Neurocomputing 171 (2016) 89–105

Contents lists available at ScienceDirect

Neurocomputing

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

An incremental meta-cognitive-based scaffolding fuzzy neural network

Mahardhika Pratama [a,*], Jie Lu [a], Sreenatha Anavatti [b], Edwin Lughofer [c], Chee-Peng Lim [d]

[a] Centre for Quantum Computation and Intelligent System, University of Technology, Sydney, Australia
[b] School of Engineering and Information Technology, University of New South Wales, Canberra, Australia
[c] Department of Knowledge-Based Mathematical Systems, Johannes Kepler University, Linz A-4040, Austria
[d] Centre for Intelligent System Research, Deakin University, Geelong Waurn Ponds Campus, Victoria 3216, Australia

ARTICLE INFO

Article history:
Received 12 November 2014
Received in revised form
20 April 2015
Accepted 14 June 2015
Communicated by M.-J. Er
Available online 9 July 2015

Keywords:
Evolving fuzzy systems
Fuzzy neural networks
Meta-cognitive learning
Sequential learning

ABSTRACT

The idea of meta-cognitive learning has enriched the landscape of evolving systems, because it emulates three fundamental aspects of human learning: *what-to-learn*; *how-to-learn*; and *when-to-learn*. However, existing meta-cognitive algorithms still exclude Scaffolding theory, which can realize a plug-and-play classifier. Consequently, these algorithms require laborious pre- and/or post-training processes to be carried out in addition to the main training process. This paper introduces a novel meta-cognitive algorithm termed GENERIC-Classifier (gClass), where the how-to-learn part constitutes a synergy of Scaffolding Theory – a tutoring theory that fosters the ability to sort out complex learning tasks, and Schema Theory – a learning theory of knowledge acquisition by humans. The what-to-learn aspect adopts an online active learning concept by virtue of an extended conflict and ignorance method, making gClass an incremental semi-supervised classifier, whereas the when-to-learn component makes use of the standard sample reserved strategy. A generalized version of the Takagi-Sugeno Kang (TSK) fuzzy system is devised to serve as the cognitive constituent. That is, the rule premise is underpinned by multivariate Gaussian functions, while the rule consequent employs a subset of the non-linear Chebyshev polynomial. Thorough empirical studies, confirmed by their corresponding statistical tests, have numerically validated the efficacy of gClass, which delivers better classification rates than state-of-the-art classifiers while having less complexity.

1. Introduction

The consolidation of the meta-cognitive aspect in machine learning was initiated by Suresh et al. [7–11] based on a prominent meta-memory model proposed by Nelson and Naren [6]. The works in [7–11] identify that the meta-cognitive component, namely *what-to-learn*, *how-to-learn* and *when-to-learn*, can respectively be modelled with sample deletion strategy, sample learning strategy and sample reserved strategy. Nevertheless, their pioneering works still discount the construct of Scaffolding theory [12,22], rendering a plug-and-play classifier. They have also not addressed the issue of semi-supervised learning, since the what-to-learn phase requires the data to be fully labelled.

A novel meta-cognitive-based Scaffolding classifier, the GENERIC-classifier (gClass), is proposed in this paper. The gClass learning engine comprises three elements: what-to-learn; how-to-learn; and when-to-learn. The underlying novelty of gClass lies on the use of Schema and Scaffolding theories in the how-to-learn component to realize it as a plug-and-play classifier. The plug-and-play learning paradigm emphasizes the need for all learning modules to be embedded in a single learning process without invoking any pre- and/or post-training processes. In respect of its cognitive constituent, the gClass fuzzy rule triggers a non-axis-orthogonal fuzzy rule in the input space, underpinned by the multivariate Gaussian function rule premise. Unlike the standard form of TSK fuzzy rule consequents, the rule consequent of gClass is built upon a non-linear function stemming from a subset of non-linear Chebyshev polynomials. All training mechanisms run in the strictly sequential learning mode to assure fast model updates and comply with the four principles of online learning [32]: (1) all training observations are sequentially presented one by one or chunk by chunk to gClass; (2) only one training datum is seen and learned in every training episode; (3) a training sample which has been seen is discarded without being reused; and (4) gClass does not require any information pertaining to the total number of training data.

The gClass learning scenario utilizes several learning modules of our previous algorithms in [18,19]: three rule growing cursors, namely Datum Significance (DS), Data Quality (DQ), and Generalized Adaptive Recursive Theory+ (GART+), are used to evolve fuzzy rules according to the Schema theory [14]; two rule pruning

* Corresponding author.
E-mail addresses: pratama@ieee.org (M. Pratama), jie.lu@uts.edu.au (J. Lu), s.anavatti@adfa.edu.au (S. Anavatti), Edwin.lughofer@jku.at (E. Lughofer), chee.lim@deakin.edu.au (C.-P. Lim).

http://dx.doi.org/10.1016/j.neucom.2015.06.022
0925-2312/

strategies, namely Extended Rule Significance (ERS) and Potential (P+) methods, are assembled to get rid of obsolete and inactive fuzzy rules and portray the fading aspect of Scaffolding theory. The P+ method also deciphers the rule recall process, manifesting the problematizing component of Scaffolding theory to cope with the recurring concept drift; the Fuzzily Weighted Generalized Recursive Least Square (FWGRLS) method is integrated to adjust the rule consequent of the fuzzy rule and in turn delineates the passive supervision of the Scaffolding theory. gClass operates as its counterparts in [7–11], where the sample reserved strategy is employed in the when-to-learn process. Nonetheless, several new learning modules are proposed in this paper:

- The what-to-learn component is built upon a new online active learning scenario, called the Extended Conflict and Ignorance (ECI) method. The ECI method is derived from the conflict and ignorance method [2], and the ignorance method is enhanced by the use of the DQ method instead of the classical rule firing strength concept. This modification makes the online active learning method more robust against outliers and more accurate in deciding the sample ignorance. Note that this mechanism can be also perceived as an enhanced version of the original what-to-learn module in [7–11]. In [7–11], the what-to-learn module is limited to ruling out redundant samples for model updates, and still assumes that data are fully labelled.
- A new fuzzy rule initialization strategy is proposed and is constructed by the potential per-class method. This method is used to avoid misclassifications caused by the class overlapping situation. A number of research efforts have been attempted in [7–10,69-71] to circumvent the class overlapping situation, however they rely on the distance ratio method, which overlooks the existence of unclean clusters. An unclean cluster is a cluster that contains supports from different classes and is prevalent in real world-problems. This learning aspect actualizes the restructuring phase of Schema theory.
- gClass is also equipped with a local forgetting scheme inspired by [28] to surmount gradual concept drift, where the forgetting intensity is enumerated by a newly developed method, called the Local Data Quality (LDQ) method. It is worth stressing that gradual concept drift is more precarious than abrupt concept drift, because gradual concept drift cannot be detected by standard drift detection or the rule generation method. On the other side, it cannot be handled by the conventional parameter learning method either. This situation entails the local forgetting scheme, which adapts fuzzy rule parameters more firmly and is thereby able to pursue changing data distributions. In the realm of Scaffolding theory, the local drift-handling strategy plays a problematizing role in the active supervision of the theory.
- gClass enhances the Fisher Separability Criterion (FSC) in the empirical feature space method with the optimization step via the gradient ascent method. This step not only alleviates the curse of dimensionality, but it also improves the discriminatory power of input features. Noticeably, it triggers a direct impact on the classifier's generalization. The online feature weighting technique is employed to address the complexity reduction scenario in the active supervision of the scaffolding concept.

The contributions of this paper are summarized as follows: (1) the paper proposes a new class of meta-cognitive classifiers, which consolidate the Schema and Scaffolding theories to drive the how-to-learn module. (2) The paper introduces a novel type of TSK fuzzy rule, crafted by the multivariable Gaussian function in the premise component and the non-linear Chebyshev polynomial in the output component. (3) Four novel learning modules in the gClass learning engine are proposed: online feature selection; online active learning; class overlapping strategy; and online feature weighting mechanism. The viability and efficacy of gClass have been numerically validated by means of thorough numerical studies in both real-world and artificial study cases. gClass has also been benchmarked against various state-of-the-art classifiers, confirmed by rigorous statistical tests in which gClass demonstrates highly encouraging generalization power while suppressing complexity to an acceptable level. The remainder of this paper is organized as follows: Section 2 discusses related works. Section 3 illustrates the gClass inference mechanism, i.e., its cognitive aspect. Section 4 outlines the algorithmic development of gClass, i.e., its meta-cognitive component. Section 5 deliberates the empirical studies and discussions of the research gap and contribution, which detail the viability and research gap of gClass. Concluding remarks are drawn in the last section of this paper.

2. Literature review

In this section, two related areas are discussed. A survey of the psychological concepts implemented in gClass is undertaken, as well as a literature review of state-of-the art evolving classifiers.

2.1. Human learning

The main challenge of learning sequentially from data streams is how to deal with the stability and plasticity dilemma [15,16,49], which requires a balance between new and old knowledge. In the realm of cognitive psychology, this dilemma is deliberated in Schema theory, which is a psychological model for human knowledge acquisition and the organization of human memory [14,66], in which knowledge is organized into units, or *schemata* (sing. *schema*). Information is stored within the schemata, and Schema theory is thus the foundation of a conceptual system for understanding knowledge representation.

In essence, Schema theory is composed of two parts: schemata construction and schemata activation. Schemata are built in the construction phase, and this is achieved by three possible learning scenarios that relate to the conflict level induced by an incoming datum – accretion, tuning and restructuring. Accretion pinpoints a conflict-free situation, where an incoming datum can be well-represented by an existing schema. Tuning represents a minor conflict circumstance in which only the adaptation of a schema is entailed. The most significant case is the restructuring phase, in which a datum induces a major conflict which demands the restructure of an existing schema or its complete replacement. Schemata activation describes a self-regulatory process to evaluate the performance of the schemata, or determines a compatible learning scenario to manage a new example.

Scaffolding theory elaborates a tutoring theory, which assists students to accomplish a complex learning task [68]. This goal is achieved by passively and actively supervising the training process. Passive supervision implements a learning strategy by virtue of the experience and consequence mechanism, and depends on the predictive quality of fresh data. Passive supervision is particularly represented by the parameter learning of the rule consequent. Active supervision makes use of more proactive mechanisms and consists of three learning scenarios: complexity reduction; problematizing; and fading [67]. The complexity reduction component aims to relieve the learning burden and can be actualized by data pre-processing and/or feature selection. Problematizing copes with concept drift and can be realized by a local forgetting mechanism and/or rule recall strategy. The fading constituent deciphers a structure simplification procedure which inhibits redundancy in the rule base; this concept is usually executed by the rule pruning technique.