Meet-up 2019 | Doctorants & Industrie

DATASET SHIFT BLACK-BOX MONITORING

Clément FEUTRY[1,2,3,4]

Pierre Duhamel[1,2,3,4], Pablo Piantanida [1,2,3], Florence Alberge[1,2,3,4]

[1] Université Paris-Sud [2] Laboratoire des Signaux et Systèmes [3] CentraleSupélec [4] CNRS

Abstract

The work presented here aims at monitoring the dataset shift in the context of a black box predictor. The only information accessible is the soft-probabilities and the training set. The data to be tested is handled in batch which is compatible with a data stream context. Two mathematically motivated tools are introduced and tested to monitor the shift.

1 CONTEXT

- Objective: A classifier has been designed based on training samples. Monitor outputs to detect discrepancy between test samples and training samples.
- Assumption: black box model with soft probabilities outputs.
- Method: Statistical studies of empirical distribution of predicted labels given a input signal.
- Note: we are not exactly checking "dataset shift", instead we are checking if the classifier is still adapted to the test samples. (this is what a user wants to check)

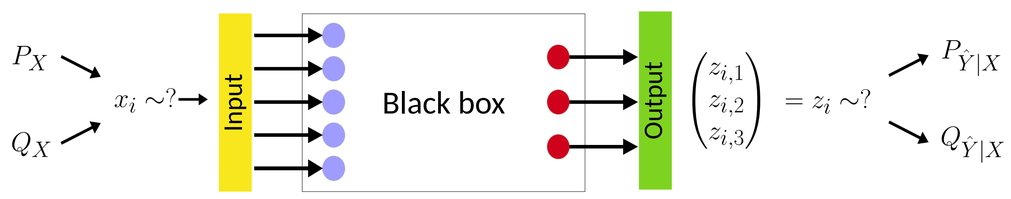

Figure 1: Presentation of the problem

2 CHALLENGES

Detect this misadaptation with as few samples as possible. The method should not rely on the precise structure of the classifier.

3 CONTRIBUTION

Obtain a clean demonstration that the following quantities contain the required information : standard deviation (SD) and geometric mean (GM):

$$SD = \mathbb{E}_X \left[\sum_{y \in \mathcal{Y}} \left(z_i - \frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} z_i \right)^2 \right]$$

$$GM = \sqrt[n]{\prod_i^n z_i}.$$

4 APPLICATION

Detect changes in the data, therefore allowing a reliable estimate of when a new training has to be done; detect outliers if the method is efficient enough; easy plugin method, can be used without digging in the training program, and on top of an already trained classifier.

5 RESULTS

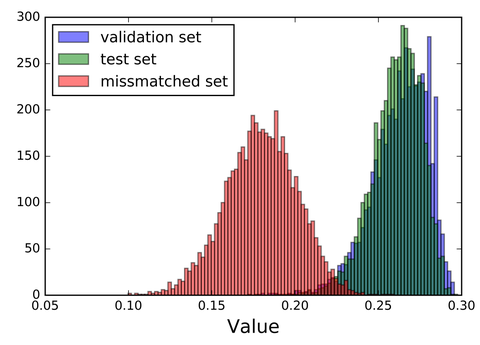

Figure 2: SD statistics of the samples for a batch of size 9. Classifier is trained on the SVHN database. Mismatch set is from the CIFAR10

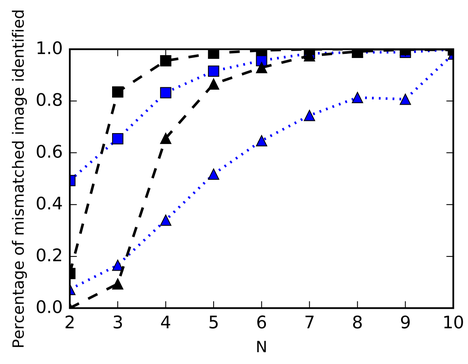

Figure 3: Mismatched data detection rate as a function of the batch size. Model trained on CIFAR-10. Mismatched data fromSVHN database. Squares: 5%; Triangles: 0.5% FPR. em-pirical GM (blue dotted); empirical SD (black dashed)

Visualisation of the statistics on the left and example of results on the right.

6 FUTURE WORK

Find improved criteria (SD, GM); consider a white-box context; check the performance in a variety of applications.

References

[1] David A Cieslak and Nitesh V Chawla. A framework for monitoring classifiers' performance: when and why failure occurs? *Knowledge and Information Systems*, 18(1):83–108, 2009.

[2] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *ICLR*, 2017.

[3] Heinrich Jiang, Been Kim, Melody Y. Guan, and Maya R. Gupta. To trust or not to trust A classifier. In *NeurIPS 2018, 3-8 December 2018, Montréal, Canada.*, pages 5546–5557, 2018.

[4] Shiyu Liang, Yixuan Li, and R Srikant. Principled detection of out-of-distribution examples in neural networks. In *ICLR*, 12 2018.

[5] Zachary C Lipton, Yu-Xiang Wang, and Alex Smola. Detecting and correcting for label shift with black box predictors. *arXiv preprint arXiv:1802.03916*, 2018.

[6] Jose G Moreno-Torres, Troy Raeder, RocíO Alaiz-RodríGuez, Nitesh V Chawla, and Francisco Herrera. A unifying view on dataset shift in classification. *Pattern Recognition*, 45(1):521–530, 2012.

[7] Gabi Shalev, Yossi Adi, and Joseph Keshet. Out-of-distribution detection using multiple semantic label representations. In *Advances in Neural Information Processing Systems*, pages 7375–7385, 2018.

[8] Masashi Sugiyama and Motoaki Kawanabe. *Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation*. The MIT Press, 2012.

[9] Gerhard Widmer and Miroslav Kubat. Learning in the presence of concept drift and hidden contexts. *Machine Learning*, 23(1):69–101, Apr 1996.

ED STIC Paris-Saclay
Université Paris-Sud
Clément Feutry- clement.feutry@ens-cachan.fr