Impact of Randomization on Ensembles for Streams with Concept Drift

Kanu Goel & Shalini Batra
Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala (Punjab), India
E-mail: kanugoel10@gmail.com
December 2020

Abstract. In the present era many real world applications are built from streaming data. The distribution of underlying data in such streams tend to change with course of time called as concept drift. A lot of algorithms are proposed in machine learning and data mining domain which are used to handle this drifting scenario. Ensembles form an integral component of such algorithms. In this paper, randomization is added using online bagging to the existing four drift handling approaches and its effect is analysed over multiple patterns of concept drift such as gradual, abrupt, recurring etc. Experimental work has been conducted over artificially generated data streams and real datasets to validate the impact of bagging in learning process.

1. Introduction

With advancement in technology, data is collected in varied forms and from multiple sources. Often the data processed in streams undergoes drift due to which the classification applications like fraud detection, intrusion detection, spam filtering, text mining etc. get affected[1][2]. Their predictive accuracy is at stake and decreases with time [3]. Ensembles where the constituent learners participate in final decision making must be up to date along with optimum learning strategy. Randomization can be added to such learning with bagging which might have good impact in increasing ensemble accuracy. Some of the concept drift handling ensemble based approaches are discussed in forthcoming section in which randomization is later incorporated.

2. Concept Drift approaches

• Weighted Majority Algorithm(WMA): This algorithm is one of the popular primitive approaches to tackle concept drift. It is based on the assumption that learners with more weight tend to give better accurate prediction results than the lower ones. Here a particular weight is assigned to each ensemble learner use to estimate final predictive capability. Weights are decreased or increased dynamically
Table 1. Characteristics of Datasets

| Dataset Name     | Drift type                             | Attributes | Classes | Generator name |
|------------------|----------------------------------------|------------|---------|----------------|
| MIXED_Incrmntl   | Incremental Drift                      | 4          | 2       | Mixed Generator[7] |
| MIXED_Abrpt      | Abrupt, Fast Moving                    | 4          | 2       | Mixed Generator |
| WAVE_Grl         | Gradual, Slow                          | 40         | 3       | Wave Generator[7] |
| WAVE_Abrpt       | Abrupt, Fast Moving                    | 40         | 3       | Wave Generator  |
| RBF_Recc+Grl     | Recurring with Gradual Drifts          | 10         | 4       | RandomRBF[7]   |
| RBF_Recc+Abrpt   | Recurring with Abrupt Drifts           | 10         | 4       | RandomRBF      |
| Electricity[8]   | Real-world                             |            |         |                |
| Sensor[9]        | Real-world                             |            |         |                |
| Weather[10]      | Real-world                             |            |         |                |

Table 2. Predictive accuracy in (%)

| Dataset             | DWM   | DWMBag | WMA | WMABag | DACC | DACCBag | EB   | EBBag |
|---------------------|-------|--------|------|--------|------|---------|------|-------|
| MIXED_Incrmntl      | 96.27 | 96.84  | 95.21| 79.00  | 89.16| 90.09   | 89.59| 86.88 |
| MIXED_Abrpt         | 96.19 | 96.88  | 95.26| 79.00  | 89.07| 90.13   | 88.44| 86.81 |
| WAVE_Grl            | 79.57 | 80.30  | 75.97| 76.78  | 72.17| 75.17   | 78.21| 79.51 |
| WAVE_Abrpt          | 81.02 | 81.98  | 76.19| 76.92  | 74.41| 77.97   | 79.33| 80.93 |
| RBF_Recc+Grl       | 89.60 | 87.28  | 88.43| 86.81  | 63.16| 68.88   | 79.78| 70.96 |
| RBF_Recc+Abrpt     | 92.09 | 90.79  | 76.55| 89.82  | 75.09| 80.82   | 80.80| 79.79 |
| Electricity         | 77.58 | 76.79  | 78.39| 73.02  | 70.66| 72.27   | 69.45| 68.68 |
| Sensor              | 79.04 | 75.61  | 39.67| 58.12  | 76.78| 78.36   | 59.16| 56.25 |
| Weather             | 72.34 | 72.53  | 71.61| 71.67  | 71.76| 64.62   | 68.90| 70.45 |

according to the output performance rendered. By this weighting strategy best of learners participate in decision making [4].

- **Dynamic Weighted Majority (DWM):** Authors Kolter and Maloof [5] proposed this algorithm which adjusts the weight of the ensemble learners according to the instance environment. Being a prediction problem, the algorithm updates itself on arrival of concept drift. It follows incremental learning perspective which involves no dependence over the learners.

- **Dynamic Adaption to concept change (DACC):** Jaber et al. [6] presented this approach that was inspired by DWM. In this approach a worst half section of ensemble learners are located, after that any learner picked out randomly is removed from the selected learners. This approach where best of learners with high accuracy form a part of current ensemble, abrupt and gradual drifts are handled well here.

- **Ensemble Building (EB):** Ramamurthy and Bhatnagar [11] proposed this chunk based approach of ensemble building. Accuracy of learners is calculated based upon the current chunk of instances. A specific threshold is defined which monitors the correctness of ensemble members. New learners are built time to time to support constant updation. This algorithm has been found to be good for abrupt and gradual drifts.

3. Proposed Work

In this paper four concept drift algorithms are considered and randomization is added to the input space of these approaches. Online bagging is used to generate the weight
which adds diversity to the training samples of the incoming stream of data. Bagging proposed by Bifet et al. [12] uses a λ value given by Poisson distribution. Lower and higher values of lambda can be used to modify the weights of the training instances. The experimental setting in this paper studies the effect of adding bagging to WMA, DWM, DACC and EB. The prediction results shown in Section 3.3 compares the performance of the above approaches with and without bagging. Four different drift forms gradual, abrupt recurring and real world datasets are included in this study. The recurring concepts are monitored both under gradual and sudden drift patterns.

3.1. Implementation details

The Massive Online Analysis (MOA) [7] software of streaming applications is used for executing the experiments. The machine used has Intel(R) Core(TM) i7-7700HQ CPU @2.80GHz with 8 GB RAM configuration. Default implementations of the considered algorithms is considered for the study. WMABag, DWMBag, DACCBag and EBBag are the four variants of the original algorithms which are implemented in MOA itself. In these variants the conventional training is replaced by the randomized training using bagging.

3.2. Datasets

The datasets used in this work belong to different generators provided by the MOA [7] software. The description is given in Table 1. Different drift patterns are mentioned to test the effect of proposed work.

3.3. Results and Discussion

This section discusses the predictive results obtained over various generators under different concept drift forms. The result figures have been separated under major four categories namely gradual drifts, abrupt, recurring concepts and two real datasets. In case of first case of gradual drifts two generators Mixed [7] and Wave ar considered whose results are shown in Fig. 1(a) and Fig. 1(b) respectively. In case of Mixed, DWM did not show much increase in accuracy due to bagging. WMA rather has shown steep fall due to randomization. DACC is the only algorithm to have witnessed increase in accuracy with the incorporation of bagging. For Wave, DWM, WMA, DACC and EB all four algorithms see an increase in their respective accuracy performance due to online bagging. In case of abrupt drift pattern, for the Mixed generator in Fig. 1(c) and Wave in Fig. 1(d) analysis is same as that in case of gradual drifts. Next set of plots depicts the occurrence of repeating concepts which are drifted gradually and abruptly as shown in Fig. 1(e) and Fig. 1(f) respectively. In case of gradually moving concepts, DWM, WMA, DACC have shown increase in accuracy with bagging while EB is shown opposite trend. In case of abrupt drifts, DWM, WMA and DACC show increase while EB shows decrease with randomization adding to ensemble learning. In this case an
4. Conclusion

This paper presents the analysis of effect of adding randomization to the ensemble learning with the use of online bagging technique. The proposed work has considered wide patterns of drift under varied number and generators. Bagging does not lead to
uniform increase or decrease of predictive accuracy. For some algorithms it has proved to be quite beneficial while for others it did not. Future works shall focus on exploring and implementing other techniques for improving ensemble accuracy and for greater range of concept drift forms.

References

[1] Rodrigo F de Mello, Ricardo A Rios, Paulo A Pagliosa, and Caio S Lopes. Concept drift detection on social network data using cross-recurrence quantification analysis. Chaos: An Interdisciplinary Journal of Nonlinear Science, 28(8):085719, 2018.

[2] Ioannis Katakis, Grigorios Tsoumakas, Evangelos Banos, Nick Bassiliades, and Ioannis Vlahavas. An adaptive personalized news dissemination system. Journal of Intelligent Information Systems, 32(2):191–212, 2009.

[3] João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. ACM Computing Surveys (CSUR), 46(4):44, 2014.

[4] Avrim Blum. Empirical support for winnow and weighted-majority algorithms: Results on a calendar scheduling domain. Machine Learning, 26(1):5–23, 1997.

[5] J Zico Kolter and Marcus A Maloof. Dynamic weighted majority: An ensemble method for drifting concepts. Journal of Machine Learning Research, 8(Dec):2755–2790, 2007.

[6] Ghazal Jaber, Antoine Cornuéjols, and Philippe Tarroux. Anticipative and dynamic adaptation to concept changes. Real-World Challenges for Data Stream Mining, page 22, 2013.

[7] Albert Bifet, Geoff Holmes, Richard Kirkby, and Bernhard Pfahringer. Moa: Massive online analysis. Journal of Machine Learning Research, 11(May):1601–1604, 2010.

[8] Michael Harries and New South Wales. Splice-2 comparative evaluation: Electricity pricing. 1999.

[9] Zahra Ahmadi and Hamid Beigy. Semi-supervised ensemble learning of data streams in the presence of concept drift. In International Conference on Hybrid Artificial Intelligence Systems, pages 526–537. Springer, 2012.

[10] Ryan Elwell and Robi Polikar. Incremental learning of concept drift in nonstationary environments. IEEE Transactions on Neural Networks, 22(10):1517–1531, 2011.

[11] Sasthakumar Ramamurthy and Raj Bhatnagar. Tracking recurrent concept drift in streaming data using ensemble classifiers. In Sixth International Conference on Machine Learning and Applications (ICMLA 2007), pages 404–409. IEEE, 2007.

[12] Nikunj C Oza. Online bagging and boosting. In Systems, man and cybernetics, 2005 IEEE international conference on, volume 3, pages 2340–2345. IEEE, 2005.