Combining Models of Approximation with Partial Learning

Ziyuan Gao[1(✉)], Frank Stephan[2,3], and Sandra Zilles[1]

[1] Department of Computer Science, University of Regina, Regina, SK S4S 0A2, Canada
{gao257,zilles}@cs.uregina.ca

[2] Department of Mathematics, National University of Singapore, Singapore 119076, Singapore

[3] Department of Computer Science, National University of Singapore, Singapore 119076, Singapore
fstephan@comp.nus.edu.sg

Abstract. In Gold's framework of inductive inference, the model of partial learning requires the learner to output exactly one correct index for the target object and only the target object infinitely often. Since infinitely many of the learner's hypotheses may be incorrect, it is not obvious whether a partial learner can be modified to "approximate" the target object.

Fulk and Jain (Approximate inference and scientific method. *Information and Computation* 114(2):179–191, 1994) introduced a model of approximate learning of recursive functions. The present work extends their research and solves an open problem of Fulk and Jain by showing that there is a learner which approximates and partially identifies every recursive function by outputting a sequence of hypotheses which, in addition, are also almost all finite variants of the target function.

The subsequent study is dedicated to the question how these findings generalise to the learning of r.e. languages from positive data. Here three variants of approximate learning will be introduced and investigated with respect to the question whether they can be combined with partial learning. Following the line of Fulk and Jain's research, further investigations provide conditions under which partial language learners can eventually output only finite variants of the target language.

1 Introduction

Gold [8] considered a learning scenario where the learner is fed with piecewise increasing amounts of finite data about a given target language L; at every stage where a new input datum is given, the learner makes a conjecture about L. If there is exactly one correct representation of L that the learner always outputs

F. Stephan was partially supported by NUS grants R146-000-181-112 and R146-000-184-112; S. Zilles was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

K. Chaudhuri et al. (Eds.): ALT 2015, LNAI 9355, pp. 56–70, 2015.
DOI: 10.1007/978-3-319-24486-0_4

after some finite time (assuming that it never stops receiving data about L), then the learner is said to have "identified L in the limit." In this paper, it is assumed that all target languages are encoded as recursively enumerable (r.e.) sets of natural numbers, and that the learner uses Gödel numbers as its hypotheses.

Gold's learning paradigm has been used as a basis for a variety of theoretical models in subjects such as human language acquisition [12] and the theory of scientific inquiry in the philosophy of science [4,11]. This paper is mainly concerned with the *partial learning* model [13], which retains several features of Gold's original framework – the modelling of learners as recursive functions, the use of texts as the mode of data presentation and the restriction of target classes to the family of all r.e. sets – while liberalising the learning criterion by only requiring the learner to output exactly one hypothesis of the target set infinitely often while it must output any other hypothesis only finitely often. It is known that partial learning is so powerful that the class of all r.e. languages can be partially learnt [13].

However, the model of partial learning puts no further constraints on those hypotheses that are output only finitely often. In particular, it offers no notion of "eventually being correct" or even "approximating" the target object. From a philosophical point of view, if partial learning is to be taken seriously as a model of language acquisition, then it is quite plausible that learners are capable of gradually improving the quality of their hypotheses over time. For instance, if the learner M sees a sentence S in the text at some point, then it is conceivable that after some finite time, M will only conjecture grammars that generate S. This leads one to consider a notion of the learner "approximating" the target language.

The central question in this paper is whether any partial learner can be redefined in a way that it approximates the target object and still partially learns it. The first results, in the context of partial learning, deal with Fulk and Jain's [5] notion of approximating recursive functions. Fulk and Jain proved the existence of a learner that "approximates" every recursive function. This result is generalised as follows: partial learners can always be made to approximate recursive functions according to their model and, in addition, eventually output only finite variants of the target function, that is, they can be designed as BC^* learners[1]. This result solves an open question posed by Fulk and Jain, namely whether recursive functions can be approximated by BC^* learners. Note that BC^* learning can also, in some sense, be considered a form of approximation, as it requires that eventually all of the hypotheses (including those output only finitely often) differ from the target object in only finitely many values. It thus is interesting to see that partial learning can be combined not only with Fulk and Jain's model of approximation, but also with BC^* learning *at the same time*. Note that in this paper, when two learning criteria A and B are said to be combinable, it is generally not assumed that the new learner is effectively constructed from the A-learner and the B-learner.

[1] BC^* is mnemonic for "behaviourally correct with finitely many anomalies" [4].

This raises the question whether partial learners can also be turned into approximate learners in the more general case of learning r.e. languages. Unfortunately, Fulk and Jain's model applies only to learning recursive functions. The second contribution is the design of three notions of approximate learning of r.e. languages, two of which are directly inspired by Fulk and Jain's model. It is then investigated under which conditions partial learners can be modified to fulfill the corresponding constraints of approximate learning. These investigations are also extended to partial learners with additional constraints, such as consistency and conservativeness. It will be shown that while partial learners can always be constructed in a way so that for any given finite set D, their hypotheses will almost always agree with the target language on D, the same does not hold if D must be a finite variant of a fixed infinite set. Thus trade-offs between certain approximate learning constraints and partial learning are sometimes unavoidable – an observation that perhaps has a broader implication in the philosophy of language learning.

Following the line of Fulk and Jain's research, conditions are investigated under which partial language learners can eventually output only finite variants of the target function. While it remains open whether or not partial learners for a given BC^*-learnable class can be made BC^*-learners for this class without losing identification power, some natural conditions on a BC^* learner M are provided under which all classes learnable by M can be learnt by some BC^* learner N that is at the same time a partial learner.

Figure 1 summarises the main results of this paper. *RECPart* and *RECApproxBC*Part* refer respectively to partial learning of recursive functions and approximate BC^* partial learning of recursive functions. The remaining learning criteria are abbreviated (see Definitions 1, 2 and 6), and denote learning of classes of r.e. languages. An arrow from criterion A to criterion B means that the collection of classes learnable under model A is contained in that learnable under model B. Each arrow is labelled with the Corollary/Example/Remark/Theorem number(s) that proves (prove) the relationship represented by the arrow. If there is no path from A to B, then the collection of classes learnable under model A is not contained in that learnable under model B.

2 Preliminaries

The notation and terminology from recursion theory adopted in this paper follows in general the book of Rogers [14]. Background on inductive inference can be found in [9]. The symbol $\mathbb{N}$ denotes the set of natural numbers, $\{0, 1, 2, \ldots\}$. Let $\varphi_0, \varphi_1, \varphi_2, \ldots$ denote a fixed acceptable numbering [14] of all partial-recursive functions over $\mathbb{N}$. Given a set S, S^* denotes the set of all finite sequences in S. Wherever no confusion may arise, S will also denote its own characteristic function, that is, for all $x \in \mathbb{N}$, $S(x) = 1$ if $x \in S$ and $S(x) = 0$ otherwise. One defines the e-th r.e. set W_e as $dom(\varphi_e)$ and the e-th canonical finite set by choosing D_e such that $\sum_{x \in D_e} 2^x = e$. This paper fixes a one-one padding function pad with $W_{pad(e,d)} = W_e$ for all e, d. Furthermore, $\langle x, y \rangle$ denotes Cantor's

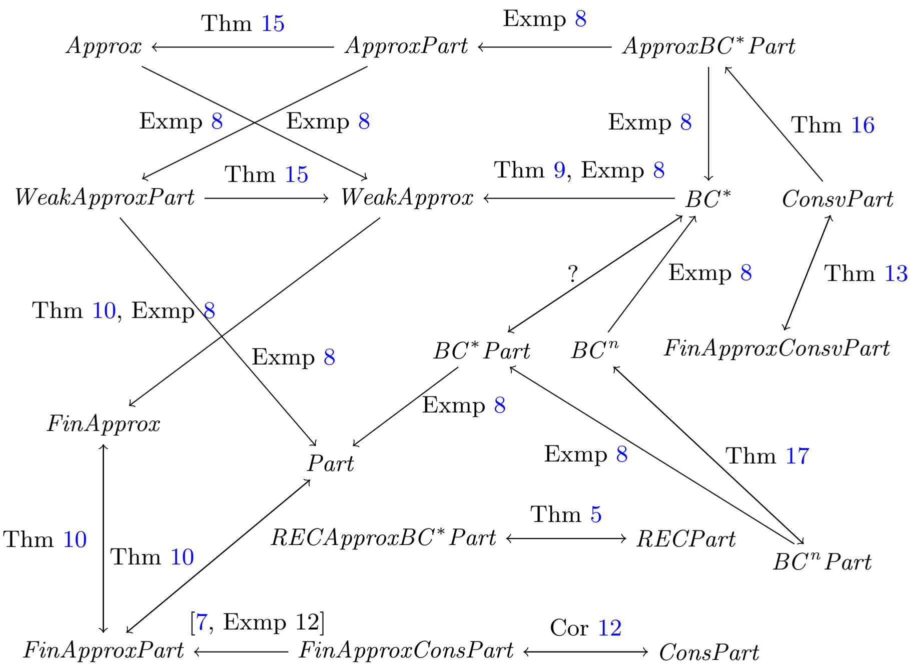

Fig. 1. Learning hierarchy

pairing function, given by $\langle x, y \rangle = \frac{1}{2}(x + y)(x + y + 1) + y$. A triple $\langle x, y, z \rangle$ denotes $\langle\langle x, y \rangle, z\rangle$.

For any $\sigma, \tau \in (\mathbb{N} \cup \{\#\})^*, \sigma \preceq \tau$ if and only if σ is a prefix of τ, $\sigma \prec \tau$ if and only if σ is a proper prefix of τ, and $\sigma(n)$ denotes the element in the nth position of σ, starting from $n = 0$. The concatenation of two strings σ and τ shall be denoted by $\sigma \circ \tau$; for convenience, and whenever there is no possibility of confusion, this is occasionally denoted by $\sigma\tau$. Let $\sigma[n]$ denote the sequence $\sigma(0) \circ \sigma(1) \circ \ldots \circ \sigma(n - 1)$. The length of σ is denoted by $|\sigma|$.

3 Learning

The basic learning paradigms studied in the present paper are *behaviourally correct learning* [2,3] and *partial learning* [13]. These learning models assume that the learner is presented with just positive examples of the target language, and that the learner is fed with a finite amount of data at every stage. They are modifications of the model of explanatory learning (or "learning in the limit"), first introduced by Gold [8], in which the learner must output in the limit a single correct representation h of the target language L; if L is an r.e. set, then h is usually an r.e. index of L with respect to the standard numbering $W_0, W_1, W_2, \ldots$ of all r.e. sets. Bārzdiņš [2] and Case [3] considered the more powerful model of behaviourally correct learning, whereby the learner must almost always output a correct hypothesis of the input set, but some of the correct hypotheses may be syntactically distinct. Case and Smith [4] also introduced a less stringent variant of BC learning of recursive functions, BC^* learning, which only requires the learner to output in the limit finite variants of the target recursive function. Still more general is the criterion of partial learning that Osherson, Stob and Weinstein [13] defined; in this model, the learner must output exactly one correct index of the input set infinitely often and output any other conjecture only finitely often.

One can also impose constraints on the quality of a learner's hypotheses. For example, Angluin [1] introduced the notion of *consistency*, which is the requirement that the learner's hypotheses must enumerate at least all the data seen up to the current stage. This seems to be a fairly natural demand on the learner, for it only requires that the learner's conjectures never contradict the available data on the target language. Angluin [1] also introduced the learning constraint of *conservativeness*; intuitively, a conservative learner never makes a mind change unless its prior conjecture does not enumerate all the current data. These two learning criteria have since been adapted to the partial learning model [6,7].

The learning criteria discussed so far (and, where applicable, their partial learning analogues) are formally introduced below.

Let $\mathcal{C}$ be a class of r.e. sets. Throughout this paper, the mode of data presentation is that of a *text*, by which is meant an infinite sequence of natural numbers and the # symbol. Formally, a *text* T_L for some L in $\mathcal{C}$ is a map $T_L : \mathbb{N} \rightarrow \mathbb{N} \cup \{\#\}$ such that $L = range(T_L)$; here, $T_L[n]$ denotes the sequence $T_L(0) \circ T_L(1) \circ \ldots \circ T_L(n-1)$ and the range of a text T, denoted $range(T)$, is the set of numbers occurring in T. Analogously, for a finite sequence σ, $range(\sigma)$ is the set of numbers occurring in σ. A text, in other words, is a presentation of positive data from the target set. A *learner*, denoted by M in the following definitions, is a recursive function mapping $(\mathbb{N} \cup \{\#\})^*$ into $\mathbb{N}$.

Definition 1. (i) [13] M *partially* (*Part*) learns $\mathcal{C}$ if, for every L in $\mathcal{C}$ and each text T_L for L, there is exactly one index e such that $M(T_L[k]) = e$ for infinitely many k; furthermore, if M outputs e infinitely often on T_L, then $L = W_e$.

(ii) [3] M *behaviourally correctly* (*BC*) learns $\mathcal{C}$ if, for every L in $\mathcal{C}$ and each text T_L for L, there is a number n for which $L = W_{M(T_L[j])}$ whenever $j \geq n$.

(iii) [1] M is *consistent* (*Cons*) if for all $\sigma \in (\mathbb{N} \cup \{\#\})^*$, $range(\sigma) \subseteq W_{M(\sigma)}$.

(iv) [1] For any text T, M is *consistent on* T if $range(T[n]) \subseteq W_{M(T[n])}$ for all $n > 0$.

(v) [7] M is said to *consistently partially* (*ConsPart*) learn $\mathcal{C}$ if it partially learns $\mathcal{C}$ from text and is consistent.

(vi) [6] M is said to *conservatively partially* (*ConsvPart*) learn $\mathcal{C}$ if it partially learns $\mathcal{C}$ and outputs on each text for every L in $\mathcal{C}$ exactly one index e with $L \subseteq W_e$.

(vii) [4] M is said to *behaviourally correctly* learn $\mathcal{C}$ *with at most a anomalies* (BC^a) iff for every $L \in \mathcal{C}$ and each text T_L for L, there is a number n for which $|(W_{M(T_L[j])} - L) \cup (L - W_{M(T_L[j])})| \leq a$ whenever $j \geq n$.

(viii) [4] M is said to *behaviourally correctly* learn $\mathcal{C}$ *with finitely many anomalies* (BC^*) iff for every $L \in \mathcal{C}$ and each text T_L for L, there is a number n for which $|(W_{M(T_L[j])} - L) \cup (L - W_{M(T_L[j])})| < \infty$ whenever $j \geq n$.

This paper will also consider combinations of different learning criteria; for learning criteria $A_1, \ldots, A_n$, a class $\mathcal{C}$ is said to be $A_1 \ldots A_n$-learnable iff there is a learner M such that M A_i-learns $\mathcal{C}$ for all $i \in \{1, \ldots, n\}$. Due to space constraints, some proofs of formal statements are omitted throughout this paper. For the full version of the paper, see http://arxiv.org/abs/1507.01215.

4 Approximate Learning of Functions

Fulk and Jain [5] proposed a mathematically rigorous definition of *approximate inference*, a notion originally motivated by studies in the philosophy of science.

Definition 2. [5] An approximate (*Approx*) learner outputs on the graph of a function f a sequence of hypotheses such that there is a sequence $S_0, S_1, \ldots$ of sets satisfying the following conditions:
(a) The S_n form an ascending sequence of sets such that their union is the set of all natural numbers;
(b) There are infinitely many n such that $S_{n+1} - S_n$ is infinite;
(c) The n-th hypothesis is correct on all $x \in S_n$ but nothing is said about the $x \notin S_n$.

The next proposition simplifies this set of conditions. The proof is omitted.

Proposition 3. *M Approx learns a recursive function f iff the following conditions hold:*
(d) For all x and almost all n, M's n-th hypothesis is correct at x;
(e) There is an infinite set S such that for almost all n and all $x \in S$, M's n-th hypothesis is correct at x.

Fulk and Jain interpreted their notion of approximation as a process in scientific inference whereby physicists take the limit of the average result of a sequence of experiments. Their result that the class of recursive functions is approximately learnable seems to justify this view.

Theorem 4 (Fulk and Jain [5]). *There is a learner M that Approx learns every recursive function.*

The following theorem answers an open question posed by Fulk and Jain [5] on whether the class of recursive functions has a learner which outputs a sequence of hypotheses that approximates the function to be learnt and almost always differs from the target only on finitely many places.

Theorem 5. *There is a learner M which learns the class of all recursive functions such that (i) M is a BC^* learner, (ii) M is a partial learner and (iii) M is an approximate learner.*

Proof. Let $\psi_0, \psi_1, \ldots$ be an enumeration of all recursive functions and some partial ones such that in every step s there is exactly one pair (e, x) for which $\psi_e(x)$ becomes defined at step s and this pair satisfies in addition that $\psi_e(y)$ is already defined by step s for all $y < x$. Furthermore, a function ψ_e is said to make progress on σ at step s iff $\psi_e(x)$ becomes defined at step s and $x \in dom(\sigma)$ and $\psi_e(y) = \sigma(y)$ for all $y \leq x$.

Now one defines for every σ a partial-recursive function $\vartheta_{e,\sigma}$ as follows:

- $\vartheta_{e,\sigma}(x) = \sigma(x)$ for all $x \in dom(\sigma)$;
- Let $e_t = e$;
- Inductively for all $s \geq t$, if some index $d < e_s$ makes progress on σ at step $s+1$ then let $e_{s+1} = d$ else let $e_{s+1} = e_s$;
- For each value $x \notin dom(\sigma)$, if there is a step $s \geq t + x$ for which $\psi_{e_s,s}(x)$ is defined then $\vartheta_{e,\sigma}(x)$ takes this value for the least such step s, else $\vartheta_{e,\sigma}(x)$ remains undefined.

The learner M, now to be constructed, uses these functions as hypothesis space; on input τ, M outputs the index of $\vartheta_{e,\sigma}$ for the unique e and shortest prefix σ of τ such that the following three conditions are satisfied at some time t:

- t is the first time such that $t \geq |\tau|$ and some function makes progress on τ;
- ψ_e is that function which makes progress at τ;
- for every $d < e$, ψ_d did not make progress on τ at any $s \in \{|\sigma|, \ldots, t\}$ and either $\psi_{d,|\sigma|}$ is inconsistent with σ or $\psi_{d,|\sigma|}(x)$ is undefined for at least one $x \in dom(\sigma)$.

For finitely many strings τ there might not be any such function $\vartheta_{e,\sigma}$, as τ is required to be longer than the largest value up to which some function has made progress at time $|\tau|$, which can be guaranteed only for almost all τ. For these finitely many exceptions, M outputs a default hypothesis, e.g., for the everywhere undefined function. Now the three conditions (i), (ii) and (iii) of M are verified. For this, let ψ_d be the function to be learnt, note that ψ_d is total.

Condition (i): M is a BC^* learner. Let d be the least index of the function ψ_d to be learnt and let u be the last step where some ψ_e with $e < d$ makes progress on ψ_d. Then every $\tau \preceq \psi_d$ with $|\tau| \geq u + 1$ satisfies that first $M(\tau)$ conjectures a function $\vartheta_{e,\sigma}$ with $e \geq d$ and $|\sigma| \geq u + 1$ and $\sigma \preceq \psi_d$ and second that almost all e_s used in the definition of $\vartheta_{e,\sigma}$ are equal to d; thus the function computed is a finite variant of ψ_d and M is a BC^* learner.

Condition (ii): M is a partial learner. Let $t_0, t_1, \ldots$ be the list of all times where ψ_d makes progress on itself with $u < t_0 < t_1 < \ldots$.; note that whenever $\tau \preceq \psi_d$ and $|\tau| = t_k$ for some k then the conjecture $\vartheta_{e,\sigma}$ made by $M(\tau)$ satisfies $e = d$ and $|\sigma| = u+1$. As none of these conjectures make progress from step $u+1$ onwards on ψ_d, they also do not make progress on σ after step $|\sigma|$ and $\vartheta_{e,\sigma} = \psi_d$; hence the learner outputs some index for ψ_d infinitely often. Furthermore, all other indices $\vartheta_{e,\sigma}$ are output only finitely often: if $e < d$ then ψ_e makes no progress on the target function ψ_d after step u; if $e > d$ then the length of σ depends on the prior progress of ψ_d on itself, and if $|\tau| > t_k$ then $|\sigma| > t_k$.

Condition (iii): M is an approximate learner. Conditions (d) and (e) in Proposition 3 are used. Now it is shown that, for all $\tau \preceq \psi_d$ with $t_k \leq |\tau| < t_{k+1}$, the hypothesis $\vartheta_{e,\sigma}$ issued by $M(\tau)$ is correct on the set $\{t_0, t_1, \ldots\}$. If $|\tau| = t_k$ then the hypothesis is correct everywhere as shown under condition (ii). So assume that $e > d$. Then $|\tau| > t_k$ and $|\sigma| > t_k$, hence $\vartheta_{e,\sigma}(x) = \psi_d(x)$ for all $x \leq t_k$. Furthermore, as ψ_d makes progress on σ in step t_{k+1} and as no ψ_c with $c < d$ makes progress on σ beyond step $|\sigma|$, it follows that the e_s defined in the algorithm of $\vartheta_{e,\sigma}$ all satisfy $e_s = d$ for $s \geq t_{k+1}$; hence $\vartheta_{e,\sigma}(x) = \psi_d(x)$ for all $x \geq t_{k+1}$. ∎

5 Approximate Learning of Languages

This section proposes three notions of approximation in language learning. The first two notions, *approximate* learning and *weak approximate* learning, are adaptations of the set of conditions for approximately learning recursive functions given in Proposition 3. Recall that a set V is a finite variant of a set W iff there is an x such that for all $y > x$ it holds that $V(y) = W(y)$.

Definition 6. Let S be a class of languages. S is *approximately (Approx) learnable* iff there is a learner M such that for every language $L \in S$ there is an infinite set W such that for all texts T and all finite variants V of W and almost all hypotheses H of M on T, $H \cap V = L \cap V$. S is *weakly approximately (WeakApprox) learnable* iff there is a learner M such that for every language $L \in S$ and for every text T for L there is an infinite set W such that for all finite variants V of W and almost all hypotheses H of M on T, $H \cap V = L \cap V$. S is *finitely approximately (FinApprox) learnable* iff there is a learner M such that for every language $L \in S$, all texts T for L, and any finite set D, it holds that for almost all hypotheses H of M on T, $H \cap D = L \cap D$.

Remark 7. Jain, Martin and Stephan [10] defined a partial-recursive function C to be an *In-classifier* for a class S of languages if, roughly speaking, for every $L \in S$, every text T for L, every finite set D and almost all n, C on $T[n]$ will correctly "classify" all $x \in D$ as either belonging to L or not belonging to L. A learner M that *FinApprox* learns a class S may be translated into a total In-classifier for S, and vice versa.

Approximate learning requires, for each target language, the existence of a set W suitable for all texts, while in weakly approximate learning the set W may depend on T. In the weakest notion, finitely approximate learning, on any text T for a target language L the learner is only required to be almost always correct on any finite set. As will be seen later, this model is so powerful that the whole class of r.e. sets can be finitely approximated by a partial learner. The following examples illustrate the models of approximate and weakly approximate learning.

Example 8. – If there is an infinite r.e. set W such that all members of $\mathcal{C}$ contain W then $\mathcal{C}$ is *Approx* learnable: the learner simply conjectures $range(\sigma) \cup W$ on any input σ. Such $\mathcal{C}$ is not necessarily BC^* learnable.

- If $\mathcal{C}$ consists only of coinfinite r.e. sets then $\mathcal{C}$ is *Approx* learnable.
- The class of all cofinite sets is BC^* learnable and *WeakApproxBC*Part* learnable but neither *Approx* learnable nor BC^n learnable for any n.
- The class of all infinite sets is *WeakApprox* learnable.
- Gold's class consisting of the set of natural numbers and all sets $\{0, 1, \ldots, m\}$ is not *WeakApprox* learnable.

The proofs are omitted. These examples establish that, in contrast to the function learning case, approximate language learnability does not imply BC^* learnability. BC^* learnability does not imply approximate learnability either, but weakly approximate learning is powerful enough to cover all BC^* learnable classes.

Theorem 9. *If $\mathcal{C}$ is BC^* learnable then $\mathcal{C}$ is WeakApprox learnable.*

Proof. By Example 8, there is a learner M that weakly approximates the class of all infinite sets. Let O be a BC^* learner for $\mathcal{C}$. Now the new learner N is given as follows: On input σ, $N(\sigma)$ outputs an index of the following set which first enumerates $range(\sigma)$ and then searches for some τ that satisfies the following conditions: (1) $range(\tau) = range(\sigma)$; (2) $|\tau| = 2 * |range(\sigma)|$; (3) $W_{O(\tau\#^s)}$ enumerates at least $|\sigma|$ many elements for all $s \leq |\sigma|$. If all three conditions are met then the set contains also all elements of $W_{M(\sigma)}$. Further details are omitted. ∎

6 Combining Partial Language Learning With Variants of Approximate Learning

This section is concerned with the question whether partial learners can always be modified to approximate the target language in the models introduced above.

6.1 Finitely Approximate Learning

The first results demonstrate the power of the model of finitely approximate learning: there is a partial learner that finitely approximates every r.e. language.

Theorem 10. *The class of all r.e. sets is FinApproxPart learnable.*

Proof. Let M_1 be a partial learner of all r.e. sets. Define a learner M_2 as follows. Given a text T, let $e_n = M_1(T[n+1])$ for all n. On input $T[n+1]$, M_2 determines the finite set $D = range(T[n+1]) \cap \{0, \ldots, m\}$, where m is the minimum $m \leq n$ with $e_m = e_n$. M_2 then outputs a canonical index for $D \cup (W_{e_n} \cap \{x : x > m\})$.

Suppose T is a text for some r.e. set L. Then there is a least l such that M_1 on T outputs e_l infinitely often and $W_{e_l} = L$. Furthermore, there is a least l' such that for all $l'' > l'$, $D_L = range(T[l''+1]) \cap \{0, \ldots, l\} = L \cap \{0, \ldots, l\}$. Hence M_2 will output a canonical index for $L = D_L \cup (W_{e_l} \cap \{x : x > l\})$ infinitely often. On the other hand, since, for every h with $e_h \neq e_l$ and $e_h \neq e_{h'}$ for all $h' < h$, M_1 outputs e_h only finitely often, M_2 will conjecture sets of the form $D' \cup (W_{e_h} \cap \{x : x > h\})$ only finitely often. Thus M_2 partially learns L.

To see that M_2 is also a finitely approximate learner, consider any number x. Suppose that M_1 on T outputs exactly one index e infinitely often; further, $W_e = L$ and j is the least index such that $e_j = e$. Let s be sufficiently large so that for all $s' > s$, $range(T[s'+1]) \cap \{0, \ldots, \max(\{x, j\})\} = L \cap \{0, \ldots, \max(\{x, j\})\}$. First, assume that M_1 outputs only finitely many distinct indices on T. It follows that M_1 on T converges to e. Thus M_2 almost always outputs a canonical index for $(L \cap \{0, \ldots, j\}) \cup (W_{e_j} \cap \{y : y > j\})$, and so it approximately learns L. Second, assume that M_1 outputs infinitely many distinct indices on T. Let $d_1, \ldots, d_x$ be the first x conjectures of M_1 that are pairwise distinct and are not equal to e. There is a stage $t > s$ large enough so that $e_{t'} \notin \{d_1, \ldots, d_x\}$ for all $t' > t$. Consequently, whenever $t' > t$, M_2 on $T[t'+1]$ will conjecture a set W such that $W \cap \{0, \ldots, x\} = L \cap \{0, \ldots, x\}$. This establishes that M_2 finitely approximately learns any r.e. set. ∎

Gao, Jain and Stephan [6] showed that consistently partial learners exist for all and only the subclasses of uniformly recursive families; the next theorem shows that such learners can even be finitely approximate at the same time, in addition to being *prudent*. A learner M is *prudent* if it learns the class $\{W_{M(\sigma)} : \sigma \in (\mathbb{N} \cup \{\#\})^*, M(\sigma) \neq ?\}$, that is, if M learns every set it conjectures [12]; the ? symbol allows M to abstain from conjecturing at any stage..

Theorem 11. *If $\mathcal{C}$ is a uniformly recursive family, then $\mathcal{C}$ is FinApproxConsPart learnable by a prudent learner.*

Proof. Let $\mathcal{C} = \{L_0, L_1, L_2, \ldots, \}$ be a uniformly recursive family. On text T, define M at each stage s as follows:

If there are $x \in \mathbb{N}$ and $i \in \{0, 1, \ldots, s\}$ such that
- $range(T[s+1]) - range(T[s]) = \{x\}$,
- $range(T[s+1]) \subseteq L_i \cup \{\#\}$ and
- $range(T[s+1]) \cap \{0, \ldots, x\} = L_i \cap \{0, \ldots, x\}$

Then M outputs the least such i
Else M outputs a canonical index for $range(T[s+1]) - \{\#\}$.

The consistency of M follows directly by construction. If T is a text for a finite set then the "Else-Case" will apply almost always and M converges to a canonical index for $range(T)$. Now consider that T is a text for some infinite set $L_m \in \mathcal{C}$ and m is the least index of itself. Let t be large enough so that for all $t' > t$, all $x \in L - range(T[t+1]) - \{\#\}$ and all $j < m$, $L_j \cap \{0, \ldots, x\} \neq range(T[t'+1]) \cap \{0, \ldots, x\}$. There are infinitely many stages $s > \max(\{t, m\})$ at which $T(s) \notin range(T[s]) \cup \{\#\}$ and $range(T[s+1]) \cap \{0, \ldots, T(s)\} = L \cap \{0, \ldots, T(s)\}$. At each of these stages, M will conjecture L_m. Thus M conjectures L_m infinitely often. Furthermore, for every x there is some s_x such that for all $y \in L -$

$range(T[s_x+1])$, it holds that $y > x$. Thus whenever $s' > s_x$, M's conjecture on $T[s'+1]$ agrees with L on $\{0,\dots,x\}$. M is therefore a finitely approximate learner, implying that it never conjectures any incorrect index infinitely often. ∎

Corollary 12. *If $\mathcal{C}$ is ConsPart learnable, then $\mathcal{C}$ is FinApproxConsPart learnable by a prudent learner.*

The following result shows that also conservative partial learning may always be combined with finitely approximate learning.

Theorem 13. *If $\mathcal{C}$ is ConsvPart learnable, then $\mathcal{C}$ is FinApproxConsvPart learnable.*

Proof. Let M_1 be a *ConsvPart* learner for $\mathcal{C}$, and suppose that M_1 outputs the sequence of conjectures $e_0, e_1, \dots$ on some given text T. The construction of a new learner M_2 is similar to that in Theorem 10; however, one has to ensure that M_2 does not output more than one index that is either equal to or a proper superset of the target language. On input $T[s+1]$, define $M_2(T[s+1])$ as follows.

1. If $range(T[s+1]) \subseteq \{\#\}$ then output a canonical index for $\emptyset$ else go to 2.
2. Let $m \le s$ be the least number such that $e_m = e_s$. If $W_{e_s,s} \cap \{0,\dots,m\} = range(T[s+1]) \cap \{0,\dots,m\} = D$ then output a canonical index for $D \cup (W_{e_m} \cap \{x : x > m\})$ else go to 3.
3. If $s \ge 1$ then output $M_2(T[s])$ else output a canonical index for $\emptyset$.

The details for verifying that M_2 is a *ConsvPart* learner for $\mathcal{C}$ are omitted. ∎

6.2 Weakly Approximate, Approximate and *BC** Learning

The next proposition shows that Theorem 11 cannot be improved and gives a negative answer to the question whether partial or consistent partial learning can be combined with weakly approximate learning.

Proposition 14. *The uniformly recursive class $\{A : A = \mathbb{N}$ or A contains all even and finitely many odd numbers or A contains finitely many even and all odd numbers$\}$ is WeakApprox learnable and ConsPart learnable, but not WeakApproxPart learnable.*

The next theorem shows that neither partial learning nor consistent partial learning can be combined with approximate learning. In fact, it establishes a stronger result: consistent partial learnability and approximate learnability are insufficient to guarantee both partial and weakly approximate learnability simultaneously.

Theorem 15. *There is a class of r.e. sets with the following properties:*
(i) The class is not BC^* learnable;
(ii) The class is not WeakApproxPart learnable;
(iii) The class is Approx learnable.

Proof. The key idea is to diagonalise against a list $M_0, M_1, \ldots$ of learners which are all total and which contains for every learner to be considered a delayed version. This permits to ignore the case that some learner is undefined on some input.

The class witnessing the claim consists of all sets L_d such that for each d, either L_d is $\{d, d+1, \ldots\}$ or L_d is a subset built by the following diagonalisation procedure: One assigns to each number $x \geq d$ a level $\ell(x)$.

- If some set $L_{d,e} = \{x \geq d : \ell(x) \leq e\}$ is infinite then
- let $L_d = L_{d,e}$ for the least such e and M_d does not partially learn L_d
- else let $L_d = \{d, d+1, \ldots\}$ and M_d does not weakly approximate L_d.

The construction of the sets is inductive over stages. For each stage $s = 0, 1, 2, \ldots$:

- Let τ_e be a sequence of all $x \in \{d, d+1, \ldots, d+s-1\}$ with $\ell(x) = e$ in ascending order;
- If there is an $e < s$ such that e has not been cancelled in any previous step and for each $\eta \preceq \tau_e$ the intersection $W_{M_d(\tau_0\tau_1\ldots\tau_{e-1}\eta),s} \cap \{y : d \leq y < d+s \wedge \ell(y) > e\}$ contains at least $|\tau_e|$ elements
  - Then choose the least such e and let $\ell(d+s) = e$ and cancel all e' with $e < e' \leq s$
  - Else let $\ell(d+s) = s$.

A text $T = \lim_e \sigma_e$ is defined as follows (where σ_0 is the empty sequence):

- Let τ_e be the sequence of all x with $\ell(x) = e$ in ascending order;
- If σ_e is finite then let $\sigma_{e+1} = \sigma_e\tau_e$ else let $\sigma_{e+1} = \sigma_e$.

In case some σ_e are infinite, let e be smallest such that σ_e is infinite. Then $T = \sigma_e$ and $L_d = L_{d,e}$ and T is a text for L_d. As $L_{d,e}$ is infinite, one can conclude that

$$\forall \eta \preceq \sigma_e \, \forall c \, [|W_{M_d(\tau_0\tau_1\ldots\tau_{e-1}\eta)} \cap \{y : \ell(y) > e\}| \geq c]$$

and thus M_d outputs on T almost always a set containing infinitely many elements outside L_d; so M_d does neither partially learn L_d nor BC^* learn L_d.

In case all σ_e are finite and therefore all $L_{d,e}$ are finite there must be infinitely many e that never get cancelled. Each such e satisfies

$$\exists \eta \preceq \tau_e \, [W_{M_d(\tau_0\tau_1\ldots\tau_{e-1}\eta)} \cap \{y : \ell(y) > e\} \text{ is finite}]$$

and therefore e also satisfies $\exists \eta \preceq \tau_e \, [W_{M_d(\tau_0\tau_1\ldots\tau_{e-1}\eta)}$ is finite]. Thus M_d outputs on the text T for the cofinite set $L_d = \{d, d+1, \ldots\}$ infinitely often a finite set and M_d is neither weakly approximately learning L_d (as there is no infinite set on which almost all conjectures are correct) nor BC^*-learning L_d. Thus claims (i) and (ii) are true.

Next it is shown that the class of all L_d is approximately learnable by some learner N. This learner N will on a text for L_d eventually find the minimum d needed to compute the function ℓ. Once N has found this d, N will on each input σ conjecture the set

$$W_{N(\sigma)} = \{x : x \geq \max(range(\sigma)) \vee \exists y \in range(\sigma) \, [\ell(x) \leq \ell(y)]\}$$

In case $L_d = L_{d,e}$ for some e, $L_{d,e}$ is infinite, and for each text for $L_{e,d}$, almost all prefixes σ of this text satisfy $\max\{\ell(y) : y \in range(\sigma)\} = e$ and $L_{d,e} \subseteq W_{N(\sigma)}$. So almost all conjectures are correct on the infinite set L_d itself. Furthermore, $W_{N(\sigma)}$ does not contain any $x < \max(range(\sigma))$ with $\ell(x) > e$, hence N eventually becomes correct also on any $x \notin L_{d,e}$ and therefore N approximates $L_{d,e} = L_d$.

In case $L_d = \{d, d+1, \ldots\}$, all $L_{d,e}$ are finite. Then consider the infinite set $S = \{x : \forall y > x\,[\ell(y) > \ell(x)]\}$. Let $x \in S$ and consider any σ with $\min(range(\sigma)) = d$. If $x \geq \max(range(\sigma))$ then $x \in W_{N(\sigma)}$. If $x < \max(range(\sigma))$ then $\ell(\max(range(\sigma))) \geq \ell(x)$ and again $x \in W_{N(\sigma)}$. Thus $W_{N(\sigma)}$ contains S. Furthermore, for all $x \geq d$ and sufficiently long prefixes σ of the text, $\ell(\max(range(\sigma))) \geq \ell(x)$ and therefore all $x \in W_{N(\sigma)}$ for almost all prefixes σ of the text. So again N approximates L_d. Thus claim (iii) is true. ∎

One can further show that the class in the above proof is explanatorily learnable if the learner has access to an oracle for the jump of the halting set.

While these negative results suggest that approximate and weakly approximate learning imposes constraints that are too stringent for combining with partial learning, at least partly positive results can be obtained. For example, the following theorem shows that *ConsvPart* learnable classes are *ApproxPart* learnable (thus dropping only the conservativeness constraint) by BC^* learners. This considerably improves an earlier result by Gao, Stephan and Zilles [7] which states that every *ConsvPart* learnable class is also BC^* learnable.

Theorem 16. *If $\mathcal{C}$ is ConsvPart learnable then $\mathcal{C}$ is ApproxPart learnable by a BC^* learner.*

Proof. Let M be a *ConsvPart* learner for $\mathcal{C}$. For a text T for a language $L \in \mathcal{C}$, one considers the sequence $e_0, e_1, \ldots$ of distinct hypotheses issued by M; it contains one correct hypothesis while all others are not indices of supersets of L. For each hypothesis e_n one has two numbers tracking its quality: $b_{n,t}$ is the maximal $s \leq n+t$ such that all $T(u)$ with $u < s$ are in $W_{e_n,n+t} \cup \{\#\}$ and $a_{n,t} = 1 + \max\{b_{m,t} : m < n\}$.

Now one defines the hypothesis set $H_{e_n,\sigma}$ for any sequence σ. Let $e_{n,0}, e_{n,1}, \ldots$ be a sequence with $e_{n,0} = e_n$ and $e_{n,u}$ be the e_m for the minimum m such that $m = n$ or W_{e_m} has enumerated all members of $range(\sigma)$ within $u+t$ time steps. The set $H_{e_n,\sigma}$ contains all x for which there is a $u \geq x$ with $x \in W_{e_{n,u}}$.

An intermediate learner O now conjectures some canonical index of a set $H_{e_n,\sigma}$ at least k times iff there is a t with $\sigma = T(0)T(1)\ldots T(a_{n,t})$ and $b_{n,t} > k$. Thus O conjectures $H_{e_n,\sigma}$ infinitely often iff W_{e_n} contains $range(T)$ and $a_{n,t} = |\sigma|$ for almost all t.

If e_n is the correct index for the set to be learnt then, by conservativeness, the sets W_{e_m} with $m < n$ are not supersets of the target set. So the values $b_{m,t}$ converge which implies that $a_{n,t}$ converges to some s. It follows that for the prefix σ of T of length s, the canonical index of $H_{e_n,\sigma}$ is conjectured infinitely often while no other index is conjectured infinitely often. Thus O is a partial learner. Furthermore, for all sets $H_{e_m,\tau}$ conjectured after $a_{n,t}$ has reached its final value s, it holds that the $e_{m,u}$ in the construction of $H_{e_m,\tau}$ converge to e_n. Thus $H_{e_m,\tau}$ is the union of W_{e_n} and a finite set. Hence O is a BC^* learner. To guarantee the third condition on approximate learning, O will be translated into another learner N.

Let $d_0, d_1, \ldots$ be the sequence of O output on the text T. Now N will copy this sequence but with some delay. Assume that $N(\sigma_k) = d_k$ and σ_k is a prefix of T. Then N will keep the hypothesis d_k until the current prefix σ_{k+1} considered satisfies either $range(\sigma_{k+1}) \nsubseteq range(\sigma_k)$ or $W_{d_k,|\sigma_{k+1}|} \neq range(\sigma_{k+1})$.

If $range(T)$ is infinite, the sequence of hypotheses of N will be the same as that of O, only with some additional delay. Furthermore, almost all W_{d_n} contain $range(T)$, thus the resulting learner N learns $range(T)$ and is almost always correct on the infinite set $range(T)$; in addition, N learns $range(T)$ partially and is also BC^*. If $range(T)$ is finite, there will be some correct index that equals infinitely many d_n. There is a step t by which all elements of $range(T)$ have been seen in the text and enumerated into W_{d_n}. Therefore, when the learner conjectures this correct index again, it will never withdraw it; furthermore, it will replace eventually every incorrect conjecture due to the comparison of the two sets. Thus the learner converges explanatorily to $range(T)$ and is also in this case learning $range(T)$ in a BC^* way, partially and approximately. From the proof of Theorem 10, one can see that N may be translated into a learner satisfying all the three requirements (a), (b) and (c). ∎

Case and Smith [4] published Harrington's observation that the class of recursive functions is BC^* learnable. This result does not carry over to the class of r.e. sets; for example, Gold's class consisting of the set of natural numbers and all finite sets is not BC^* learnable. In light of Theorem 5, which established that the class of recursive functions can be BC^* and *Part* learnt simultaneously, it is interesting to know whether *any* BC^* learnable class of r.e. sets can be both BC^* and *Part* learnt at the same time. While this question in its general form remains open, the next result shows that BC^n learning is indeed combinable with partial learning.

Theorem 17. *Let $n \in \mathbb{N}$. If $\mathcal{C}$ is BC^n learnable, then $\mathcal{C}$ is Part learnable by a BC^n learner.*

Proof. Fix any n such that $\mathcal{C}$ is BC^n learnable. Given a recursive BC^n learner M of $\mathcal{C}$, one can construct a new learner N_1 as follows. First, let $F_0, F_1, F_2, \ldots$ be a one-one enumeration of all finite sets such that $|F_i| \leq n$ for all i. Fix a text T, and let $e_0, e_1, e_2, \ldots$ be the sequence of M's conjectures on T.

For each set of the form $W_{e_i} \cup F_j$ (respectively $W_{e_i} - F_j$), N_1 outputs a canonical index for $W_{e_i} \cup F_j$ (respectively $W_{e_i} - F_j$) at least m times iff the following two conditions hold.

1. There is a stage $s > j$ for which the number of distinct $x < j$ such that either $x \in W_{e_i,s} \wedge x \notin range(T[s+1])$ or $x \in range(T[s+1]) \wedge x \notin W_{e_i,s}$ holds does not exceed n.
2. There is a stage $t > m$ such that for all $x < m$, $x \in W_{e_i,t} \cup F_j$ iff $x \in range(T[t+1])$ (respectively $x \in W_{e_i,t} - F_j$ iff $x \in range(T[t+1])$).

At any stage $T[s+1]$ where no set of the form $W_{e_i} \cup F_j$ or $W_{e_i} - F_j$ satisfies the conditions above, or each such set has already been output the required number of times (up to the present stage), N_1 outputs $M(T[s+1])$. The details showing that a $BC^n Part$ learner N for $\mathcal{C}$ can be constructed from N_1 are omitted. ∎

Theorems 18 and 19 show that partial BC^* learning is possible for classes that can be BC^* learned by learners that satisfy some additional constraints. The proofs are omitted.

Theorem 18. *Assume that $\mathcal{C}$ is BC^* learnable by a learner that outputs on each text for any $L \in \mathcal{C}$ at least once a fully correct hypothesis. Then $\mathcal{C}$ is Part learnable by a BC^* learner.*

Theorem 19. *Suppose there is a recursive learner that BC^* learns $\mathcal{C}$ and outputs on every text for any $L \in \mathcal{C}$ at least one index infinitely often. Then there is a recursive learner for $\mathcal{C}$ that BC^* and Part learns $\mathcal{C}$.*

References

1. Angluin, D.: Inductive inference of formal languages from positive data. Information and Control **45**(2), 117–135 (1980)
2. Bārzdiņš, J.: Two theorems on the limiting synthesis of functions. In: Theory of Algorithms and Programs, vol. 1, pp. 82–88. Latvian State University (1974) (in Russian)
3. Case, J., Lynes, C.: Machine inductive inference and language identification. In: Nielsen, M., Schmidt, E.M. (eds.) Automata, Languages and Programming. LNCS, vol. 140, pp. 107–115. Springer, Heidelberg (1982)
4. Case, J., Smith, C.: Comparison of identification criteria for machine inductive inference. Theoretical Computer Science **25**, 193–220 (1983)
5. Fulk, M., Jain, S.: Approximate inference and scientific method. Information and Computation **114**, 179–191 (1994)
6. Gao, Z., Jain, S., Stephan, F.: On conservative learning of recursively enumerable languages. In: Bonizzoni, P., Brattka, V., Löwe, B. (eds.) CiE 2013. LNCS, vol. 7921, pp. 181–190. Springer, Heidelberg (2013)
7. Gao, Z., Stephan, F., Zilles, S.: Partial learning of recursively enumerable languages. In: Jain, S., Munos, R., Stephan, F., Zeugmann, T. (eds.) ALT 2013. LNCS, vol. 8139, pp. 113–127. Springer, Heidelberg (2013)
8. Gold, E.M.: Language identification in the limit. Information and Control **10**, 447–474 (1967)
9. Jain, S., Osherson, D., Royer, J.S., Sharma, A.: Systems that learn: an introduction to learning theory. MIT Press (1999)
10. Jain, S., Martin, E., Stephan, F.: Learning and classifying. Theoretical Computer Science **482**, 73–85 (2013)
11. Martin, E., Osherson, D.N.: Elements of scientific inquiry. MIT Press (1998)
12. Osherson, D.N., Stob, M., Weinstein, S.: Learning strategies. Information and Control **53**, 32–51 (1982)
13. Osherson, D.N., Stob, M., Weinstein, S.: Systems that learn: an introduction to learning theory for cognitive and computer scientists. MIT Press (1986)
14. Rogers, Jr., H.: Theory of recursive functions and effective computability. MIT Press (1987)