

Drift Severity Metric

Petr Kosina¹ and João Gama² and Raquel Sebastião³

Abstract. Concept drift in data is usually considered only as abrupt or gradual thus referring to the speed of change. Such simple distinguishing by speed is sufficient for most of the problems, but there might be situations for which a finer representation would be of use.

This paper studies further the phenomenon of concept drift and introduces a simple measure which is relevant to the speed and amount of change between different concepts.

Keywords: Data Streams, Concept Drift, Drift Rate

1 Motivation

In many practical situations where there is a vast amount of data and it flows continuously, such as in biomedicine, fraud detection etc., there is a need for tools that are taking the change detection into account. Such applications are of high importance and interest. Therefore the change in data has been subject of research in many works that are dealing with data stream mining, mostly in a pursuit of fast and adequate reaction.

The process of change does not need to be the same and the ability and complexity to detect a difference in the underlying concept varies. In this paper we are interested in the speed and amount of change between different concepts in data. We introduce a simple method for measuring the change that can describe the difficulty to recognize concept drift with respect to the number of examples needed to be seen before the detection. Such an indicator may be of use during reaction to concept drift when selecting classifier, either single or as a part of ensemble, or possibly setting a sensitivity of drift detection method.

2 Related Work

Concepts drift is a topic in many existing works dealing with data stream mining [3, 7]. In [7], the phenomenon of concept drift is defined as number of instances Δx between two stable concepts. When $\Delta x = 1$ drift is instantaneous and if $\Delta x > 1$ the concept is changing over a number of instances. The higher Δx the more gradual (slower) the drift is. Besides instantaneous drift, moderate and slow drifts as subsets of gradual drift are presented.

The rest of the related work can be grouped into two main categories: drift detection methods and severity and rate of drift definitions and works.

Drift Detection Methods.

DDM. The first drift detection method used in experiments in this paper was DDM [3]. It assumes that if the distribution of the examples is stationary, the error-rate of the learner will decrease when the number of examples increases. Description can be found in [3].

Page-Hinkley Test. Another drift detection method is based on Page-Hinkley (PH) statistics, described in [6], commonly used for change detection in signal processing. For the purpose of measuring the proposed metric, two different threshold values, as opposed to one in the original approach, were set. The first, λ_w , signals the warning phase and the second, λ_d , signals the drift in data.

Rate and Severity of Drift.

Drift rate. The drift rate in [1] is defined as the probability of two consecutive concept functions to provide different label on random example $Prob(f_t \neq f_{t+1})$, where $f_t = X \rightarrow \{0, 1\}$ assigns the label 0 or 1 to the referred example x_t from some domain X at time t . **Severity.** In the paper [5] we can encounter the severity of drift. Suppose examples of context C_i and examples of the following context C_{i+1} are from some input space S . During the drift examples from both contexts, C_i and C_{i+1} , can appear. The severity of drift is defined as the percentage of S that has its class label changed after the drift is complete, e.g. examples of C_{i+1} are emitted.

3 Proposed Metric

The drift rate of evolving data significantly affects the detection difficulty. For some artificial datasets it is possible to obtain some characteristics, like compute the severity or the information about the speed is already known. But not many works deal with providing more detailed information about the event and we lack measure that could be used for different datasets.

Considering the speed, drifts are commonly divided into two basic groups, abrupt and gradual. In abrupt drift the amount of change is immediately present whereas in gradual drift it is distributed over longer period of time e.g very small amount of immediate change. A value describing the change may be of use in studying and comparing behaviour of algorithms on different datasets and could reveal some interesting characteristics. This measure shows how difficult is for the system to recognize that there was a drift.

The drift detection methods from the previous part have two states relevant for the metric. The first one is a warning state during which we keep simple statistics about the number of processed examples and the number of misclassified ones. Once we reach the second state e.g. drift is detected, we easily compute the metric as $\frac{\#misclassified}{\#examples_in_warning}$. Only statistics from warning followed by drift are taken into account, therefore the examples which caused false alarm are not included in the computation.

4 Experiments

In the experiments the drift detection methods were used with two different classifiers. The first one was an incremental version of Naive Bayes and Hoeffding Tree [2] was the second. The settings of the Hoeffding Tree classifier were following: $\delta = 10^{-6}$, $\tau = 0.05$,

¹ Fac. of Informatics, Masaryk University, Brno email: soso@mail.muni.cz

² LIAAD-INESC Porto, FEP-University of Porto email: jgama@fep.up.pt

³ LIAAD-INESC Porto, FC - University of Porto email: raquel@liaad.up.pt

where δ was the split confidence constant and τ the tie breaking constant. Numeric attributes were handled by binary tree structures [2].

The PH test detection was used only with Naive Bayes classifier since slower learning process of Hoeffding Tree in some of the datasets disallowed its usage. The values of PH test thresholds were set to $\lambda_d = 2.5$, and $\lambda_w = 2$, e.g. the value of λ_w was 80 % of λ_d .

Experimental Methodology. The severity of drift in SEA dataset is computed as follows. The 2 relevant attributes of the data defines 2-dimensional area. The threshold line divides the area into 2 parts, one for each class. By subtracting the areas defined by 2 following thresholds we obtain the area affected by drift.

Datasets. The 9 artificial datasets used in the experiments. SEA was generated based on [8], STAGGER and LED was generated by [4], and other 6 were taken from the [3].

Results. The SEA datasets, that were generated with different thresholds Θ , contained additional information to examine - the area difference of two concepts represented by the severity. The proposed metric does not measure the severity, however it is highly correlated to it as could be seen in Tab. 1.

Considering speed and severity the warning length indicates the severity. It is longer in low and shorter in high severity data. The measure is influenced by both, but the dependency is stronger for the severity, as illustrated on Fig.1 with 2 abrupt and 2 gradual drifts (change over 100 examples) with different severities.

Table 1. SEA Concepts results showing the dependency of the metric to the severity in two datasets generated with the same threshold Θ values.

Severity	Measured Value			Θ	
	Hoeff. Tree	Naive Bayes	NB with PH		
3.125	-	0.15298	0.16431	6	6.5
	-	0.15206	0.12		
4.725	-	-	-	9	8.5
	-	0.13034	-		
9	-	0.18715	0.20792	8.5	9.5
	0.17755	0.18966	0.18396		
11.625	0.18538	0.18557	0.15589	8.5	7
	0.19795	0.17957	0.15601		
13.125	0.20737	0.21883	0.24324	8	9.5
	0.21964	0.21418	0.22436		
15	0.25545	0.25440	0.25000	6.5	8.5
	0.24650	0.25549	0.28966		
16	0.25710	0.26804	0.22819	7	9
	0.18538	0.27254	0.27684		
24	0.34591	0.31944	0.26772	9.5	6.5
	0.29712	0.29620	0.28467		
27.125	0.34383	0.29018	0.26050	9.5	6
	0.29541	0.26596	0.27397		
corr	0.90345	0.90766	0.78851		

Datasets from [3] had different number of concepts but each concept had 1,000 examples. 4 datasets based on sine function contained abrupt change and their class label was reversed, therefore it had high severity. The measure was very high meaning the change was detected very quickly and easily (results omitted due to space limitation).

LED and STAGGER dataset results (omitted due to space limitation) also confirmed expected - the larger the difference between concepts the higher the value of the metric.

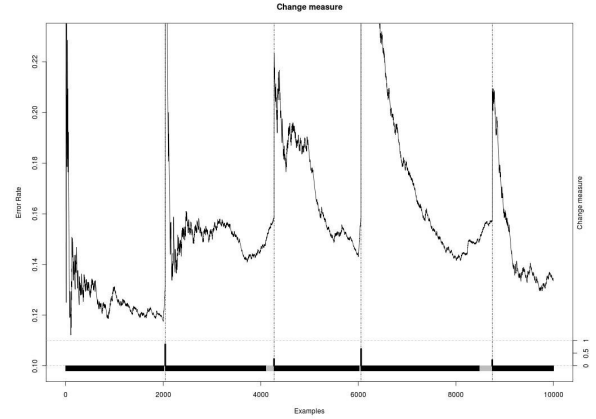


Figure 1. Two sudden drifts (at 2,000 and 4,000) and two gradual (6,000 and 8,000) with severities (75 % and 15 %). The bar on the bottom is black during stationary learning, grey in warning and height of the bar at drift represents the change measure. It indicates small difference between speeds, bigger between severities.

5 Conclusions

In this paper we introduced a metric that was a simple, yet effective, indicator of amount and speed of change between different concepts in data. We experimentally proved on artificial datasets that the higher the measured value was the bigger the change between concepts was, which was also supported by high correlation of metric to the severity of drift, and the easier it was to recognize the drift.

This measure provides valuable information for users about the characteristics of the changes that are hidden in their data.

Acknowledgements

This work was supported by Research Project of FI MUNI and KDUS (PTDC/EIA-EIA/098355/2008). The work of Raquel Sebastião is supported by the Portuguese Foundation for Science and Technology (FCT) under the PhD Grant SFRH/BD/41569/2007.

REFERENCES

- [1] Bartlett P. L., Ben-David S., and Kulkarni S. R. *Learning Changing Concepts by Exploiting the Structure of Change*. Mach. Learn. 41, 2 (Nov. 2000), 153–174.
- [2] Gama J., Fernandes R., and Rocha R. *Decision Trees for Mining Data Streams*. Intelligent Data Analysis, IOS Press, pp. 23–45, 2006.
- [3] Gama J., Medas P., Castillo G., and Rodrigues P. *Learning with Drift Detection*. Proc. of the 17th Brazilian Symposium on Artificial Intelligence, Vol. 3171, pp. 286–295 - October 2004.
- [4] Kirkby R. *Massive Online Analysis*. University of Waikato, Hamilton, New Zealand, 2007. <http://sourceforge.net/projects/moa-datastream/>
- [5] Minku F. L. and Yao X. *Using diversity to handle concept drift in on-line learning*. In Proc. of the 2009 international Joint Conference on Neural Networks (Atlanta, Georgia, USA, 2009). IEEE Press, Piscataway, NJ, 2968–2975.
- [6] Page E. S. *Continuous inspection schemes*. Biometrika, 41(1/2):100–115, 1954.
- [7] Stanley K. O., *Learning concept drift with a committee of decision trees*. Tech. Report UTAI- TR-03-302, Department of Computer Sciences, University of Texas at Austin, USA, 2003.
- [8] Street W. N., and Kim Y. *A streaming ensemble algorithm SEA for large-scale classification*. In Proc. 7th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 377–382. ACM Press, 2001.