

# Incremental Learning

**Abdelhamid Bouchachia**

*University of Klagenfurt, Austria*

## INTRODUCTION

Data mining and knowledge discovery is about creating a comprehensible model of the data. Such a model may take different forms going from simple association rules to complex reasoning system. One of the fundamental aspects this model has to fulfill is adaptivity. This aspect aims at making the process of knowledge extraction continually maintainable and subject to future update as new data become available. We refer to this process as *knowledge learning*.

Knowledge learning systems are traditionally built from data samples in an off-line one-shot experiment. Once the learning phase is exhausted, the learning system is no longer capable of learning further knowledge from new data nor is it able to update itself in the future. In this chapter, we consider the problem of incremental learning (IL). We show how, in contrast to off-line or batch learning, IL learns knowledge, be it symbolic (e.g., rules) or sub-symbolic (e.g., numerical values) from data that evolves over time. The basic idea motivating IL is that as new data points arrive, new knowledge elements may be created and existing ones may be modified allowing the knowledge base (respectively, the system) to evolve over time. Thus, the acquired knowledge becomes self-corrective in light of new evidence. This update is of paramount importance to ensure the adaptivity of the system. However, it should be meaningful (by capturing only interesting events brought by the arriving data) and sensitive (by safely ignoring unimportant events). Perceptually, IL is a fundamental problem of cognitive development. Indeed, the perceiver usually learns how to make sense of its sensory inputs in an incremental manner via a filtering procedure.

In this chapter, we will outline the background of IL from different perspectives: machine learning and data mining before highlighting our IL research, the challenges, and the future trends of IL.

## BACKGROUND

IL is a key issue in applications where data arrives over long periods of time and/or where storage capacities are very limited. Most of the knowledge learning literature reports on learning models that are one-shot experience. Once the learning stage is exhausted, the induced knowledge is no more updated. Thus, the performance of the system depends heavily on the data used during the learning (knowledge extraction) phase. Shifts of trends in the arriving data cannot be accounted for.

Algorithms with an IL ability are of increasing importance in many innovative applications, e.g., video streams, stock market indexes, intelligent agents, user profile learning, etc. Hence, there is a need to devise learning mechanisms that are able of accommodating new data in an incremental way, while keeping the system under use. Such a problem has been studied in the framework of adaptive resonance theory (Carpenter et al., 1991). This theory has been proposed to efficiently deal with the stability-plasticity dilemma. Formally, a learning algorithm is totally stable if it keeps the acquired knowledge in memory without any catastrophic forgetting. However, it is not required to accommodate new knowledge. On the contrary, a learning algorithm is completely plastic if it is able to continually learn new knowledge without any requirement on preserving the knowledge previously learned. The dilemma aims at accommodating new data (plasticity) without forgetting (stability) by generating knowledge elements over time whenever the new data conveys new knowledge elements worth considering.

Basically there are two schemes to accommodate new data. To retrain the algorithm from scratch using both old and new data is known as revolutionary strategy. In contrast, an evolutionary continues to train the algorithm using only the new data (Michalski, 1985). The first scheme fulfills only the stability requirement, whereas the second is a typical IL scheme that is able to fulfill both stability and plasticity. The goal is to make

a tradeoff between the stability and plasticity ends of the learning spectrum as shown in Figure 1.

As noted in (Polikar et al., 2000), there are many approaches referring to some aspects of IL. They exist under different names like *on-line learning*, *constructive learning*, *lifelong learning*, and *evolutionary learning*. Therefore, a definition of IL turns out to be vital:

- IL should be able to accommodate plasticity by learning knowledge from new data. This data can refer either to the already known structure or to a new structure of the system.
- IL can use only new data and should not have access at any time to the previously used data to update the existing system.
- IL should be able to observe the stability of the system by avoiding forgetting.

It is worth noting that the IL research flows in three directions: clustering, classification, and rule associations mining. In the context of classification and clustering, many IL approaches have been introduced. A typical incremental approach is discussed in (Parikh & Polikar, 2007) which consists of combining an ensemble of multilayer perceptron networks (MLP) to accommodate new data. Note that stand-alone MLPs, like many other neural networks, need retraining in order to learn from the new data. Other IL algorithms were also proposed as in (Domeniconi & Gunopulos, 2001), where the aim is to construct incremental support vector machine classifiers. Actually, there exist four neural models that are inherently incremental: (i) adaptive resonance theory (ART) (Carpenter et al., 1991), (ii) min-max neural networks (Simpson, 1992), (iii) nearest generalized exemplar (Salzberg, 1991), and (iv) neural gas model (Fritzke, 1995). The first three

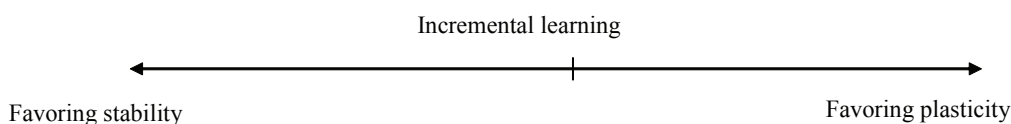
incremental models aim at learning hyper-rectangle categories, while the last one aims at building point-prototyped categories.

It is important to mention that there exist many classification approaches that are referred to as IL approaches and which rely on neural networks. These range from retraining misclassified samples to various weighing schemes (Freeman & Saad, 1997; Grippo, 2000). All of them are about sequential learning where input samples are sequentially, but iteratively, presented to the algorithm. However, sequential learning works only in close-ended environments where classes to be learned have to be reflected by the readily available training data and more important prior knowledge can also be forgotten if the classes are unbalanced.

In contrast to sub-symbolic learning, few authors have studied incremental symbolic learning, where the problem is incrementally learning simple classification rules (Maloof & Michalski, 2004; Reinke & Michalski, 1988).

In addition, the concept of incrementality has been discussed in the context of association rules mining (ARM). The goal of ARM is to generate all association rules in the form of  $X \rightarrow Y$  that have support and confidence greater than a user-specified minimum support and minimum confidence respectively. The motivation underlying incremental ARM stems from the fact that databases grow over time. The association rules mined need to be updated as new items are inserted in the database. Incremental ARM aims at using only the incremental part to infer new rules. However, this is usually done by processing the incremental part separately and scanning the older database if necessary. Some of the algorithms proposed are FUP (Cheung et al., 1996), temporal windowing (Rainsford et al., 1997), and DELI (Lee & Cheung, 1997).

Figure 1. Learning spectrum



5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/incremental-learning/10944](http://www.igi-global.com/chapter/incremental-learning/10944)

## Related Content

---

### Reasoning about Frequent Patterns with Negation

Marzena Kryszkiewicz (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1667-1674).

[www.irma-international.org/chapter/reasoning-frequent-patterns-negation/11042](http://www.irma-international.org/chapter/reasoning-frequent-patterns-negation/11042)

### Neural Networks and Graph Transformations

Ingrid Fischer (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1403-1408).

[www.irma-international.org/chapter/neural-networks-graph-transformations/11005](http://www.irma-international.org/chapter/neural-networks-graph-transformations/11005)

### Program Comprehension through Data Mining

Ioannis N. Kouris (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1603-1609).

[www.irma-international.org/chapter/program-comprehension-through-data-mining/11033](http://www.irma-international.org/chapter/program-comprehension-through-data-mining/11033)

### A Novel Approach on Negative Association Rules

Ioannis N. Kouris (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1425-1430).

[www.irma-international.org/chapter/novel-approach-negative-association-rules/11008](http://www.irma-international.org/chapter/novel-approach-negative-association-rules/11008)

### Transferable Belief Model

Philippe Smets (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1985-1989).

[www.irma-international.org/chapter/transferable-belief-model/11091](http://www.irma-international.org/chapter/transferable-belief-model/11091)