Neural Networks 80 (2016) 79–94

Contents lists available at ScienceDirect

Neural Networks

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

Meta-cognitive online sequential extreme learning machine for imbalanced and concept-drifting data classification

Bilal Mirza, Zhiping Lin *

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore

ARTICLE INFO

Article history:
Received 3 July 2015
Received in revised form 8 April 2016
Accepted 21 April 2016
Available online 28 April 2016

Keywords:
Multi-class imbalance
Concept drift
Extreme learning machine
Meta-cognition
Sequential learning

ABSTRACT

In this paper, a meta-cognitive online sequential extreme learning machine (MOS-ELM) is proposed for class imbalance and concept drift learning. In MOS-ELM, meta-cognition is used to self-regulate the learning by selecting suitable learning strategies for class imbalance and concept drift problems. MOS-ELM is the first sequential learning method to alleviate the imbalance problem for both binary class and multi-class data streams with concept drift. In MOS-ELM, a new adaptive window approach is proposed for concept drift learning. A single output update equation is also proposed which unifies various application specific OS-ELM methods. The performance of MOS-ELM is evaluated under different conditions and compared with methods each specific to some of the conditions. On most of the datasets in comparison, MOS-ELM outperforms the competing methods.

1. Introduction

Class imbalance and concept drift are two problem domains in sequential learning which have recently attracted much attention in various fields, see, e.g., Ditzler and Polikar (2012) and Hoens, Polikar, and Chawla (2012). Class imbalance and concept drift can exist either separately or concurrently in a data stream. Although a lot of work has been done focusing on the concept drift problem (Hoens et al., 2012), the class imbalance problem (He & Garcia, 2009; Wasikowski & Chen, 2010; Zhou & Liu, 2006) and in particular, the combination of the two problems in a data stream are largely unexplored.

Sequential (or incremental) learners (Kim, Toh, Teoh, Eng, & Yau, 2013; Lian, Huang, Saratchandran, & Sundararajan, 2006) are used when a complete dataset is not available at once and is received over time as a stream of samples. In the context of sequential learning, data sampling methods are generally used to tackle the class imbalance problem (Chen & He, 2009, 2011; Ditzler, Polikar, & Chawla, 2010; Gao, Fan, Han, & Yu, 2007). For data sampling methods, the imbalance in data streams is alleviated by undersampling the majority class or oversampling the minority class in every chunk. Recently, a weighted online sequential extreme learning machine (WOS-ELM) algorithm was proposed in Mirza, Lin, and Toh (2013) as a cost-sensitive learning method (Ling & Sheng, 2008). WOS-ELM is an extension of online sequential extreme learning machine (OS-ELM) (Lian et al., 2006) to imbalance learning. OS-ELM has now become one of the standard algorithms for sequential learning due to its fast learning speed in both chunk-by-chunk and one-by-one modes (Kim et al., 2013; Mirza et al., 2013; Vaughan & Bohac, 2015).

In non-stationary environments, the statistical characteristics of training data streams may change, resulting in concept drift. Concept drifts generally occur when the underlying distribution changes over time. Various methods have been proposed for learning from drifting concepts (Elwell & Polikar, 2011; Gama, Medas, Castillo, & Rodrigues, 2004; Moreno-Torres, Raeder, Alaiz-RodríGuez, Chawla, & Herrera, 2012). Some OS-ELM specific methods also exist which address the non-stationary problem in online sequential learning. For example, an OS-ELM based method, referred to as timeliness online sequential extreme learning machine (TOSELM) (Gu, Liu, Chen, Jiang, & Yu, 2014), responds to the changing data distribution using adjustable weighting mechanism. For applications where data has a period of validity, i.e., after certain time period the old data becoming irrelevant, an online sequential extreme learning machine with forgetting mechanism (FOS-ELM) was proposed in Zhao, Wang, and Park (2012) to remove the outdated data after a fixed period. However, both TOSELM (Gu et al., 2014) and FOS-ELM (Zhao et al., 2012) methods have not been applied to class imbalance learning.

Recently, the class imbalance problem in concept-drifting environments has received some attention for chunk-by-chunk

* Corresponding author. Tel.: +65 67906857; fax: +65 6793 3318.
E-mail addresses: bilal2@e.ntu.edu.sg (B. Mirza), ezplin@ntu.edu.sg (Z. Lin).

http://dx.doi.org/10.1016/j.neunet.2016.04.008
0893-6080/

learning see, e.g., Ditzler and Polikar (2012) and Hoens et al. (2012). Learn^{++}.NIE (Ditzler & Polikar, 2012) is considered as the state-of-the-art method in such environments. However, most of the existing methods, including Learn^{++}.NIE, assume that a full chunk of data is available for training. Recently, an ensemble of subset online sequential extreme learning machine (ESOS-ELM) (Mirza, Lin, & Nan, 2015) was proposed by the authors to tackle class imbalance and concept drift in both one-by-one and chunk-by-chunk modes. However, both ESOS-ELM and Learn^{++}.NIE are proposed specifically for data streams with two classes and the extension of these two methods to the multi-class case is non-trivial. Very recently, multi-class imbalance problem in sequential learning has been addressed by the authors in Mirza, Lin, Cao, and Lai (2015) for datasets without concept drift. To the best of our knowledge, there is presently no multi-class sequential learner that can tackle the combined problem of class imbalance and concept drift in data streams. Moreover, ESOS-ELM and Learn^{++}.NIE are primarily proposed for data streams with imbalanced classes and drifting concepts and may not adapt efficiently to balanced data or stationary applications. In view of the increasing popularity of sequential learning and its wide applications (Ditzler & Polikar, 2012; Hoens et al., 2012; Lan, Soh, & Huang, 2009), it is desirable to have a self-regulatory classifier which is applicable to balance or imbalance and stationary or concept drift learning from binary or multi-class data streams.

Meta-cognitive learning (Babu & Suresh, 2012, 2013; Cox, 2005; Isaacson & Fujita, 2006; Nelson & Narens, 1980; Savitha, Suresh, & Kim, 2014) has been used to self-regulate the learning process. It decides which sample to learn, when and how to learn. A recently proposed multi-class ELM based method, called Meta-cognitive ELM (McELM) (Savitha et al., 2014), obtains better generalization performance than self-regulatory resource allocation network (SRAN), ELM, SVM and another meta-cognitive neural network (Babu & Suresh, 2012). However, there is no mechanism to tackle the concept drift problem in McELM. Moreover, although imbalance learning was discussed, McELM was not suitable for highly imbalanced data streams since samples from minority and majority classes were treated equally (Savitha et al., 2014).

With the above background, in this paper we propose a sequential learner for tackling the class imbalance problem in binary or multi-class data streams with or without concept drift. The method is referred to as meta-cognitive online sequential extreme learning machine (MOS-ELM). Unlike McELM (Savitha et al., 2014), MOS-ELM can tackle the class imbalance and concept drift problems effectively. MOS-ELM is capable of solving the multi-class concept drift problem for which ESOS-ELM and Learn^{++}.NIE methods fall short. Moreover, WOS-ELM (Mirza, Lin, Cao et al., 2015; Mirza et al., 2013) only tackles the multi-class imbalance problem without concept drift, while MOS-ELM can be effectively applied to multi-class imbalance learning with or without concept drift. The main contributions of this paper are enumerated as follows:

(1) The first meta-cognitive framework to address both class imbalance and concept drift problems is proposed. It incorporates data sampling and cost-sensitive weighting for multi-class imbalance learning and a window-based approach for concept drift learning.
(2) A new adaptive window approach is introduced for concept drift learning.
(3) A new OS-ELM based output update equation is derived for imbalanced or balanced data streams, with or without concept drift. This equation is comprehensive as it unifies various existing application specific OS-ELM methods.

This paper is organized as follows. Section 2 gives the preliminaries. Section 3 presents the details of the MOS-ELM method. This is followed by experiments for validating the performance of the proposed framework in Section 4. Finally, Section 5 concludes the paper.

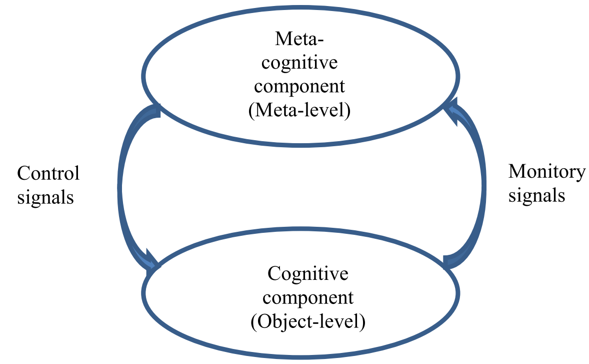

Fig. 1. Nelson and Narens model of meta-cognition.

2. Preliminaries on meta-cognitive learning and related knowledge measures

According to a recent paper in human learning (Isaacson & Fujita, 2006), learning is most effective when the learner is capable of assessing what-to-learn, when-to-learn and how-to-learn. This type of learning is referred to as meta-cognitive learning and is effective because the learner adopts self-regulation in learning. Meta-cognition means knowledge about knowledge. It is the learner's knowledge about his own cognitive processes. In meta-cognitive learning, the learner can control the learning process by selecting strategies, monitor their progress, and adapt them appropriately when necessary. A review on meta-cognition models for computation is provided in Cox (2005).

Most of the machine learning algorithms address only how-to-learn problem because they lack the capability to assess their knowledge with respect to the knowledge in the training data. Consequently, meta-cognitive or self-regulated learning algorithms (Babu & Suresh, 2012, 2013; Savitha et al., 2014) were proposed recently which are inspired by the Nelson and Narens model of the human meta-cognition as shown in Fig. 1 (Isaacson & Fujita, 2006; Nelson & Narens, 1980). These algorithms improve and control the cognitive process by selecting the best learning strategy among various strategies stored in the memory. A meta-cognitive framework typically consists of a cognitive component which constitutes the object-level processes and a meta-cognitive component which constitutes the meta-level processes. Cognitive processes are monitored and controlled by feedforward and feedback loops with meta-cognitive processors. The flow of information from the meta-cognitive component to the cognitive component is called the control signal and the flow in the reverse direction is called the monitory signal.

The meta-cognitive component provides self-regulation for the cognitive component, where the cognitive component is a standard learning algorithm. The meta-cognitive component has a dynamic model of the cognitive component and a self-regulatory mechanism to decide what-to-learn, when-to-learn and how-to-learn based on the knowledge of the cognitive component. The meta-cognitive component improves the generalization capability of the learner by preventing it from learning repetitive knowledge. It monitors the performance of the cognitive component using certain knowledge measures and consequently controls the cognitive component by selecting appropriate learning strategies. Knowledge measures used in this work include the class label predicted by the cognitive component, the maximum hinge-loss error and a class imbalance learning (CIL) performance measure, e.g. the geometric mean in this work. The knowledge measures for