Information and Computation 247 (2016) 152–168

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

On the role of update constraints and text-types in iterative learning

Sanjay Jain [a,*,1], Timo Kötzing [b,*], Junqi Ma [a,*], Frank Stephan [c,*,2]

[a] Department of Computer Science, National University of Singapore, Singapore 117417, Republic of Singapore
[b] Friedrich-Schiller University, Jena, Germany
[c] Department of Mathematics, National University of Singapore, Singapore 119076, Republic of Singapore

ARTICLE INFO

Article history:
Received 19 July 2015
Received in revised form 16 November 2015
Available online 4 January 2016

Keywords:
Inductive inference
Iterative learning
Learning constraints
Decisive learning
Conservative learning

ABSTRACT

The present work investigates the relationship of iterative learning with other learning criteria such as decisiveness, caution, reliability, non-U-shapedness, monotonicity, strong monotonicity and conservativeness. Building on the result of Case and Moelius that iterative learners can be made non-U-shaped, we show that they also can be made cautious and decisive. Furthermore, we obtain various special results with respect to one-one texts, fat texts and one-one hypothesis spaces.

1. Introduction

In this paper we consider *inductive inference*, a branch of algorithmic learning theory. This branch analyses the problem of algorithmically learning a description for a formal language (a recursively enumerable subset of the set of natural numbers) when presented successively all and only the elements of that language. For example, a learner M might be presented more and more even numbers. After each new number, M outputs a description for a language as its conjecture. The learner M might decide to output a program for the set of all multiples of 4, as long as all numbers presented are divisible by 4. Later, when M sees an even number not divisible by 4, it might change this guess to a program for the set of all multiples of 2.

Gold, in his seminal paper [12], introduced this idea of learning a language in the limit formally. His first and simple learning criterion was **TxtGEx**-*learning*,[3] where a learner is *successful* iff, on every *text* for L (listing of all and only the elements of L) it eventually stops changing its conjectures, and its final conjecture is a correct description (an *explanation*) for the input sequence. Trivially, each single, describable language L has a suitable constant function as a **TxtGEx**-learner (this learner constantly outputs a description for L). Thus, we are interested in analysing for which *classes of languages* $\mathcal{L}$ there is a *single learner* h learning *each* member of $\mathcal{L}$. This framework is also sometimes known as *language learning in the limit* and has been studied extensively, using a wide range of learning criteria similar to **TxtGEx**-learning (see, for example, the textbook [17]).

* Corresponding authors.
E-mail addresses: sanjay@comp.nus.edu.sg (S. Jain), timo.koetzing@uni-jena.de (T. Kötzing), ma.junqi@nus.edu.sg (J. Ma), fstephan@comp.nus.edu.sg (F. Stephan).

[1] Supported by NUS grants C252-000-087-001 and R146-000-181-112.
[2] Supported in part by NUS grant R146-000-181-112.

[3] **Txt** stands for learning from a *text* of positive examples; **G** stands for Gold, who introduced this model, and is used to indicate full-information learning; **Ex** stands for *explanatory*.

http://dx.doi.org/10.1016/j.ic.2015.12.005
0890-5401/

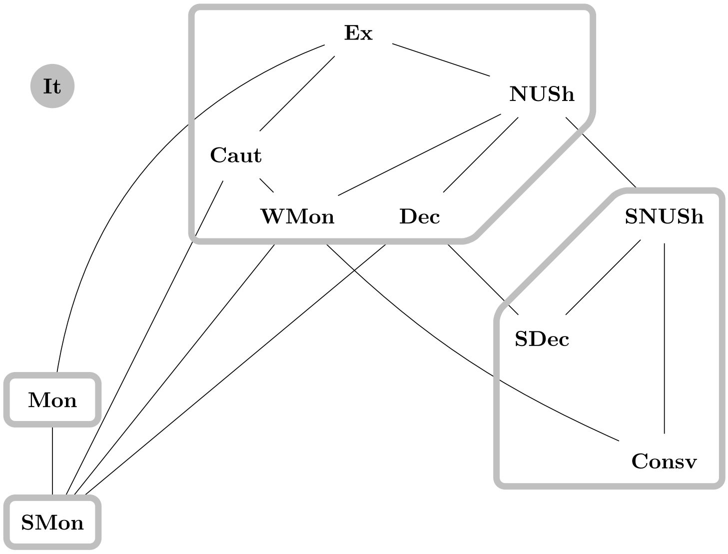

Fig. 1. Relation of criteria combined with iterative learning.

It is easy to see from the definition of **TxtGEx**-learning that the learner can be arbitrarily inefficient: the learner can postpone computations and decisions until more data has been shown; no restrictions on the computing time (beyond linear) in each update step will restrict the learner's abilities. One way to address this problem is to restrict the access to past data. The most common formalisation of this idea is *iterative learning* (**TxtItEx**) [30], where the learner, in each iteration, gets to see only the new data item and its previous hypothesis. Due to the padding lemma, this memory of the previous hypothesis is still not void, but finitely many data can be memorised by padding the hypothesis. In effect, syntactic changes of the hypothesis, which do not affect its semantics, are used as a memory.

There are several approaches which aim to make updates more meaningful. One direction is to consider one-one hypothesis spaces where the learner cannot do padding without changing the semantics of the previous hypothesis. Other restrictions on the updates are requiring that they respect some semantic constraints towards preserving already achieved quality of the previous hypothesis and avoiding obvious errors. For example,

- updates have to be motivated by inconsistent data observed (syntactic conservativeness) [1,27],
- semantic updates have to be motivated by inconsistent data observed (semantic conservativeness) [11],
- updates cannot repeat semantically abandoned conjectures (decisiveness) [26],
- updates cannot go from correct to incorrect hypotheses (non-U-shapedness) [4],
- later conjectures cannot be proper subsets of earlier conjectures (cautiousness) [27] or
- conjectures have to contain all the data observed so far (consistency) [2].

In particular those constraints in the list which rule out updates without a semantic improvement of the hypothesis do in some cases effectively hinder padding and are therefore restrictive compared to plain iterative learning.

There is already a quite comprehensive body of work on how iterativeness relates with various combinations of these constraints [7,13,16–18,20,21,23,24]. However, the work in this area had two shortcomings: (a) it was not clear how strong an update restriction is necessary to actually restrict the learning power below full iterative learning; and (b) there was no complete picture of the relations of the mentioned update restrictions in the setting of iterative learning. With this paper we eliminate these shortcomings: Regarding (a), we show that strong decisiveness, but not decisiveness restricts iterative learning in learning power. Regarding (b) we completely characterise the relationship of the iterative learning criteria with the different restrictions. This is depicted in the diagram in Fig. 1, representing a *map* of the update constraints and their relations. A black line between two learning criteria indicates a trivial inclusion (where the inclusion follows directly from the definition of the restriction). A grey box around criteria indicates equality of these criteria, as found in this work.

Our work extends a breakthrough result by Case and Moelius [9] who showed that iterative learners can be made non-U-shaped. The present work improves this result by showing that iterative learners can also be made decisive — this stands in contrast to the case of the usual non-iterative framework where decisiveness is a real restriction in learning [4]. This result is given in Theorem 10 in Section 4. Also in that section are the other results giving the complete characterisation indicated in Fig. 1.

Further sections give additional results, complementing the statements shown in Section 4. Section 5 considers alternative text-types, such as *fat texts* and *one-one texts*. Here a text is *fat* if every datum appears infinitely often and is *one-one* if every datum appears exactly once. It is interesting to see that, for iterative learning from fat texts, the divide between decisive and strongly decisive learning vanishes and instead neither update constraint restricts the learning power of iterative