift. Conventional Random Forest employ the so called proximity measure. It uses a tree structure to obtain similarity in the following way: if two different sample are in the same terminal node, their proximity will be increased by one. At the end, proximities are normalised by dividing by the number of trees [16].

3.3 Base classifier weighting function

Following the AWE approach proposed in [12] we use an error rate to produce weights (1) of classifiers, where E is an new block testing error for i -th classifier, ε is a small parameter.

$$w_i = 1/(E^2 + \varepsilon) \quad (1)$$

3.4 Forgetting strategy

One of the main problems in concept-drifting learning is to select the proper forgetting strategy and forgetting rate [8]. The classifier should be adaptive enough to handle changes. In this case different strategies can be more appropriate to different types of drift (for example, sudden and gradual drifts). In this paper we focus on gradual changes only.

We propose two different ways to handle the concept-drift:

- temporal sample weighting,
- ensemble pruning technique.

Data forgetting through temporal sample weighting. We use sample weighing to decrease the influence of old samples to learn the new trees (2).

$$w_x(t) = \exp(-\alpha t) \quad (2)$$

where α is the sample weighting rate which can be selected in experiment by cross-validation.

Knowledge forgetting through ensemble growing and pruning In this part we apply the classic *replace-the-looser* approach [8] to discard trees with high error on new block samples.

3.5 Algorithm

To predict the sample we Algorithm 1. First we use a stored window to find similar items using the specified similarity metric. Second we evaluate our current ensemble on similar examples. Then we compute weights adherence to errors on k similar samples.

On every chunk the algorithm tests all the trees to choose the poorest base learner and replace it with new one trained on new block data. This process is iterative while the ensemble error on new block samples is higher than a specified threshold.

Input:

S : data stream chunks which sequentially produce training examples $\langle x, y \rangle$

S_i : data stream chunks of $blockSize$

k : number of the nearest neighbours

$W(E)$: Classifier weighting function

Output: Class probability vector

$nearest \leftarrow$ find k nearest samples from cache using proximity metric.

for *all* $x_i \in S_i$ **do**

for *all* trees c_i in ensemble **do**

 Get average error $E_i^{nearest}$ of c_i on nearest samples Add weighted by (1)
 base classifier probability to class probability vector

end

end

return *Probability Vector*

Algorithm 1: PDSRF prediction algorithm

4 Experimental Evaluation

In this experiment, we evaluate our algorithm on various publicly available datasets like CoverType [19], and compared it to the most popular concept-drift classifiers. We compare our algorithm with SEA [20], Hoeffding Adaptive tree [21], Online Bagging [22] and AWE [6] implemented in MOA [11]. Our algorithm was implemented natively in C++ according to the same testing methodology. Extremely randomized trees is used as a base learner. With adherence to this methodology classification accuracy was calculated using the *data block evaluation method*, which exploits the *test-then-train* approach. The data block evaluation method reads incoming samples without processing them, until they form a data block of size d . Each new data block is first used to test the existing classifier, then it updates the classifier [23].

Proposed proximity driven streaming random forest was tested with $blockSize = 300$, $windowSize = 1500$, number of nearest neighbours is 20 and ensemble consists of 30 decision trees. In order to make the results more interpretable we also test original random forest with incorporated “*replace-the-looser*” forgetting.

Method	Mean accuracy
RF	0.54
PDSRF	0.63
AVE2(buffered)	0.81

Table 1. Cover type dataset mean accuracy results

4.1 Datasets

We follow the literature on adaptive ensemble classifiers and select the publicly available benchmark datasets with concept drift. In this paper the proposed method was tested on the *cover type dataset* from Jock A. Blackard, Colorado State University [19]. *Cover type dataset* contains the forest cover type for 30×30 meter cells obtained from US Forest Service Region 2 Resource Information System (RIS) data. It contains 581 012 instances and 54 attributes, and it has been used as a benchmark in several papers on data stream classification.

4.2 Results

The proposed approach shows results similar to the above mentioned AUE2 [24]. Next comparison (ref. Fig. 1) of the original random forest with incorporated “replace-the-looser” forgetting and proposed proximity driven streaming random forest are presented. In Table 1 the mean accuracy is shown.

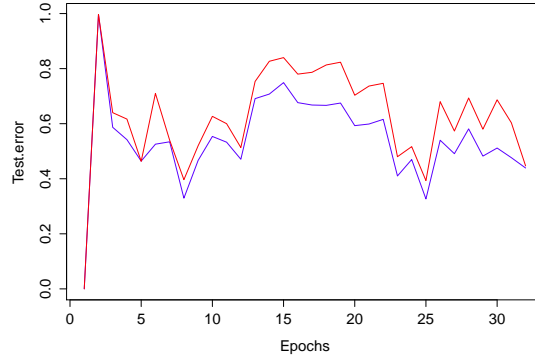


Fig. 1. Original random forest with incorporated “replace-the-looser” (blue) and PDSRF (red) accuracy.

Remark 1. For sake of space, we do not report the complete tests in this paper. For more details are complete results of our approach evaluation readers may refer to <http://mmwind.github.io/pdsrf>.

5 Discussion and Conclusions

In comparison to other concept-drift approaches like Online Random Forest and the AWE, our approach needs more computational resources and thus more time for both the training and prediction stages. But the proposed approach is highly parallelisable and can be implemented using the GPGPU. It must be noted that the proposed approach can be efficiently applied only to gradual concept drifts. PDSRF is senisible to all the parameters changes and all of these parameters must be accurately tuned. As it shown the proposed approach significantly exceeds the original random forest with incorporated “replace-the-looser” forgetting. Although the presented results show that the accuracy is lower than AWE2, the approach has some promising directions is that the Random Forest can be used in unsupervised and allows to work with missing data, which is an issue with smart grid datasets and wind power forecasting where telecommunications signals and data recording is not 100% robust.s

Acknowledgment. This work is partly funded by the RSF grant No. 14-19-00054.

References

1. Zhukov, A., Kurbatsky, V., Tomin, N., Sidorov, D., Panasetsky, D., Foley, A.: Ensemble methods of classification for power systems security assessment. arXiv, Artificial Intelligence (cs.AI) **arXiv:1601.01675** (2016) 1–6
2. Tomin, N., Zhukov, A., Sidorov, D., Kurbatsky, V., Panasetsky, D., Spiryaev, V.: Random forest based model for preventing large-scale emergencies in power systems. *International Journal of Artificial Intelligence* **13** (2015) 211–228
3. Sidorov, D.: Modeling of nonlinear time-dependent dynamical systems by Volterra series: identification and application. *Sib. Zh. Ind. Mat.* **1** (2000) 182–194
4. Sidorov, D.: Modelling of non-linear dynamic systems by Volterra series. In: *Attractors, Signals, and Synergetics Workshop, 2000*, Pabst Science Publ. USA-Germany (2002) 276–282
5. Sidorov, D.: *Integral dynamical models: singularities, signals and control*. Singapore: World Scientific Publ. (2015)
6. Wang, H., Fan, W., Yu, P.S., Han, J.: Mining concept-drifting data streams using ensemble classifiers. In: *Proceedings of SIGKDD, August 24–27, 2003, Washington, DC, USA*. (2003) 226–235
7. Gama, J.: *Knowledge discovery from data streams*. Singapore: CRC Press Publ. (2010)
8. Kuncheva, L.: Classifier ensembles for changing environment. In Roli, F., Kittler, J., Windeatt, T., eds.: *Multiple Classifier Systems, 2004 5th Intl. Workshop on*, Springer-Verlag (2004) 1–15

9. Turkov, P., Krasotkina, O., Mottl, V.: Dynamic programming for Bayesian logistic regression learning under concept drift. In: Pattern Recognition and Machine Intelligence: 5th International Conference, PReMI 2013, Kolkata, India, December 10-14, 2013. Proceedings. Springer Berlin Heidelberg, Berlin, Heidelberg (2013) 190–195
10. Saffari, A., Leistner, C., Santner, J., Godec, M., Bischof, H.: On-line random forests. In: Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, IEEE (2009) 1393–1400
11. Bifet, A., Holmes, G., Kirkby, R., Pfahringer, B.: MOA: massive online analysis. *Journal of Machine Learning Research* **11** (2010) 1601–1604
12. Wang, H., Fan, W., Yu, P.S., Han, J.: Mining concept-drifting data streams using ensemble classifiers. In: Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining, ACM (2003) 226–235
13. Breiman, L.: Bagging predictors. *Machine Learning* **24**(2) (1996) 123–140
14. Ho, T.K.: The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **20**(8) (Aug 1998) 832–844
15. Zhukov, A., Kurbatsky, V., Tomin, N., Sidorov, D., Panasetsky, D., Spiryaev, V.: Random forest based model for emergency state monitoring in power systems. In: Mathematical Method for Pattern Recognition: Book of abstract of the 17th All-Russian Conference with Interneational Participation, Svetlogorsk, TORUS PRESS (2015) 274
16. Breiman, L.: Random forests. *Machine Learning* **45**(1) (2001) 5–32
17. Breiman, L., Friedman, J., Olshen, R., Stone, C., Steinberg, D., Colla, P.: *Cart: Classification and regression trees*. Wadsworth: Belmont, CA **156** (1983)
18. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. *Machine learning* **63**(1) (2006) 3–42
19. Blake, C.L., Merz, C.J.: *UCI repository of machine learning databases* (1998)
20. Street, W.N., Kim, Y.: A streaming ensemble algorithm (SEA) for large-scale classification. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '01, New York, NY, USA, ACM (2001) 377–382
21. Bifet, A., Gavaldà, R.: Adaptive Learning from Evolving Data Streams. In: Advances in Intelligent Data Analysis VIII: 8th International Symposium on Intelligent Data Analysis, IDA 2009, Lyon, France, August 31 - September 2, 2009. Proceedings. Springer Berlin Heidelberg, Berlin, Heidelberg (2009) 249–260
22. Oza, N.: Online bagging and boosting. In: Systems, Man and Cybernetics, 2005 IEEE International Conference on. Volume 3. (Oct 2005) 2340–2345 Vol. 3
23. Brzezinski, D.: Mining data streams with concept drift. Diss. MS thesis. Dept. of Computing Science and Management. Poznan University of Technology (2010)
24. Brzezinski, D., Stefanowski, J.: Reacting to different types of concept drift: The accuracy updated ensemble algorithm. *Neural Networks and Learning Systems, IEEE Transactions on* **25**(1) (Jan 2014) 81–94