Engineering Applications of Artificial Intelligence 65 (2017) 336–345

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Adaptive learning for dynamic environments: A comparative approach

Joana Costa [a,b,*], Catarina Silva [a,b], Mário Antunes [a,c], Bernardete Ribeiro [b]

[a] School of Technology and Management, Polytechnic Institute of Leiria, Portugal
[b] Center for Informatics and Systems (CISUC), Department of Informatics Engineering, University of Coimbra, Portugal
[c] Center for Research in Advanced Computing Systems (CRACS), INESC-TEC University of Porto, Portugal

ARTICLE INFO

Keywords:
Dynamic environments
Ensembles
Learn++.NSE
Twitter

ABSTRACT

Nowadays most learning problems demand adaptive solutions. Current challenges include temporal data streams, drift and non-stationary scenarios, often with text data, whether in social networks or in business systems. Various efforts have been pursued in machine learning settings to learn in such environments, specially because of their non-trivial nature, since changes occur between the distribution data used to define the model and the current environment.

In this work we present the Drift Adaptive Retain Knowledge (DARK) framework to tackle adaptive learning in dynamic environments based on recent and retained knowledge. DARK handles an ensemble of multiple Support Vector Machine (SVM) models that are dynamically weighted and have distinct training window sizes. A comparative study with benchmark solutions in the field, namely the Learn++.NSE algorithm, is also presented. Experimental results revealed that DARK outperforms Learn++.NSE with two different base classifiers, an SVM and a Classification and Regression Tree (CART).

1. Introduction

Streaming sources are becoming ubiquitous. Ranging from data generated by sensors on the Internet of Things (IoT) to social media platforms increasingly accessed with mobile devices, such deluge of data streams is becoming one of the greatest challenges in terms of learning and information extraction (Huijse et al., 2014). Hence, nowadays most learning problems demand dynamic models, adaptive to new circumstances as they emerge. Paradigmatic to this setting are social networks as Twitter, where new information appears all the time. Albeit we can undoubtedly benefit from all these data, one major drawback of such overflow is the inability to easily perceive important, significant and accurate information. This challenge arises not only because the amount of data is overwhelming to process, but also because time plays an important role by fast out-dating information (Costa et al., 2016).

To handle such challenges of dynamic environments we have to address some innovative models that are able to deal with models ageing as, so far, the deployed models performance is reduced because they are not able to deal with dynamic environments.

Additionally, drifts can have different patterns and thus must be treated differently. The most significant types of drifts are depicted in Fig. 1, namely sudden, gradual, incremental and reoccurring (Zliobaite, 2010). Sudden drift is present when the occurring rate of the drift is high and a concept appears or disappears abruptly. Although it is mostly stated as sudden or abrupt drift, it can also be referred as concept change. Gradual drift is characterized by a low drift rate and occurs when the probability of a given context to be associated with a concept increases or decreases during a certain period of time. Additionally, the probability to be associated with another context increases proportionally. Incremental drift can be considered as a subgroup of gradual drift, through the main difference is that the change between the two concepts is much slower and only perceived when looking to what is occurring during longer periods of time. Reoccurring drift occurs when a previously active concept reappears after a period of time. It is important to refer that although it appears seasonally its periodicity must be unknown, otherwise the core assumption of the uncertainty about the future could be compromised.

Different approaches have been pursued with the above goals, like ensemble systems for classification problems (Kuncheva, 2004), proposed and discussed in this work. We present the DARK framework, Drift Adaptive Retain Knowledge framework, that uses an ensemble of Support Vector Machines with dynamic weighting schemes and variable training window sizes for text classification scenarios. A comparative

* Corresponding author.
E-mail addresses: joana.costa@ipleiria.pt, joanamc@dei.uc.pt (J. Costa), catarina@ipleiria.pt, catarina@dei.uc.pt (C. Silva), mario.antunes@ipleiria.pt, mantunes@dcc.fc.up.pt (M. Antunes), bribeiro@dei.uc.pt (B. Ribeiro).

http://dx.doi.org/10.1016/j.engappai.2017.08.004
Received 5 March 2017; Received in revised form 13 June 2017; Accepted 7 August 2017
0952-1976/

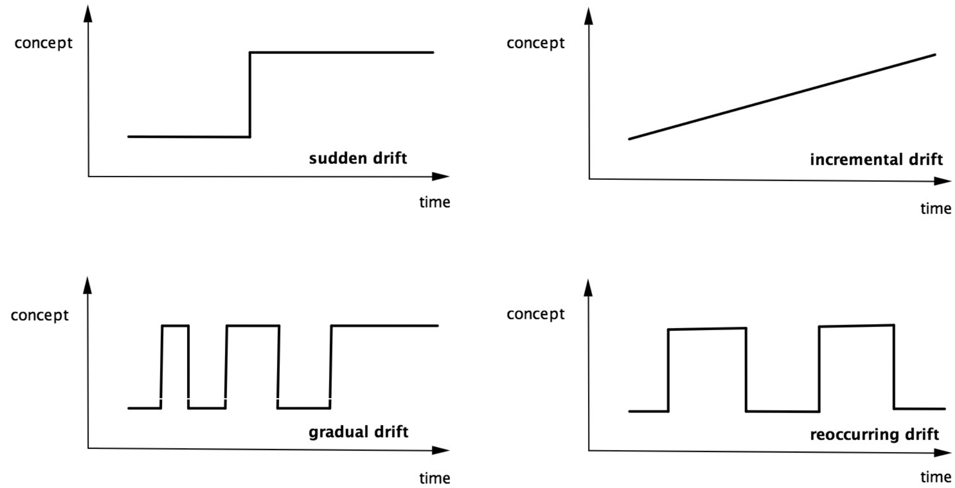

Fig. 1. Different types of drift.

study with benchmark solution in the field is also put forward and the experimental results attest the potential of DARK, as it outperforms both Learn++.NSE with two different base classifiers, an SVM and a Classification and Regression Tree (CART).

There are tree main contributions in this paper: to infer about the influence of recent examples for the overall learning and classification performances; to validate the DARK framework with text classification scenarios, by applying it to text datasets based on Twitter social network public stream and present a comparative study with benchmark solutions in the field, namely Learn++.NSE algorithm.

The rest of the paper is organized as follows. Section 2 presents active and passive approaches for handling drift in dynamic environments. Section 3 defines and details the proposed DARK framework. In Section 4 we introduce the experimental setup, including the Twitter case study and a description of Learn++.NSE. Section 5 presents and discusses the obtained results. Finally, we address conclusions and future lines of research in Section 6.

2. Approaches for drift detection, adaptation and learning

Different approaches exist for learning in nonstationary environments that can be casted as active or passive approaches, that are described and summarized in Table 1.

2.1. Active approaches

Active approaches for learning in nonstationary and dynamic environments are used to detect changes in the environment and react adaptively, updating or building a new classifier. Features are extracted for change detection and, once a change is detected, the classifier model is updated or rebuilt by discarding the obsolete knowledge and adapting to the new environment. The whole process involves change detection and adaptation methods (Ditzler et al., 2015).

Change detection approaches inspect extracted features and variations in the underlying distribution data using theoretically-grounded statistical techniques and include (Ditzler et al., 2015):

1. *Hypothesis Tests* assess the validity of a hypothesis by controlling the false positive rate in change detection based on predetermined confidence calculations and using statistical techniques. The confidence threshold can be based on the mean value with which a set of samples has been drawn from a specific distribution as in Patist (2007) and Nishida and Yamauchi (2007);
2. *Change-Point Methods* use a fixed data sequence to verify if a given sequence contains a change-point, by analysing all possible partitions of the data stream. This statistical technique is highly computational bounded, nevertheless it has the ability to detect the presence of a change and estimate the instant where the change occurred, as in Ross et al. (2011);
3. *Sequential Hypothesis Tests* inspect sequentially incoming examples, one at a time, until there are enough examples to determine the presence of a change or not. Some examples of this technique are probability ratio test (Wald, 1992) and repeated significance test (Armitage, 1960);
4. *Change Detection Tests* overcome limitations of the previous technique by sequentially analysing the statistical behaviour of data streams. This method consists on a change detection based on a threshold as in Harel et al. (2014); Haque et al. (2015). The limitation of this method is the difficulty to set the threshold to an optimal value with which we may have a reasonable classification performance.

The **Adaptation** phase occurs after a change in environment is observed and detected. It consists on adapting the classifier to the change by learning from the new available information and discarding the obsolete (Gama et al., 2014). Adaptation mechanisms can be grouped into the following three main categories (Ditzler et al., 2015):

- *Windowing* is the most used and easiest mechanism. It is based on a sliding window that includes, at each given moment, the most recent and up-to-date examples, while the obsolete ones are discarded. With this mechanism the up-to-date examples are used to retrain the classifier and thus enhance its performance for the next batch(es). The choice of the appropriate window length is a critical issue and can itself be adaptively calculated (Alippi et al., 2013, 2012; Bifet and Gavalda, 2007) or determined by the expected change ratio (Alippi and Roveri, 2008; Cohen et al., 2008b). Just-In-Time (JIT) adaptive classifier, a new generation of adaptive classifiers that are able to operate in nonstationary environments is proposed in Alippi and Roveri (2008).
- *Weighting*, unlike windowing mechanisms, takes into account all the examples weighted according to some rule, like their age or relevancy with respect to the recent classification accuracy performance (Koychev, 2000). Several approaches can be found in the literature regarding the weighting mechanisms used: gradual forgetting (Koychev, 2000); time-based weighting (Datar and Motwani, 2016), change index which measures the variation of data processing over time (Alippi et al., 2009); and based on the accuracy/error calculated in the last batch of supervised data (Klinkenberg, 2004);