LEARNING PARAMETER DISTRIBUTIONS TO DETECT CONCEPT DRIFT IN DATA STREAMS

Johannes Haug, Gjergji Kasneci
Data Science and Analytics, University of Tuebingen, Germany

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

BACKGROUND & OBJECTIVES

Streaming data is potentially unbounded and underlies distributional change, called concept drift. To maintain high predictive performance, we aim to detect concept drift …

- … both globally and w.r.t. specific input features,
- … with a short delay,
- … while avoiding false alarms.

INTUITION & METHODOLOGY

We train the parameters of a predictive model at each time step w.r.t. the active concept. Hence, **changes** in the distribution (uncertainty) of **optimal parameters** $\theta \sim P(\theta;\psi)$ may indicate **real concept drift**.

Inspired by [1], we model concept drift between time step t and u as a difference in the marginal likelihood w.r.t. ψ_t and ψ_u. This may be expressed in terms of the differential entropy h and KL-divergence D_{KL}:

$$|h[P(\theta;\psi_u)] - h[P(\theta;\psi_t)] + D_{KL}[P(\theta;\psi_u)||P(\theta;\psi_t)]| > 0 \quad (1)$$

For more robust and continuous drift detection, we investigate the moving average of (1) in a shifting window. Besides, we define a dynamic threshold.

Our framework **ERICS** is **model-agnostic**.

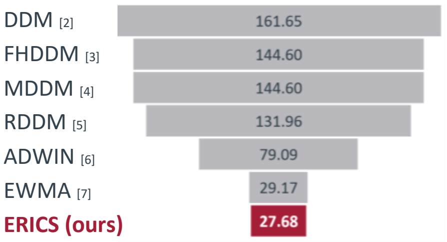
EXPERIMENTS & RESULTS

For the evaluation of ERICS, we used a Probit model with Gaussian parameters $P(\theta;\psi_t) = N(\mu_t, \Sigma_t)$.

We compared ERICS to **6 related methods**, using **4 synthetic** and **6 real-world data sets**. We trained a Hoeffding Tree to provide the related methods with predictions.

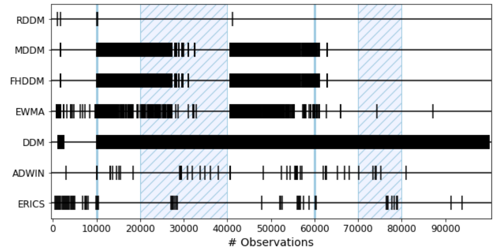
Delay

We measured the avg. delay across data sets (in no. of batches) of each related method:



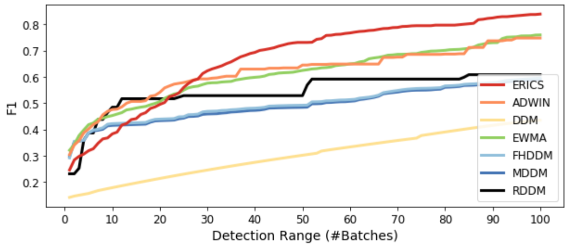
Drift Alerts (Example)

Drift alerts (black lines) for a synthetic data set with sudden drifts (blue lines) and gradual drifts (blue areas):



Precision and Recall

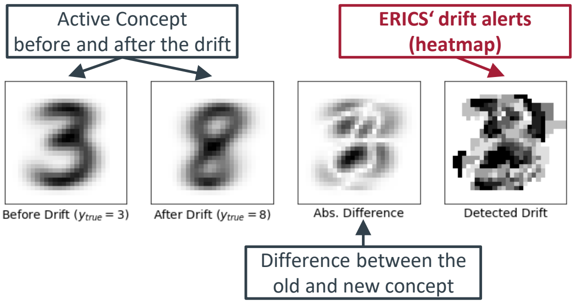
We monitored the precision and recall of each concept drift method w.r.t. different detection ranges (i.e. permitted delay). Below, we show the harmonic mean of precision and recall, i.e. the F1 score:



Since the distribution of the Probit parameters does not change immediately after a concept drift, ERICS required more observations before eventually achieving **higher recall and precision than related methods**.

Partial Drift Detection

With MNIST as illustrative example, we investigated the partial drift detection capability of ERICS:



[1] Haug, Johannes, et al. "Leveraging Model Inherent Variable Importance for Stable Online Feature Selection." Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020.
[2] Gama, Joao, et al. "Learning with drift detection." Brazilian symposium on artificial intelligence. Springer, Berlin, Heidelberg, 2004.
[3] Pesaranghader, Ali, and Herna L. Viktor. "Fast hoeffding drift detection method for evolving data streams." Joint European conference on machine learning and knowledge discovery in databases. Springer, Cham, 2016.
[4] Pesaranghader, Ali, Herna L. Viktor, and Eric Paquet. "McDiarmid drift detection methods for evolving data streams." 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, 2018.
[5] Barros, Roberto SM, et al. "RDDM: Reactive drift detection method." Expert Systems with Applications 90 (2017): 344-355.
[6] Bifet, Albert, and Ricard Gavalda. "Learning from time-changing data with adaptive windowing." Proceedings of the 2007 SIAM international conference on data mining. Society for Industrial and Applied Mathematics, 2007.
[7] Ross, Gordon J., et al. "Exponentially weighted moving average charts for detecting concept drift." Pattern recognition letters 33.2 (2012): 191-198.

https://github.com/haugjo/erics https://uni-tuebingen.de/dsar johannes-christian.haug@uni-tuebingen.de