Mozafari, Niloofar; Hashemi, Sattar; Hamzeh, Ali

Summary: Recently data streams have been extensively explored due to their emergence in a great deal of applications such as sensor networks, web click streams and network flows. One of the most important challenges in data streams is concept change where data underlying distributions change from time to time. A vast majority of researches in the context of data stream mining are devoted to labeled data, whereas, in real world human practice labels of data are rarely available to learning algorithms. Moreover, most of the methods that detect changes in unlabeled data stream merely deal with numerical data sets, and also are facing considerable difficulty when the dimension of data tends to increase. We present a Precise Statistical approach for Concept Change Detection in unlabeled data streams, which, abbreviated as PSCCD, detects changes using an exchangeable test. This hypothesis test is driven from a martingale which is based on Doob’s maximal inequality. The advantages of our approach are three fold. First, it does not require a sliding window on the data stream whose size is a well-known challenging issue; second, it works well in multi-dimensional data stream, and last but not the least, it is applicable to different types of data including categorical, numerical and mixed-attribute data streams. To explore the advantages of our approach, quite a lot of experiments with different settings and specifications are conducted. The obtained results are very promising.

MSC:

62-07 Data analysis (statistics) (MSC2010)
62G10 Nonparametric hypothesis testing
60G09 Exchangeability for stochastic processes
60G42 Martingales with discrete parameter
68M99 Computer system organization

Keywords:
martingales; exchangeability; hypothesis testing; strangeness measure

Full Text: DOI

References:
[9] Ho, S.-S., A martingale framework for concept change detection time-varying data streams, (), 321-327


[16] Bickel, P.J.; Doksum, K., Mathematical statistics: basic ideas and selected topics, (1977), Holden-Day, Inc. · Zbl 0403.62001

[17] ()


[22] Wang, H.; Fan, W.; Yu, P.S.; Han, J., Mining concept-drifting data streams using ensemble classifiers, (), 226-235


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.