Deferral Classification of Evolving Temporal Dependent Data Streams

Michael Mayo
Department of Computer Science
University of Waikato
New Zealand
mmayo@waikato.ac.nz

Albert Bifet
LTCI, CNRS, Télécom ParisTech
Université Paris-Saclay
75013, Paris, France
albert.bifet@telecom-paristech.fr

ABSTRACT

Data streams generated in real-time can be strongly temporally dependent. In this case, standard techniques where we suppose that class labels are not correlated may produce sub-optimal performance because the assumption is incorrect. To deal with this problem, we present in this paper a new algorithm to classify temporally correlated data based on deferral learning. This approach is suitable for learning over time-varying streams. We show how simple classifiers such as Naive Bayes can boost their performance using this new meta-learning methodology. We give an empirical validation of our new algorithm over several real and artificial datasets.

Keywords

data streams, classification, temporal dependence

1. INTRODUCTION

Nowadays, *Big Data* applications can be found in many diverse fields that require deep insight, such as financial market data, energy data, and many others [5]. These sources of data generated in real-time may have a strong dependence from one instance to the next: instances labeled with a specific class label may be more likely to be followed by instances with the same class label.

In this type of data, we can observe that there is short term memory in the stream, i.e. in the case of binary classification, the probability of a positive or negative is more or less dependent on the instance that has just been observed. This seems to be the case in financial and energy related market data streams.

It is very important to acknowledge this type of temporal dependence, since the performance of the classifier depends on it. A classifier dealing with temporal dependent data should always be compared with the *no-change* classifier, i.e. a classifier that simply predicts the last class label seen on the stream. This is due to the fact that this very simple classifier, in datasets with strong temporal dependence may be simply more accurate than a more complex classifier such as a decision tree or Naive Bayes [3].

In this paper we present a new meta classifier that can boost the performance of any classifier to be able to deal with this temporal data dependency. We then run an empirical evaluation to show its benefits. The paper is structured as follows: in Section 2 we discuss related work, and in Section 3 we present our new meta classifier. We perform and discuss an experimental evaluation in Section 4. We end the paper giving some conclusions in Section 5.

SAC 2016, April 04 - 08, 2016, Pisa, Italy

ISBN 978-1-4503-3739-7/16/04. . . $15.00

DOI: http://dx.doi.org/10.1145/2851613.2851890

2. RELATED WORK

The temporal dependency of data has been studied in time series [4], and in regression [10].

In classification, a solution based on using the non-change classifier was used in the winning solution of the EUNITE2003 competition [11], which was about predicting glass quality from a glass production line. The problem needed multi-step prediction into the future, and the winners used linear interpolation starting from the last value (as the prediction for the first value at $t + 1$) to the global mean (as the prediction for the value at $t + 20$, the one furthest into the future). This was described as the *Naive Rule*.

Zliobaite [12] detected the temporal dependence component in the Electricity Dataset. The Electricity Dataset due to [7] is a popular benchmark for testing adaptive classifiers. It has been used in over 40 concept drift experiments [6, 8, 2, 9]. The class label identifies the change of the price (UP or DOWN) related to a moving average of the last 24 hours, and is subject to concept drift due to changing consumption habits, unexpected events and seasonality.

The *Kappa Plus Statistic* measure was proposed in [3] to measure the performance of a classifier with temporal dependence. As the standard Kappa Statistic normalizes the accuracy of a classifier with the performance of a majority class based classifier, this new measure, normalizes the accuracy of a classifier against the no-change classifier.

3. DEFERRAL CLASSIFIER

In this section we propose a new algorithm to classify evolving data streams, assuming that the performance of our new classifier is going to be at least as good as that of the no-change classifier.

Our new *deferral classifier* is based on this pseudo-code:

- Each time a new instance arrives, make a prediction using a data stream classifier
  - If the prediction is sufficiently certain (based on a probability threshold t) then accept it;
  - If the prediction is not sufficiently certain, default to predicting the outcome of the last instance

We further extend this approach with two new algorithms:

- the first algorithm tunes the threshold parameter t automatically as the stream is processed, keeping records of the accuracy over a sliding window for different values of t (such as

$0.1, 0.2, \ldots, 1.0$); the value of t that historically would have produced the lowest error is chosen

- the second algorithm uses a consensus of the class labels over the last n instances instead of the single most recent class label.

The first algorithm is implemented by computing an exponential average on the error for each possible threshold t. It simply always sets t to whatever value would have historically resulted in the with lowest average error.

We implemented the second algorithm as a meta-classifier that has a parameter α, which maintains a running numeric "vote" or prediction for each class. When a new instance arrives, the meta classifier multiplies all votes by $(1 - \alpha)$, and adds α to the vote of the correct class of the instance. This approach has the advantage that it is identical to a classifier that predicts the last class label observed when $\alpha = 1$, but when $\alpha < 1$, it gives more weight to previous instances beyond the most recent.

4. EXPERIMENTAL EVALUATION

In this section, we perform two evaluations to compare the new classification schema with previous state-of-the art strategies:

- comparison with standard real and artificial datasets
- comparison with artificial streams generated adding strong temporal dependency

Massive **O**nline **A**nalysis (MOA) [1] is a software environment for implementing algorithms and running experiments for online learning from data streams. All algorithms evaluated in this paper were implemented in the Java programming language by extending the MOA software. The synthetic datasets were generated using MOA generators, and the real-world datasets used are available from the MOA website.

4.1 Results

We ran an experimental evaluation to test our new deferral classifier. We compare the original Naive Bayes, with the following classifiers:

- deferral classifier with $\alpha = 1$,
- deferral classifier with $\alpha = 0.5$,
- temporal augmented classifier [3], where the class label of the previous instance is used as an additional attribute.

We use the datasets introduced previously for evaluation. The experiments were performed on 2.66 GHz Core 2 Duo E6750 machines with 4 GB of memory.

The evaluation methodology used was Interleaved Test-Then-Train: every example was used for testing the model before using it to train. This interleaved test followed by train procedure was carried out on one million examples from the hyperplane and RandomRBF datasets. The parameters of these streams are the following:

- RBF(x,v): RandomRBF data stream of 5 classes with x centroids moving at speed v.
- HYP(x,v): Hyperplane data stream of 5 classes with x attributes changing at speed v.

We report the following measures based on the accuracy of the classifiers compared with very simple classifiers:

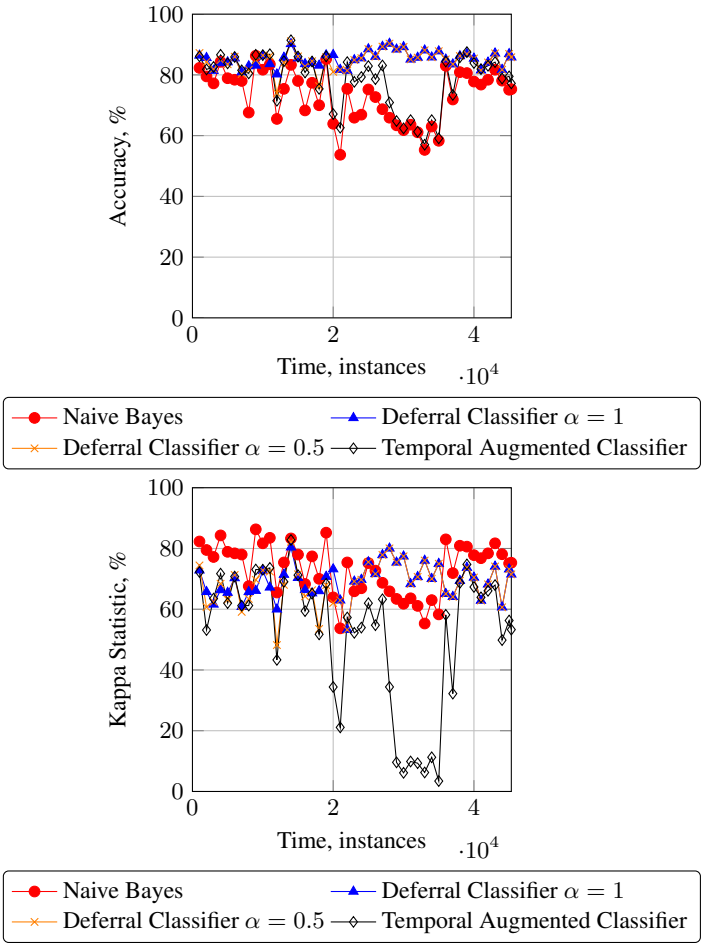

Figure 1: Accuracy and Kappa Statistic on the Electricity Market Dataset

- final accuracy p_0
- $\kappa = (p_0 - p_c)/(1 - p_c)$ where p_c is the probability that the classifiers predicts correctly by chance
- $\kappa^+ = (p_0 - p'_e)/(1 - p'_e)$ where p'_e is the accuracy of the classifier that predicts using the label of the last instance seen.

We plot the learning curves for the ELECTRICITY dataset in Figure 1 and for the FOREST COVERTYPE dataset in Figure 2. We observe that the two deferral classifier performs similarly, but with higher accuracy and κ^+ than the single Naive Bayes and the temporal augmented Naive Bayes.

Tables 1 reports the performance of the classification models induced on the synthetic data and the real datasets: FOREST COVERTYPE, POKER HAND, ELECTRICITY and COVPOKELEC. The performance is measured as the final percentage of examples correctly classified over the test/train interleaved evaluation.

We see that the deferral classifiers are superior to the Temporal Augmented Classifier and the Naive Bayes classifier alone. On Table 1, we see that classifier have positive values over artificial datasets, as they don't have strong temporal dependence, but they have in some cases negative values to indicate that the performance is not good. We see that the classifier with better Kappa Plus Statistic is the deferral classifier using $\alpha = 1$.

	Naive Bayes	Deferral Classifier $\alpha = 1$	Deferral Classifier $\alpha = 0.5$	Temporal Augmented Classifier
CovType	-699.37	0.14	-10.65	-190.18
Electricity	-81.59	-0.84	-3.29	-46.70
Poker	-58.84	-40.54	-44.55	0.75
CovPokElec	-338.46	-192.05	-203.62	-279.35
HYP(10,0.001)	41.83	41.83	41.83	41.82
HYP(10,0.0001)	82.50	82.50	82.50	82.50
RBF(0,0)	36.87	36.85	36.86	36.87
RBF(50,0.001)	8.31	8.01	8.02	8.29
RBF(10,0.001)	37.85	37.85	37.85	37.83
RBF(50,0.0001)	10.72	10.16	10.24	10.71
RBF(10,0.0001)	38.03	38.02	38.03	38.02
Average	-83.83	1.99	-0.62	-23.58

Table 1: Kappa Plus statistic. Higher is better. A positive value indicates that on average the classifier outperforms the no-change classifier.

5. CONCLUSIONS

In this paper we presented a new deferral classifier, to address the problem of temporal dependence on evolving data streams. We showed the benefits of the new method running an empirical evaluation over several datasets, using our method as a meta-classifier over the Naive Bayes classifier.

As future work, we would like to continue studying this problem in more depth, and try to apply these techniques to the more challenging setting of evolving data stream multi-label classification, where the number of labels is not fixed, and the probability distribution that is generating the data may be evolving.

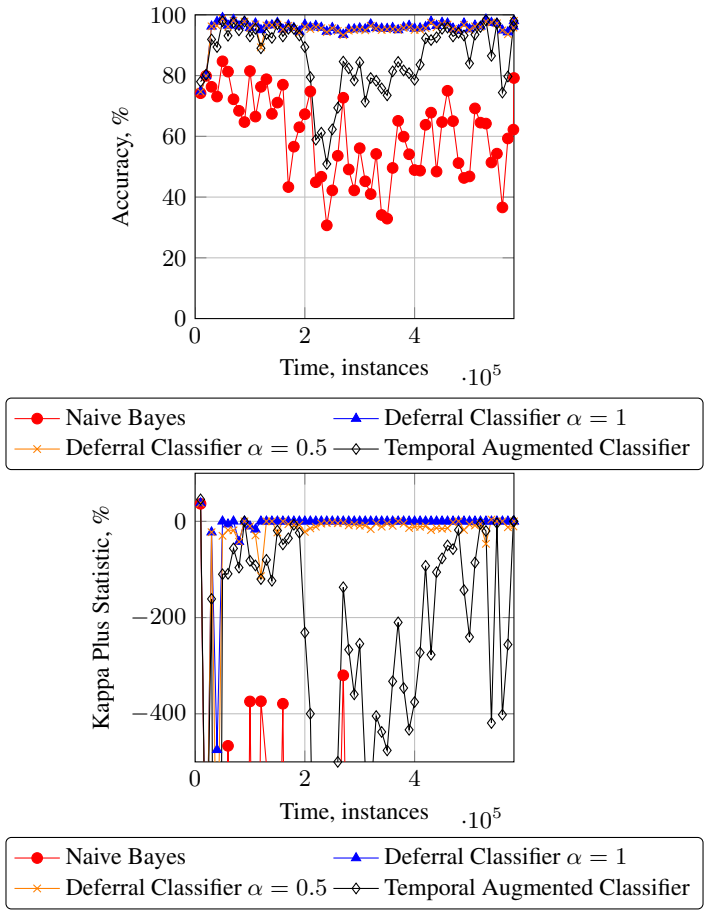

Figure 2: Accuracy and κ^+ on the Forest Covertype dataset

6. REFERENCES

[1] A. Bifet, G. Holmes, R. Kirkby, and B. Pfahringer. MOA: Massive online analysis. *J. of Mach. Learn. Res.*, 11:1601–1604, 2010.

[2] A. Bifet, G. Holmes, B. Pfahringer, R. Kirkby, and R. Gavaldà. New ensemble methods for evolving data streams. In *KDD*, pages 139–148, 2009.

[3] A. Bifet, J. Read, I. Zliobaite, B. Pfahringer, and G. Holmes. Pitfalls in benchmarking data stream classification and how to avoid them. In *ECMLPKDD*, pages 465–479, 2013.

[4] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 3rd edition, 1994.

[5] K. Cukier. Data, data everywhere. The Economist Report, 2010.

[6] J. Gama, P. Medas, G. Castillo, and P. Rodrigues. Learning with drift detection. In *Proc. of the 7th Brazilian Symp. on Artificial Intelligence*, SBIA, pages 286–295, 2004.

[7] M. Harries. SPLICE-2 comparative evaluation: Electricity pricing. Tech. report, University of New South Wales, 1999.

[8] J. Kolter and M. Maloof. Dynamic weighted majority: An ensemble method for drifting concepts. *J. of Mach. Learn. Res.*, 8:2755–2790, 2007.

[9] G. Ross, N. Adams, D. Tasoulis, and D. Hand. Exponentially weighted moving average charts for detecting concept drift. *Pattern Recogn. Lett*, 33:191–198, 2012.

[10] G. Seber and C. Wild. *Nonlinear Regression*. Wiley Series in Probability and Statistics. Wiley, 2003.

[11] M. Wojnarski. Prediction of product quality in glass manufacturing process using LTF-A neural network.On Bagging and Nonlinear Estimation. Technical report, EUNITE Competition, 2003.

[12] I. Zliobaite. How good is the electricity benchmark for evaluating concept drift adaptation. *CoRR*, abs/1301.3524, 2013.