Neurocomputing 74 (2011) 1785–1799

Contents lists available at ScienceDirect

Neurocomputing

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

Incremental learning with multi-level adaptation

Abdelhamid Bouchachia

Department of Informatics, University of Klagenfurt, Austria

ARTICLE INFO

Available online 24 February 2011

Keywords:
Adaptation
Incremental learning
Classification
Ensemble learning
Dynamic environments
Drift

ABSTRACT

Self-adaptation is an inherent part of any natural and intelligent system. Specifically, it is about the ability of a system to reconcile its requirements or goal of existence with the environment it is interacting with, by adopting an optimal behavior. Self-adaptation becomes crucial when the environment changes dynamically over time. In this paper, we investigate self-adaptation of classification systems at three levels: (1) natural adaptation of the base learners to change in the environment, (2) contributive adaptation when combining the base learners in an ensemble, and (3) structural adaptation of the combination as a form of dynamic ensemble. The present study focuses on neural network classification systems to handle a special facet of self-adaptation, that is, incremental learning (IL). With IL, the system self-adjusts to accommodate new and possibly non-stationary data samples arriving over time. The paper discusses various IL algorithms and shows how the three adaptation levels are inherent in the system's architecture proposed and how this architecture is efficient in dealing with dynamic change in the presence of various types of data drift when applying these IL algorithms.

1. Introduction

Adaptivity is particularly manifest in intelligent applications where learning from data is at the heart of system modeling and identification. The goal is to cope with non-stationary changing situations by employing adaptive mechanisms to accommodate changes in the data. This becomes more important when storage capacities (memory) are very limited and when data arrives over long periods of time. In such situations, the system should adapt itself to the new data samples which may convey a changing situation and at the same time should keep in memory relevant information that had been learned in the remote past.

In this contribution, we aim at studying one of the fundamental aspects of adaptivity, that is, adaptive incremental learning (AIL) which seeks to deal with data arriving over time or with (static) huge amounts of data that exceed the storage capacities. Thus, processing of data at once is not feasible. Most of the available literature on machine learning reports on learning models that are one-shot experiment lacking adaptivity. Therefore, learning algorithms with an adaptive incremental learning ability are of increasing importance in many nowadays online data streams and time series applications. The neural learning systems discussed here are classification oriented architectures that suggest an adaptive incrementality based on algorithmic tuning. A detailed definition of incrementality can be found in [5].

All incremental learning algorithms are confronted with the plasticity-stability dilemma. This dilemma establishes the trade-off between catastrophic interference (or forgetting) on one hand and the ability to incrementally and continually accommodate new knowledge in the future whenever new data samples become available. The former aspect is referred to as stability, while the latter is referred to as plasticity. In a nutshell, the stability-plasticity dilemma is concerned with learning new knowledge without forgetting the previously learned one. This problem has been thoroughly studied by many researchers [16,20,42].

From another perspective, incrementality assumes phenomena that evolve over time and change their known evolution schemes. This refers in the first place to the problem of concept drift. To deal with such a problem, most often dedicated techniques to drift detection and handling use either full memory (e.g., the system has a memory and therefore has access to already seen data in the past) or partial memory (e.g., temporal window of data or space for storing (part of) the system's knowledge). It seems, however, quite appealing to investigate the problem of concept drift with no memory (i.e., data is processed online without any full or temporal storage). This paper aims at looking closely at this approach.

Moreover, as we are interested in studying a collection of incremental learning algorithms, it sounds legitimate to observe adaptivity from the perspective of ensemble learning. By considering such line of investigation, the aim of this paper is to achieve a *three-level* adaptivity mechanism as shown in Fig. 1.

1. Adaptivity due to the nature of the classifiers. The classifiers are self-adaptive by construction.
2. Adaptivity due to proportional (weighted) contribution of each classifier in the ensemble decision.
3. Adaptivity due to the structural update (dynamically changing structure) of the ensemble.

E-mail address: hamid@isys.uni-klu.ac.at

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.12.017

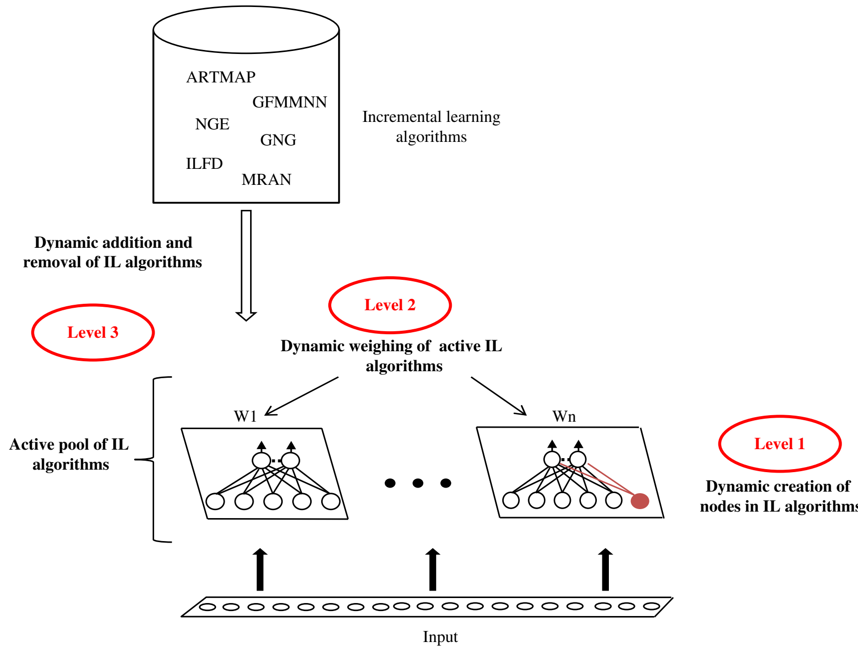

Fig. 1. Multi-level adaptation.

While levels 1 and 2 per see are not original in this contribution, level 3 is an original proposal. However, combining levels 1 and 2 at the same time is indeed an original contribution that is worth investigating.

In addition to the novelties mentioned, it is also important to stress that the present approach is distinct to other studies discussing incrementality and ensemble learning [28,35,44]. The present study differs in the following aspects:

- base classifiers are distinct,
- base classifiers are adaptive and incremental by nature,
- base classifiers handle novelty detection by nature,
- ensemble method with dynamic combination, new classifiers may be added others may be removed, and
- ensemble method proposed deals with data drift.

It is important to underline that our first and primary aim in this paper is to consider:

- incremental neural networks (there exist incremental statistical/probabilistic [7] and incremental fuzzy classifiers [8], but these are not relevant to this study),
- well-established algorithms—that are widely applied, and
- a high number of such algorithms.

The question for us is: *How to build machinery that is able to self-adapt at various levels*. To answer this question, we define in this paper a layered architecture of adaptation mechanisms. In such an architecture the base learner can be changed at wish, but the adaptation strata remain. One may think about it as a service oriented architecture, where a service is delivered by a learner. These services can either compete against each other or cooperate with each other.

Before delving into details of each adaptivity level, we highlight the structure of the paper. Section 2 describes the incremental classifiers used, their differences and similarities. Section 4 looks at the problem of ensemble classifiers before discussing the problem of concept drift and adaptivity consequences in Section 5. Section 6 describes an approach that unifies ensemble learning and drift handling from the perspective of adaptivity. Section 7 provides an evaluation of the various adaptivity levels mentioned earlier.

2. Roadmap through AIL algorithms

There exists a certain number of incremental learning algorithms that are known to be lifelong learning algorithms. For the sake of exhaustiveness, we select six most illustrative algorithms [5,9]. These include: minimal resource allocating network (MRAN) [49,54], fuzzy ARTMAP (FAM) [12,46,20], nearest generalized exemplar (NGE) [40], generalized fuzzy min-max neural networks (GFMMNN) [6,19], growing neural gas (GNG) [14,18,45], and incremental learning based on function decomposition (ILFD) [3]. These incremental algorithms are chosen due to their characteristics including different types of prototypes, generation mechanisms, operations on the prototypes (shrinking, deletion, growing, overlap control), noise resistance, and data normalization requirement. It is, however, important to recall that some of the algorithms require recycling over the data to achieve more stability. This will be avoided in this study, so that the spirit of incremental learning as defined earlier is preserved.

Table 1 shows some of the characteristics of the studied algorithms. Each of these algorithms is capable of online learning and produces a set of prototypes per class (in the case of classification). The algorithmic steps in all these algorithms are theoretically the same. A prototype is generated when the incoming data point is sufficiently dissimilar to the existing prototypes. Otherwise, an adjustment of some of the existing prototypes is conducted. The first characteristic that distinguishes these algorithms is the type of prototypes. In fact, we propose to categorize them into two classes: *hyperbox-based algorithms* (*HBAs*) and *point-based algorithms* (*PAs*). The HBAs class includes FAM, NGE and GFMMNN. Many variations of these algorithms exist. For instance, there exist some attempts to generalize the FAM categories to different shapes. The PAs class includes GNG and ILFD. While prototypes in GNG are nodes of a graph, in ILFD and MRAN they are cluster centers (in the sense of radial basis