

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
zhukovalex13@gmail.com

² Irkutsk State University, Irkutsk, Russia
dsidorov@isem.irk.ru

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 ensemble pruning as a forgetting strategy. We present results of empirical comparison of our method and other state-of-the-art concept-drift classifiers.

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 papers 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 [3] and principles of the random subspace method [4] to build a highly decorrelated ensemble of decision trees [5]. There are some attempts [6] 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 wind power 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 [7], 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 [8] and its bibliography.

This paper is organised into five sections. In Sect. 2, a brief review of ensemble streaming classifiers based on random forest is given. Section 3 delivers detailed presentation of proposed algorithm. In Sect. 4 experiments on both machine learning and tracking tasks are provided. Finally, the paper concludes with Sect. 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 [9]. 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 [10] and review [11]. Bayesian logistic regression was used in [12] to handle drifting concepts in terms of dynamical programming. Concept drifting handling is close to methodology of on-line random forest algorithm [6] 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 [9]. 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.

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 [13] 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 classification 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 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 [3] or RSM [4]), 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 [14]. The length of this window is fixed and can be estimated by cross-validation. For the sake of time and memory optimization Extremely randomized trees [15] is used as a base learner instead of original randomized trees.

3.2 Base Classifier Weighting Function

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 [5].

Following the AWE approach proposed in [13] 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.

$$w_i = 1/(E^2 + \varepsilon) \tag{1}$$

3.3 Forgetting Strategy

One of the main problems in concept-drifting learning is to select the proper forgetting strategy and forgetting rate [11]. 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.

In this paper we propose ensemble pruning technique to handle the concept-drift. This technique uses the classic *replace-the-looser* approach [11] to discard trees with high error on new block samples.

3.4 Algorithm

To predict the sample we propose the following algorithm. 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.

4 Experimental Evaluation

In this experiment, we evaluate our algorithm on publicly available dataset CoverType [16], and compared it to the most popular concept-drift classifiers. Our algorithm was implemented natively in C++ according to the same testing methodology. 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 [17].

Proposed proximity driven streaming random forest was tested with different sizes of block, window, nearest neighbours number and ensemble size (Table 2). In order to make the results more interpretable we also test original random forest with incorporated “*replace-the-looser*” forgetting.

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 [16]. *Cover type dataset* contains the forest cover type for 30×30 m 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 [18]. 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. It must be noted that in case of high window size optimal number of nearest neighbours tends to one (Fig. 2).

Remark 1. For more details of our approach evaluation readers may refer to <http://mmwind.github.io/pdsrf>.

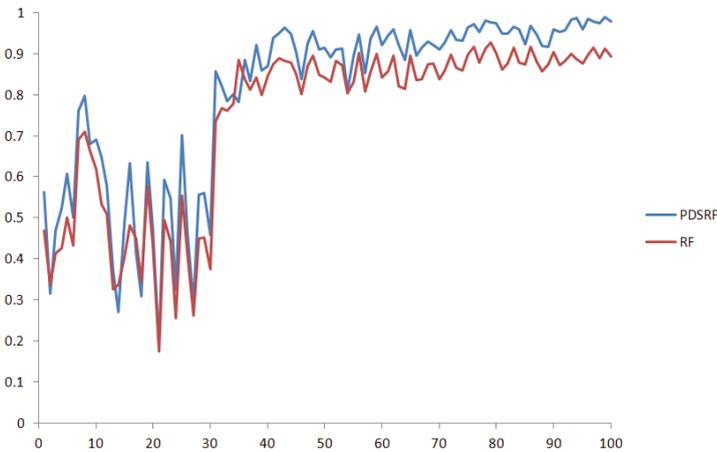


Fig. 1. Original random forest with incorporated “replace-the-looser” and PDSRF accuracy on first 100 chunks of the dataset.

Table 1. Cover type dataset mean accuracy results. Some of the results obtained using settings from [18].

Method	Mean accuracy
PDSRF	87.42
HOT	86.48
AUE2	85.20
PDSRF(without weighting)	82.79
AUE1	81.24
Lev	81.04
DWM	80.84
Oza	80.40
AWE	79.34
Win	77.19

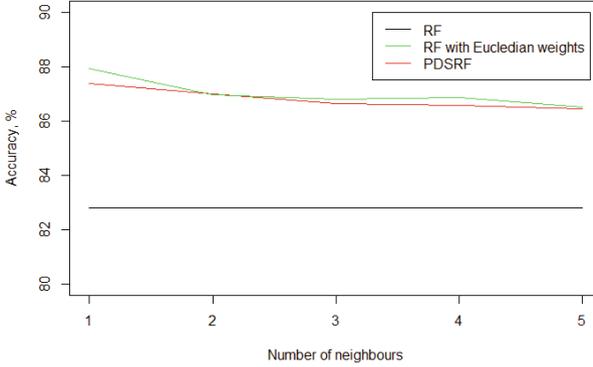


Fig. 2. Mean accuracy for Random Forest, and proposed method with intrinsic proximity and euclidean measures for block size equal to 500 and window size 1500.

Table 2. Cover type dataset mean accuracy results where k is a number of nearest neighbours.

Block size	Window size	k	PDSRF without weighting	PDSRF
300	1000	5	77.65	81.15
300	1000	10	77.69	81.05
300	1000	20	77.69	80.52
300	1500	5	77.67	81.11
300	1500	10	77.89	81.21
300	1500	20	77.8	80.83
500	500	5	82.76	86.38
500	500	10	82.73	86.16
500	500	20	82.68	86.02
500	1000	5	82.76	86.45
500	1000	10	82.87	86.27
500	1000	20	82.75	86.04
500	1500	5	82.75	86.49
500	1500	10	82.74	86.29
500	1500	20	82.7	85.96

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 PDSRF can outperform the most of modern methods as it is demonstrated in the experiment. The proposed

Table 3. Mean accuracy on first 100 blocks of cover type dataset.

Block size	Window size	k	Number of trees	PDSRF (without weighting)	PDSRF
500	1000	1	5	75.03	79.25
500	1000	1	7	74.37	79.42
500	1000	1	10	74.34	78.76
500	1000	1	13	74.22	79.56
500	1000	1	15	74.00	79.72
500	1000	1	17	74.18	80.09
500	1000	1	20	74.23	79.90
500	1000	2	5	74.96	78.77
500	1000	2	7	74.36	79.33
500	1000	2	10	74.14	79.51
500	1000	2	13	74.03	79.02
500	1000	2	15	74.02	79.51
500	1000	2	17	74.35	79.68
500	1000	2	20	74.09	79.85
500	1000	3	5	74.83	77.78
500	1000	3	7	74.66	79.59
500	1000	3	10	74.34	79.14
500	1000	3	13	74.35	79.17
500	1000	3	15	74.09	79.32
500	1000	3	17	74.09	79.68
500	1000	3	20	74.13	80.22
500	1000	5	5	74.59	78.70
500	1000	5	7	74.29	78.76
500	1000	5	10	74.26	78.66
500	1000	5	13	73.94	79.52
500	1000	5	15	74.37	80.16
500	1000	5	17	74.09	79.41
500	1000	5	20	74.01	80.16

approach is highly parallelizable and can be implemented using the GPGPU. As it was shown in Table 3 high accuracy can be achieved with relatively small number of base classifiers. It must be noted that the proposed approach can be efficiently applied only to gradual concept drifts. PDSRF is sensible 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 higher than AUE2, 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.

Acknowledgment. This work is funded by the RSF grant No. 14-19-00054 and by the International science and technology cooperation program of China, project 2015DFR70850, NSFC Grant No. 61673398.

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), pp. 1–6. [arXiv:1601.01675](https://arxiv.org/abs/1601.01675) (2016)
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. *Int. J. Artif. Intell.* **13**, 211–228 (2015)
3. Breiman, L.: Bagging predictors. *Mach. Learn.* **24**(2), 123–140 (1996)
4. Ho, T.K.: The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(8), 832–844 (1998)
5. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001)
6. Saffari, A., Leistner, C., Santner, J., Godec, M., Bischof, H.: On-line random forests. In: 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), 1393–1400. IEEE (2009)
7. Sidorov, D.: Modelling of non-linear dynamic systems by Volterra series. In: *Attractors, Signals, and Synergetics Workshop*, vol. 2000, pp. 276–282. Pabst Science Publication, USA-Germany (2002)
8. Sidorov, D.: *Integral Dynamical Models: Singularities, Signals and Control*. World Scientific Publishing, Singapore (2015)
9. Wang, H., Fan, W., Yu, P.S., Han, J.: Mining concept-drifting data streams using ensemble classifiers. In: *Proceedings of SIGKDD*, 24–27 August 2003, Washington, DC, USA, pp. 226–235 (2003)
10. Gama, J.: *Knowledge discovery from data streams*. CRC Press Publishing, Singapore (2010)
11. Kuncheva, L.: Classifier ensembles for changing environment. In: Roli, F., Kittler, J., Windeatt, T. (eds.) *2004 5th International Workshop on Multiple Classifier Systems*, pp. 1–15. Springer, Heidelberg (2004)
12. Turkov, P., Krasotkina, O., Mottl, V.: Dynamic programming for Bayesian logistic regression learning under concept drift. In: Maji, P., Ghosh, A., Murty, M.N., Ghosh, K., Pal, S.K. (eds.) *PREMI 2013*. LNCS, vol. 8251, pp. 190–195. Springer, Heidelberg (2013)
13. 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*, pp. 226–235. ACM (2003)
14. 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 International Participation*, p. 274. TORUS PRESS, Svetlogorsk (2015)

15. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. *Mach. Learn.* **63**(1), 3–42 (2006)
16. Blake, C.L., Merz, C.J.: UCI repository of machine learning databases (1998)
17. Brzezinski, D.: Mining data streams with concept drift. Dissertation MS thesis. Department of Computing Science and Management, Poznan University of Technology (2010)
18. Brzezinski, D., Stefanowski, J.: Reacting to different types of concept drift: the accuracy updated ensemble algorithm. *IEEE Trans. Neural Netw. Learn. Syst.* **25**(1), 81–94 (2014)