Theoretical Computer Science 410 (2009) 1796–1806

Contents lists available at ScienceDirect

Theoretical Computer Science

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

Prescribed learning of r.e. classes

Sanjay Jain [a,*], Frank Stephan [a,b], Nan Ye [a]

[a] *Department of Computer Science, National University of Singapore, Singapore 117417, Republic of Singapore*

[b] *Department of Mathematics, National University of Singapore, Singapore 117543, Republic of Singapore*

ARTICLE INFO

Keywords:
Inductive inference
Prescribed learning
Formal languages
Recursion theory

ABSTRACT

This work extends studies of Angluin, Lange and Zeugmann on the dependence of learning on the hypothesis space chosen for the language class in the case of learning uniformly recursive language classes. The concepts of class-comprising (where the learner can choose a uniformly recursively enumerable superclass as the hypothesis space) and class-preserving (where the learner has to choose a uniformly recursively enumerable hypothesis space of the same class) are formulated in their study. In subsequent investigations, uniformly recursively enumerable hypothesis spaces have been considered. In the present work, we extend the above works by considering the question of whether learners can be effectively synthesized from a given hypothesis space in the context of learning uniformly recursively enumerable language classes. In our study, we introduce the concepts of *prescribed learning* (where there must be a learner for every uniformly recursively enumerable hypothesis space of the same class) and *uniform learning* (like prescribed, but the learner has to be synthesized effectively from an index of the hypothesis space). It is shown that while for explanatory learning, these four types of learnability coincide, some or all are different for other learning criteria. For example, for conservative learning, all four types are different. Several results are obtained for vacillatory and behaviourally correct learning; three of the four types can be separated, however the relation between prescribed and uniform learning remains open. It is also shown that every (not necessarily uniformly recursively enumerable) behaviourally correct learnable class has a prudent learner, that is, a learner using a hypothesis space such that the learner learns every set in the hypothesis space. Moreover the prudent learner can be effectively built from any learner for the class.

1. Introduction

Human learning is a process in which general principles, which are useful for reasoning about human activities and the external environments, are synthesized from a limited amount of data. Illustrative examples include the acquisition of languages by children and the discovery of scientific laws. What is the mechanism that enables human beings to learn so many different and apparently unrelated things that they are able to learn? Understanding such mechanism, and hopefully automating such mechanism using machines, are of fundamental importance.

Although it is still not yet clear how human beings learn, it is generally agreed that an important mechanism in human learning is inductive inference [2,26], which is the process of observing more and more examples, and forming a sequence of conjectures which eventually converges to one single conjecture that explains all the examples. A natural question is how powerful inductive inference is. Put it in another way, what are the classes of concepts that can be learned using

* Corresponding author.
E-mail addresses: sanjay@comp.nus.edu.sg (S. Jain), fstephan@comp.nus.edu.sg (F. Stephan), g0701171@nus.edu.sg (N. Ye).

0304-3975/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2009.01.011

inductive inference? To answer this question, Gold formulated a recursion-theoretic framework of inductive inference [11]. A recursion-theoretic framework is natural and significant in two ways.

- First, if we assume that human learning is a computational process and we accept the Church–Turing thesis, then a recursion-theoretic model for inductive inference is sound and suitable for answering what classes of concepts are learnable using inductive inference.
- Second, a recursion-theoretic model at least tells us the limits of what computers can learn automatically, irrespective of any assumption on the nature of human learning.

Since the publication of Gold's framework, many works have been done to extend it. We briefly mention a few examples below. Gold considered only the simplest success criteria for learning: the sequence of conjectures eventually converges to a single correct conjecture. But just as done in machine learning, it is natural to require some performance metrics in the learning process, and thus various natural success criteria have been studied, such as conservative learning [1] and monotonic learning [24]. While Gold used acceptable numberings of r.e. languages as possible hypothesis spaces, Angluin considered using indexed families (or uniformly recursive language classes) as hypothesis spaces for learning of indexed families [1], and she characterized learnable indexed families. Many other interesting questions have been studied and a survey of the subject can be found in [12].

A basic characteristic of inductive inference is that inductive inference is always based on some *inductive bias*, which is the assumption on what kinds of hypotheses are possible. For example, in machine learning, when we use decision trees or neural networks to learn some concepts, we are assuming that the concepts can be represented using decision trees or neural networks. The same remark applies to human learning as well.

A natural question stemming from this observation is whether the inductive bias, that is, the choice of hypothesis space is critical in learning. In addition, are we able to learn in some systematic way such that whenever we have a hypothesis space covering the concepts to be learned, then we are guaranteed to learn the target concepts correctly? That is, are we able to *effectively synthesize* a correct learning method for any given hypothesis space covering all target concepts?

Lange and Zeugmann have shown that the choice of hypothesis space can be critical, in the sense that the hypothesis space chosen can affect whether a language class is learnable, even when the hypothesis space covers all target languages. In this paper, we extend the work of Lange and Zeugmann and study the question of synthesizing learners from hypothesis spaces in the context of learning uniformly r.e. language classes. It is shown that although it is not always possible to effectively synthesize learners from hypothesis spaces, there indeed exist non-trivial language classes for which this can be done.

Note that previously learnability of uniformly r.e. classes had been considered by de Jongh, Kanazawa [8] and Zilles [28,29]. De Jongh and Kanazawa [8] gave a characterization of when uniformly r.e. classes of languages are learnable in the limit. Zilles [28,29] had considered effective synthesis of learners from indices for r.e. classes of languages (although in her case, it was the target class which varied with the index, rather than the hypothesis space as in our case).

Another interesting question on inductive inference is whether a learnable class can always be learned by a prudent learner, that is, a learner which only outputs conjectures for languages which can be learned by it. A general observation is that machine learning algorithms are generally prudent. It has already been shown that for several learning criteria, a prudent learner exists for a learnable class. We prove that the answer is positive for the case of behaviourally correct learning, which has been open since 1988.

We proceed formally in the following paragraphs.

Remark 1. We first introduce some basic recursion-theoretic notations and some basic notations for inductive inference. Let $W_0, W_1, W_2, \ldots$ be an acceptable enumeration of all r.e. subsets of the set of natural numbers $\mathbb{N}$. A language is an r.e. subset of natural numbers. Let φ_e denote the e-th partial-recursive function, again from an acceptable numbering. For more information on recursion theory, the reader is referred to standard textbooks like the ones of Odifreddi [19] and Soare [23]. The function $\langle e, x\rangle = \frac{1}{2}\cdot(e+x)(e+x+1)+x$ is Cantor's pairing function. A family $L_0, L_1, L_2, \ldots$ is uniformly recursively enumerable iff $\{\langle e, x\rangle : x \in L_e\}$ is a recursively enumerable set. For ease of notation, uniformly r.e. classes are just called *r.e. classes*. Note that in this paper, notations like $\{L_0, L_1, L_2, \ldots\}$ are used as a short-hand for both, the family as well as for the class of the sets; so set-theoretic comparisons like $\{L_0, L_1, L_2, \ldots\} \subseteq \{H_0, H_1, H_2, \ldots\}$ and $\{L_0, L_1, L_2, \ldots\} = \{H_0, H_1, H_2, \ldots\}$ ignore the ordering of the sets inside the class. Furthermore, let $W_{e,s}, L_{e,s}, H_{e,s}$ be the elements enumerated within time s into W_e, L_e, H_e, respectively. Without loss of generality, $W_{e,s}, L_{e,s}, H_{e,s}$ are subsets of $\{0, 1, \ldots, s\}$.

Let σ, τ range over $(\mathbb{N}\cup\{\#\})^*$. Furthermore, let $\sigma \subseteq \tau$ denote that τ is an extension of σ as a string. content(σ) denotes the set of natural numbers in the range of σ. T is a text if T maps $\mathbb{N}$ to $\mathbb{N}\cup\{\#\}$ and T is a text for L_a iff the numbers occurring in T are exactly those in L_a. content(T) denotes the set of natural numbers in the range of T. $T[n]$ denotes the string consisting of the first n members of the text T, so $T[0]$ is the empty string and $T[2] = T(0)T(1)$.

Definition 2. A learner is a recursive function from $(\mathbb{N}\cup\{\#\})^*$ to $\mathbb{N}\cup\{?\}$. In the following, let M be a learner and let $\{L_0, L_1, L_2, \ldots\}, \{H_0, H_1, H_2, \ldots\}$ be r.e. classes. Here $\{L_0, L_1, L_2, \ldots\}$ is the class M should learn and $\{H_0, H_1, H_2, \ldots\}$ is the hypothesis space used by M.

The learner M *converges* on T to b if there is an n with $M(T[m]) = b$ for all $m \geq n$.

The learner M is *finite* [11] if for every text T there is one index e such that for all n, either $M(T[n]) = ?$ or $M(T[n]) = e$.

The learner M is *confident* [21] if M converges on every text T to a hypothesis.

The learner M is *conservative* [1] if for all σ, τ with $M(\sigma\tau) \neq M(\sigma)$ there is an x occurring in $\sigma\tau$ such that $x \notin H_{M(\sigma)}$.

The learner M *semantically identifies* L_a if, given any text T for L_a, $H_{M(T[n])} = L_a$ for almost all n. The learner M *syntactically identifies* L_a if, given any text T for L_a, there is a b with $H_b = L_a$ and $M(T[n]) = b$ for almost all n.

The learner M is a *behaviourally correct learner* for $\{L_0, L_1, L_2, \ldots\}$ iff M semantically identifies every L_a [4,7]. M is an *explanatory learner* for $\{L_0, L_1, L_2, \ldots\}$ if M syntactically identifies every L_a [5,11]. M is a *vacillatory learner* for $\{L_0, L_1, L_2, \ldots\}$ iff M is a behaviourally correct learner for $\{L_0, L_1, L_2, \ldots\}$ which on every text for a language L_a outputs only finitely many syntactically different hypotheses [6].

The learner M is *prudent* [10,21] if it learns all languages in its hypothesis space $\{H_0, H_1, H_2, \ldots\}$.

In the first three sections, all classes considered are recursively enumerable, only in Section 4 learnability of general classes is investigated.

Remark 3. Let M be a learner for $\{L_0, L_1, L_2, \ldots\}$ using $\{H_0, H_1, H_2, \ldots\}$ as the hypothesis space. A sequence σ is called a *syntactic stabilizing sequence* for M on a set L iff $\sigma \in (L \cup \{\#\})^*$ and for all $\tau \in (L \cup \{\#\})^*$, $M(\sigma\tau) = M(\sigma)$. A sequence σ is called a *semantic stabilizing sequence* for M on a set L iff $\sigma \in (L \cup \{\#\})^*$ and for all $\tau \in (L \cup \{\#\})^*$, $H_{M(\sigma\tau)} = H_{M(\sigma)}$. Stabilizing sequences are called *locking sequences* for M on L, if in addition to the above conditions it holds that $H_{M(\sigma)} = L$. Note that, if M learns L then stabilizing sequences for M on L are also locking sequences for M on L.

Let K denote the halting problem. Let K' denote the halting problem relative to K. There is a partial-K-recursive function Γ which assigns to each e the length-lexicographically least syntactic stabilizing sequence for M on L_e; $\Gamma(e)$ is defined iff such a sequence exists. Γ has a two-place approximation $\gamma(e, t)$ which converges to $\Gamma(e)$ if $\Gamma(e)$ is defined and diverges otherwise. Note that Γ and γ can be obtained effectively from an index for M and an index e' with $W_{e'} = \{\langle e, x\rangle : x \in L_e\}$. Blum and Blum [5] introduced the notion of locking sequences and Fulk [10] introduced the notion of stabilizing sequences.

Angluin [1] initiated the study of learning indexed families with respect to uniformly recursive hypothesis spaces, rather than uniformly r.e. hypothesis spaces. Lange and Zeugmann [16,25,26] considered the effect of allowing hypothesis spaces to contain languages not in the target language class and allowing the ordering of languages in the hypothesis space to be changed. They investigated the relationship between three types of learning: *exact learning*, where the hypothesis space is just the language class to be learned; *class-preserving learning*, where the hypothesis space consists of the same sets as $\{L_0, L_1, L_2, \ldots\}$; and *class-comprising learning*, where the hypothesis space may contain besides the sets from $\{L_0, L_1, L_2, \ldots\}$, also some other sets. Later, Lange, Kapur and Zeugmann [17,27] extend these studies. In the present work, we extend the study on the dependence of learnability on hypothesis space by introducing the new notions of uniform and prescribed learning to address the problem of synthesizing learners from any given class-preserving hypothesis space. In the following definition, let I range over properties of learners as defined in Remark 2; that is, I can stand for "finite", "explanatory", "conservatively explanatory", "confidently explanatory", "vacillatory" and "behaviourally correct".

Definition 4. $\{L_0, L_1, L_2, \ldots\}$ is *class-comprisingly I* learnable (see [16]) iff it is I learnable with respect to some hypothesis space $\{H_0, H_1, H_2, \ldots\}$; note that learnability automatically implies $\{L_0, L_1, L_2, \ldots\} \subseteq \{H_0, H_1, H_2, \ldots\}$.

$\{L_0, L_1, L_2, \ldots\}$ is *class-preservingly I* learnable (see [16]) iff it is I learnable with respect to some hypothesis space $\{H_0, H_1, H_2, \ldots\}$ satisfying $\{H_0, H_1, H_2, \ldots\} = \{L_0, L_1, L_2, \ldots\}$.

$\{L_0, L_1, L_2, \ldots\}$ is *prescribed I* learnable iff it is I learnable with respect to every hypothesis space $\{H_0, H_1, H_2, \ldots\}$ such that $\{L_0, L_1, L_2, \ldots\} = \{H_0, H_1, H_2, \ldots\}$.

$\{L_0, L_1, L_2, \ldots\}$ is *uniformly I* learnable iff there is a recursive enumeration of partial-recursive functions $M_0, M_1, M_2, \ldots$ such that the following holds: Whenever $\{H_0, H_1, H_2, \ldots\} = \{L_0, L_1, L_2, \ldots\}$ and $W_e = \{\langle d, x\rangle : x \in H_d\}$ then M_e is total and an I learner for $\{L_0, L_1, L_2, \ldots\}$ with respect to this hypothesis space $\{H_0, H_1, H_2, \ldots\}$.

Remark 5. Note that exact learning requires the ordering of the languages in $\{L_0, L_1, L_2, \ldots\}$ be taken into account, while all other definitions hold without paying attention to the specific ordering of the sets inside $\{L_0, L_1, L_2, \ldots\}$. Exact learning and prescribed learning are related : a class $\{L_0, L_1, L_2, \ldots\}$ is prescribed I learnable iff every family $\{H_0, H_1, H_2, \ldots\}$ with $\{H_0, H_1, H_2, \ldots\} = \{L_0, L_1, L_2, \ldots\}$ is exactly I learnable.

The question whether a class can be learned using any given representation is quite natural. It reflects the situation where a company building learners cannot enforce its representation of the data/hypothesis on the clients but has to make for each client a learning algorithm using the client's representation. The difference between prescribed and uniform learning would then be that in the first case the programmers have to adjust for each client the learning program by hand, while in the second case there is some synthesizer which reads the client's requirements from some file and then adapts the learner automatically.

Remark 6. Note that in the case of learning with respect to r.e. families, uniform learning and prescribed learning are defined in a class-preserving way. Jain and Stephan [14] showed that there is a one–one numbering of all r.e. sets (that is a Friedberg numbering [9]) such that only classes with finitely many infinite sets can be behaviourally correctly learned with respect to this numbering as hypothesis space.

Furthermore, the above result can be strengthened to uniform learning by showing that only classes consisting of finite sets are class-comprising uniformly behaviourally correct learnable. To see this, let $\{H_0, H_1, H_2, \ldots\}$ be a Friedberg numbering [9]. For a given parameter e, a family $\{G_0, G_1, G_2, \ldots\}$ is constructed from $\{H_0, H_1, H_2, \ldots\}$ such that the following hold for all a:

- For all b, $G_{\langle a,b\rangle} \subseteq H_a$;

- $G_{\langle a,b\rangle} = H_a$ if either $b = 0 \wedge |W_e| = \infty$ or $b = |W_e| + 1$;
- $G_{\langle a,b\rangle}$ is finite if either $b > 0 \wedge |W_e| = \infty$ or $b \neq |W_e| + 1 \wedge |W_e| < \infty$.

Suppose by way of contradiction that there is an r.e. infinite set H_a such that some class containing H_a can be class-comprising uniformly behaviourally correctly learned. Note that for any fixed e and the class $\{G_0, G_1, G_2, \ldots\}$ with parameter e built as above, there exists exactly one index $\langle f(e), g(e)\rangle$ with $G_{\langle f(e),g(e)\rangle} = H_a$. By construction, $f(e) = a$. By the assumption on uniform learnability, there is a recursive enumeration of learners $N_0, N_1, N_2, \ldots$ such that each N_e learns the given class with respect to the hypothesis space $\{G_0, G_1, G_2, \ldots\}$ built with parameter e. As there is a fixed recursive text T for H_a and one can simulate N_e on T, the function g is limit-recursive (that is, there exists a recursive function h such that $g(x) = \lim_{t\to\infty} h(x, t)$). Note that W_e is infinite iff $g(e) = 0$. As $\{e : |W_e| = \infty\} \not\leq_T K$, this gives a contradiction. So class-comprising uniform behaviourally correct learning only permits learning classes of finite sets.

Thus it is reasonable to restrict oneself to the class-preserving versions of prescribed and uniform learning; this convention has already been adapted in Definition 4.

The next result is obvious from the definitions.

Proposition 7. *For any notion I of learning and any class $\mathcal{L}$, the following implications hold: $\mathcal{L}$ is uniformly I learnable $\Rightarrow$ $\mathcal{L}$ is prescribed I learnable $\Rightarrow$ $\mathcal{L}$ is class-preservingly I learnable $\Rightarrow$ $\mathcal{L}$ is class-comprisingly I learnable.*

It depends on the chosen learning criterion I, which of the implications can be reversed. For finite and explanatory learning, all four notions are the same, as shown in Theorems 8 and 9. A lot of research [12] deals with requiring additional constraints on how hypotheses are chosen during explanatory learning. Such requirements change also the relations between the four types of learning. For confident learning, Theorem 10 shows that the uniform, prescribed and class-preserving types coincide while class-comprising confident learning is more general. For conservative learning, Example 11 gives classes which separate all four types of conservative learning. Theorems 12, 13, 15 and 16 deal with vacillatory and behaviourally correct learning. They give classes which, for these criteria, are class-comprisingly but not class-preservingly learnable as well as classes which are class-preservingly but not prescribed learnable. The separation of prescribed from uniform is open for these two criteria.

The concept of prudent learning in inductive inference was first formalized by Osherson, Stob and Weinstein [20]. Fulk showed that prudence is not restrictive for explanatory learning [10]. Jain and Sharma [13] showed that prudence is not restrictive for vacillatory learning. In Theorem 17 it is shown that prudence is not restrictive for behaviourally correct learning. In 1988, Kurtz and Royer [15] had claimed to have this result, but their proof had a bug and the problem had remained open since then. Furthermore, the construction of the prudent learner is effective in the original learner for a behaviourally correct learnable class. It is still open whether prudence for explanatory and vacillatory learning can be effectivized.

2. Finite and explanatory learning

Finitely learnable classes can be learned uniformly, because a finite learner essentially associates each language with a characteristic finite subset.

Theorem 8. *Every class-comprisingly finitely learnable class is also uniformly finitely learnable.*

Proof. Let M be a finite learner for $\{L_0, L_1, L_2, \ldots\}$ using a class-comprising hypothesis space. Let e be an index for a hypothesis space $\{H_0, H_1, H_2, \ldots\}$. That is, $W_e = \{\langle b, x\rangle : x \in H_b\}$. Further suppose $\{H_0, H_1, H_2, \ldots\} = \{L_0, L_1, L_2, \ldots\}$. Then a learner M_e is defined as follows. $M_e(T[n])$ is defined by the first case below which applies:

- If there is an $m < n$ with $M_e(T[m]) \neq ?$ then $M_e(T[n]) = M_e(T[m])$ for the least such m;
- If there are $m \leq n$ and $b \leq n$ with $M(T[m]) \neq ?$ and $\text{content}(T[m]) \subseteq H_{b,n}$ then $M_e(T[n]) = b$;
- Otherwise $M_e(T[n]) = ?$.

The first condition guarantees that M_e outputs on T at most one hypothesis besides the symbol ?. Hence every M_e is a finite learner. It follows from the definition of finite learning that $H_b = H_c$ whenever $M(T[m]) \neq ?$, $\text{content}(T[m]) \subseteq H_b$ and $\text{content}(T[m]) \subseteq H_c$. Hence the b chosen in the second case is a correct hypothesis whenever this case applies. Furthermore, this case eventually applies on texts for languages in $\{L_0, L_1, L_2, \ldots\}$. This completes the proof that $\{L_0, L_1, L_2, \ldots\}$ is uniformly finitely learnable. □

The same result holds for explanatory learning.

Theorem 9. *Every class-comprisingly explanatorily learnable class is also uniformly explanatorily learnable.*

Proof. Let $\mathcal{L}$ be given and let M be a learner using a hypothesis space $\{L_0, L_1, L_2, \ldots\}$ containing $\mathcal{L}$ and perhaps other languages. Choose i such that $W_i = \{\langle a, x\rangle : x \in L_a\}$.

Fix any j and assume that j is an index of a hypothesis space $\{H_0, H_1, H_2, \ldots\}$ for $\mathcal{L}$, that is, assume $\{H_0, H_1, H_2, \ldots\} = \mathcal{L}$ and $W_j = \{\langle b, x\rangle : x \in H_b\}$. Let Γ_j be the function from Remark 3 which assigns to the members of $\{H_0, H_1, H_2, \ldots\}$ the length-lexicographically least syntactic stabilizing sequences with respect to the learner M. $\gamma_j(b, t)$ is then the t-th approximation of $\Gamma_j(b)$ as defined in Remark 3.

The learner M_j is constructed as follows: $M_j(\sigma)$ is the least b such that either $\gamma_i(M(\sigma), |\sigma|) = \gamma_j(b, |\sigma|)$ or $b = |\sigma|$. The latter condition is just to make M_j total and to terminate the search.

Assume that M converges on some text T to an index a of a language $L_a \in \mathcal{L}$. As $L_a \in \mathcal{L}$, there is a b with $H_b = L_a$; assume that b is the least such index. As $\{H_0, H_1, H_2, \ldots\} = \mathcal{L}$ and M is a learner for $\{H_0, H_1, H_2, \ldots\}$, an index c satisfies $\Gamma_j(c) = \Gamma_i(a)$ iff $H_c = L_a$. Hence M_j converges on T to b as, for all $c < b$ and almost all s, $\gamma_j(b, s) = \gamma_i(a, s)$ and $\gamma_j(c, s) \neq \gamma_i(a, s)$. It follows that M_j learns $\mathcal{L}$ using the hypothesis space $\{H_0, H_1, H_2, \ldots\}$. □

The next result shows that class-preserving confident learning coincides with uniform confident learning. The proof of the second part shows that class-preserving confident learning is not closed under taking subclasses.

Theorem 10. (a) *Every class-preservingly confidently learnable class $\mathcal{L}$ is also uniformly confidently learnable.*

(b) *The class $\{D : |D| = 2 \vee (|D| = 1 \wedge D \subseteq K')\}$ is class-comprisingly but not class-preservingly confidently learnable.*

Proof. (a) Reviewing the proof of Theorem 9, the additional constraints to those given there on M and $\{L_0, L_1, L_2, \ldots\}$ are that $\{L_0, L_1, L_2, \ldots\} = \mathcal{L}$ and M converges on every text to some index. Assume again that j and $\{H_0, H_1, H_2, \ldots\}$ satisfy $\{L_0, L_1, L_2, \ldots\} = \{H_0, H_1, H_2, \ldots\}$ and $W_j = \{\langle b, x\rangle : x \in H_b\}$. Assume that T is any text. Then M converges on T to some index a as M is confident. By construction, M_j converges then to the least index b with $L_a = H_b$. Hence M_j also converges on all texts and hence M_j is confident. Furthermore, M_j learns $\mathcal{L}$ explanatorily with respect to the hypothesis space $\{H_0, H_1, H_2, \ldots\}$.

(b) The class $\{D : |D| = 2 \vee (|D| = 1 \wedge D \subseteq K')\}$ is class-comprisingly confidently learnable as follows. On a text for a set with up to two elements, the learner converges to an index for this set using $\{W_0, W_1, W_2, \ldots\}$ as the hypothesis space. The learner does not revise its hypothesis after seeing three elements in the input, in order to obtain confidence.

Note that $\{D : |D| = 2 \vee (|D| = 1 \wedge D \subseteq K')\}$ is an r.e. class. To see this, note that there is a two-place recursive function g with $x \in K'$ iff $g(x, y) = 1$ for almost all y and $x \notin K'$ iff $g(x, y) = 0$ for infinitely many y. Now let

$$L_{2\langle x,y\rangle} = \{x, x + y + 1\} \text{ and}$$

$$L_{2\langle x,y\rangle+1} = \begin{cases} \{x, x + z + 1\} & \text{if } z \text{ is the least number with } z > y \text{ and } g(x, z) \neq 1; \\ \{x\} & \text{if } g(x, z) = 1 \text{ for all } z > y. \end{cases}$$

It is easy to verify that $\{L_0, L_1, \ldots\} = \{D : |D| = 2 \vee (|D| = 1 \wedge D \subseteq K')\}$. It is easy to see that $\{L_0, L_1, L_2, \ldots\}$ is even an indexed family for the given class. Now assume that some confident learner M for $\{L_0, L_1, L_2, \ldots\}$ uses some hypothesis space $\{H_0, H_1, H_2, \ldots\}$ with $\{H_0, H_1, H_2, \ldots\} = \{L_0, L_1, L_2, \ldots\}$. Then one can define the K-recursive function f with $f(x)$ being the hypothesis to which M converges on the text x^∞. If $x \in K'$ then $H_{f(x)} = \{x\}$ as M learns this set. If $x \notin K'$ then $H_{f(x)} \neq \{x\}$ as no member of $\{H_0, H_1, H_2, \ldots\}$ equals $\{x\}$. The test whether $H_{f(x)} = \{x\}$ is also K-recursive. This would give a contradiction to $K' \not\leq_T K$. Thus there is no class-preserving confident learner for $\{L_0, L_1, L_2, \ldots\}$. □

For conservative learning, a strict hierarchy can be established. Note that the following example can be transferred to many related notions like monotonic [24] and non-U-shaped learning [3] without giving more insight. Therefore, these learning criteria are not considered in the present work.

Example 11. (a) The class $\{D : |D| \leq 1\}$ is prescribed conservatively but not uniformly conservatively learnable.

(b) The class $\{D : |D| < \infty\}$ is class-preservingly conservatively but not prescribed conservatively learnable.

(c) The class $\{D : |D| = 2 \vee (|D| = 1 \wedge D \subseteq K')\}$ is class-comprisingly conservatively but not class-preservingly conservatively learnable.

Proof. (a) The prescribed learner knows the index a of $\emptyset$ in the given numbering $\{H_0, H_1, H_2, \ldots\}$. So it conjectures H_a until a number x occurs in the input and an index b is found with $x \in H_b$. Then the learner makes one mind change to b and keeps this index forever. This learner is conservative and correct as $\{x\}$ is the only set in $\{H_0, H_1, H_2, \ldots\}$ containing x. For the second part, let S be a simple set [22], that is, S is r.e., co-infinite and intersects every infinite r.e. set. For each e, let $S^e = S \cup \{0, 1, \ldots, e\}$ and let $s_0^e, e_1^e, s_2^e, \ldots$ be a uniformly recursive one–one enumeration of S^e. Now define class-preserving hypothesis spaces $\mathcal{H}^0, \mathcal{H}^1, \mathcal{H}^2, \ldots$, where $\mathcal{H}^e = \{H_0^e, H_1^e, H_2^e, \ldots\}$ with $H_x^e = \{y\}$ if $x = s_y^e$ and $H_x^e = \emptyset$ if $x \notin S^e$.

If $\{D : |D| \leq 1\}$ is uniformly conservatively learnable, then there exists a recursive family of learners $N_0, N_1, N_2, \ldots$ such that for all $e \in \mathbb{N}$, N_e conservatively learns the class $\{D : |D| \leq 1\}$ with respect to $\mathcal{H}^e$. The r.e. set $A = \{x :$ for some e, N_e outputs x on $\#^\infty\}$ is infinite (as for all e, N_e outputs an index larger than e) and disjoint to S. This contradicts the fact that S is simple.

Note that the hypothesis spaces $\mathcal{H}^e$ constructed above are uniformly recursive. Thus, $\{D : |D| \leq 1\}$ is not even uniformly class-preservingly conservatively learnable when the hypothesis spaces must be uniformly recursive. A similar observation holds for part (b), too.

(b) The class of all finite sets is clearly conservatively learnable in the canonical numbering of the finite sets. Now let $I_0, I_1, I_2, \ldots$ be a recursive partition of the natural numbers into intervals such that there is a simple set A with $I_n \not\subseteq A$ for all n. Let A_t denote A enumerated within t steps. Let $\{L_0, L_1, L_2, \ldots\}$ be the canonical numbering of the finite sets and let $H_m = L_n$ for $m \in I_n - A$ and $H_m = L_n \cup \{m+n+t, m+n+t+1\}$ for $m \in I_n \cap A$, with $m \in A_t - A_{t-1}$. It is easy to see that $\{H_0, H_1, H_2, \ldots\}$ is also a numbering of all finite sets. Assume now that M is a learner using the hypothesis space $\{H_0, H_1, H_2, \ldots\}$. Let T_x denote the text x^∞. Then one defines a recursive function f as follows: $f(x) = b$ for the first b found such that $x \in H_b$ and $M(T_x[k]) = b$ for some k. As all H_b are finite, the set $\{f(0), f(1), f(2), \ldots\}$ contains infinitely many indices and is recursively enumerable. Hence there is an x with $f(x) \in A$. It follows that $\{x\} \subset H_{f(x)}$ as $H_{f(x)}$ contains at least two elements. So the learner M overgeneralizes on $T_x[k]$ and is not conservative.

(c) In Theorem 10, it has been shown that the class $\{D : |D| = 2 \vee (|D| = 1 \wedge D \subseteq K')\}$ is an r.e. class. The class-comprising confident learner given there is also conservative. Now assume that some conservative learner M for this class uses some class-preserving hypothesis space $\{H_0, H_1, H_2, \ldots\}$. Then one can again define $f(x)$, this time only partial-recursive, to be the b found such that M outputs b on the text x^∞ and $x \in H_b$. Now $x \in K'$ iff $f(x)$ is defined and $H_{f(x)} = \{x\}$. This condition can be checked with oracle K although $K' \not\leq_T K$. From this contradiction it follows that there is no class-preserving conservative learner for $\{D : |D| = 2 \vee (|D| = 1 \wedge D \subseteq K')\}$. □

3. Vacillatory and behaviourally correct learning

For vacillatory and behaviourally correct learning, a strict hierarchy from prescribed to class-preserving to class-comprising learning can be established. It remains open whether uniform learning is more restrictive than prescribed learning.

Theorem 12. *Let $L_{2a} = \{\langle a, b\rangle : b \in \mathbb{N}\}$ and $L_{2a+1} = \{\langle a, b\rangle : b \leq |W_a|\}$. Then $\{L_0, L_1, L_2, \ldots\}$ is uniformly behaviourally correct learnable and class-preservingly vacillatorily learnable but neither prescribed vacillatorily learnable nor class-comprisingly explanatorily learnable.*

Proof. Assume that $\{H_0, H_1, H_2, \ldots\} = \{L_0, L_1, L_2, \ldots\}$ and $W_e = \{\langle b, x\rangle : x \in H_b\}$. Let s be the length and D be the content of the input. Now, uniformly in e, a behaviourally correct learner M_e using the hypothesis space $\{H_0, H_1, H_2, \ldots\}$ is constructed. M_e first computes the sets

- $A = \{c \leq s : D = H_{c,s}\}$ and
- $B = \{c \leq s : D \cap H_{c,s} \neq \emptyset\}$;

then M_e follows the first of the following cases which applies:

- If $D = \emptyset$ then M_e outputs ?;
- If $A \neq \emptyset$ then M_e outputs $\min(A)$;
- If $B \neq \emptyset$ then M_e outputs some $c \in B$ for which $H_{c,s}$ has the largest number of elements;
- Otherwise M_e repeats the previous conjecture.

The first case, together with the last, makes sure that M_e is total, starts with ? and never returns to ? once it has taken another hypothesis. Assume now that M_e sees a text for a language $H_b \in \{L_{2a}, L_{2a+1}\}$ and that b is the least index of H_b in $\{H_0, H_1, H_2, \ldots\}$. Furthermore, assume that so much data have been observed such that the following four conditions hold:

- $s \geq b$;
- The datum $\langle a, 0\rangle$ is in both, D and $H_{b,s}$;
- If $H_b \neq L_{2a+1}$ (and thus, W_a and L_{2a+1} are finite), then $|H_{b,s}| > |L_{2a+1}|$ and $|D| > |L_{2a+1}|$;
- If H_b is finite then $H_b = H_{b,s} = D$ and, for all $d < b$ and $t \geq s$, $H_{d,t} \neq D$.

Note that $D \neq \emptyset$ and $B \neq \emptyset$ and therefore M_e outputs a hypothesis c different from ?. Now it is shown that $H_c = H_b$: First note that $\langle a, 0\rangle \in D$ and $b \in B$, hence the algorithm chooses c either by the second or the third condition in the algorithm. It follows that $H_c = L_{2a}$ or $H_c = L_{2a+1}$. If H_b is finite, it follows directly from the learning algorithm that $b = \min(A)$ for the set A considered there and hence $c = b$. If H_b is infinite and L_{2a+1} is finite, then $|H_c| \geq |H_{b,s}| > |L_{2a+1}|$ and $H_c = L_{2a} = H_b$. If H_b and L_{2a+1} are both infinite then $H_b = L_{2a} = L_{2a+1}$ and $H_c = H_b$. So M_e is a behaviourally correct learner for $\{L_0, L_1, L_2, \ldots\}$ using the hypothesis space $\{H_0, H_1, H_2, \ldots\}$.

To see that $\{L_0, L_1, L_2, \ldots\}$ is class-preservingly vacillatorily learnable, take $H_b = L_b$ for all b. For each language there are at most 2 indices in $\{H_0, H_1, H_2, \ldots\}$ and therefore the above described behaviourally correct learner is also a vacillatory one.

To see that $\{L_0, L_1, L_2, \ldots\}$ is not prescribed vacillatory learnable, one constructs a suitable hypothesis space as follows:

$$H_{\langle a,b\rangle} = \begin{cases} L_{2a+1} & \text{if } b = \min(\{s : |W_{a,s}| = |W_a|\}); \\ L_{2a} & \text{otherwise.} \end{cases}$$

For each a there is a b with $H_{\langle a,b\rangle} = L_{2a+1}$; if W_a is finite then one can take b as the minimum of the non-empty set $\{s : |W_{a,s}| = |W_a|\}$; if W_a is infinite then one can take $b = 0$. The reason for the latter case is that then $L_{2a} = L_{2a+1}$. Furthermore, all but at most one of the b satisfy $L_{2a} = H_{\langle a,b\rangle}$. Hence $\{H_0, H_1, H_2, \ldots\}$ is a hypothesis space for $\{L_0, L_1, L_2, \ldots\}$. If there were a prescribed vacillatory learner using $\{H_0, H_1, H_2, \ldots\}$ as the hypothesis space then there would also be a K-recursive function f such that $f(a)$ is the maximal element output by this learner on the canonical text for L_{2a+1}. It would follow that W_a is finite iff $W_{a,f(a)} = W_a$; note that $f(a) \geq \langle a, b\rangle \geq b$ for the least b such that $L_{2a+1} = H_{\langle a,b\rangle}$. But then a K-recursive procedure could check, given a, whether W_a is finite. As such a procedure does not exist [23], $\{L_0, L_1, L_2, \ldots\}$ is not vacillatorily learnable with respect to the hypothesis space $\{H_0, H_1, H_2, \ldots\}$.

As just seen, $\{L_0, L_1, L_2, \ldots\}$ is not prescribed vacillatorily learnable and hence also not prescribed explanatorily learnable. It follows using Theorem 9 that $\{L_0, L_1, L_2, \ldots\}$ is also not class-comprisingly explanatorily learnable. □

Theorem 13. *For all a, b let*

$$L_{\langle a,b\rangle} = \begin{cases} \{\langle a,c\rangle : c \in \mathbb{N}\} & \text{if } b = 0; \\ \{\langle a,c\rangle : c \le |W_a|\} & \text{if } b = 1; \\ \{\langle a,c\rangle : c \le |W_{a,d}|\} \cup \{\langle a+1, |W_{a,d}|+e+1\rangle\} & \text{if } b = 2+\langle d,e\rangle. \end{cases}$$

The class $\{L_0, L_1, L_2, \ldots\}$ *is class-preservingly behaviourally correct learnable but not prescribed behaviourally correct learnable.*

Proof. Recall that $|W_{a,d}| \le d+1$ for all d. It is easy to see that $\{L_0, L_1, L_2, \ldots\}$ is a uniformly r.e. class. Assume that an input of length s and content D is given. A behaviourally correct learner takes now the first case which applies.

- If there is a pair $\langle a,b\rangle$ such that $\langle a+1, a+b+2\rangle < s$ and $L_{\langle a,b\rangle,s} = D$ then output $\langle a,b\rangle$ for the least pair where these conditions are true.
- If there is an a such that $\{\langle a,0\rangle\} \subseteq D \subseteq L_{\langle a,0\rangle}$ then output $\langle a,0\rangle$.
- Otherwise output ?.

In this context it is assumed that for $b > 1$ and $s > \langle a+1, a+b+2\rangle$, $L_{\langle a,b\rangle,s} = L_{\langle a,b\rangle}$ as one can compute all members directly from the parameters a, b. It is easy to see that this learner succeeds on all finite sets from $\{L_0, L_1, L_2, \ldots\}$. So assume that an infinite set $L_{\langle a,0\rangle}$ is given. If $L_{\langle a,1\rangle} = L_{\langle a,0\rangle}$ then the learner will eventually vacillate between these two indices. If $L_{\langle a,1\rangle} \subset L_{\langle a,0\rangle}$ then $L_{\langle a,1\rangle}$ is finite and as the learner eventually sees an element of $L_{\langle a,0\rangle} - L_{\langle a,1\rangle}$, it will converge to $\langle a,0\rangle$. So $\{L_0, L_1, L_2, \ldots\}$ is class-preservingly behaviourally correct learnable.

Now a hypothesis space is constructed using which $\{L_0, L_1, L_2, \ldots\}$ cannot be behaviourally correctly learned. For all a, b let

$$\begin{aligned} H_{\langle a,0\rangle} &= L_{\langle a,0\rangle}; \\ H_{\langle a,2b+1\rangle} &= L_{\langle a,b+2\rangle}; \\ H_{\langle a,2b+2\rangle} &= \begin{cases} \{\langle a,c\rangle : c \le |W_{a,b}|\} & \text{if } W_{a,b} = W_a; \\ \{\langle a,c\rangle : c \le |W_{a,b}|\} \cup \{\langle a+1, |W_{a,b}|+s+1\rangle\} & \text{if } s \text{ is the least number with } W_{a,b} \subset W_{a,s}. \end{cases} \end{aligned}$$

It is easy to check that this class is an indexed family, that is, $\{H_0, H_1, H_2, \ldots\}$ is uniformly recursive. Thus, if one could behaviourally correctly learn $\{L_0, L_1, L_2, \ldots\}$ using $\{H_0, H_1, H_2, \ldots\}$ as the hypothesis space, one could also explanatorily learn $\{L_0, L_1, L_2, \ldots\}$ using $\{H_0, H_1, H_2, \ldots\}$ (this folklore result is based on the observation that, for the hypothesis space being an indexed family, the mind changes can be delayed until it can be verified that the latter hypothesis differs from the earlier one). Using Theorem 9, this would imply that the class from Theorem 12 (which is contained in $\{L_0, L_1, L_2, \ldots\}$) is prescribed explanatorily learnable and hence prescribed vacillatory learnable. This contradicts Theorem 12. So $\{L_0, L_1, L_2, \ldots\}$ is not prescribed behaviourally correct learnable. □

Corollary 14. *Let* $\{L_0, L_1, L_2, \ldots\}$ *be as defined in Theorem 13. Then* $\{L_0, L_1, L_2, \ldots\} \cup \{\mathbb{N}\}$ *is class-preserving behaviourally correct learnable. Furthermore, no* $\{F_0, F_1, F_2, \ldots\} \supseteq \{L_0, L_1, L_2, \ldots\} \cup \{\mathbb{N}\}$ *is prescribed behaviourally correct learnable.*

Proof. The class-preserving behaviourally correct learner for $\{L_0, L_1, L_2, \ldots\}$ from Theorem 13 can easily be extended to one for $\{L_0, L_1, L_2, \ldots\} \cup \{\mathbb{N}\}$. Let $\{H_0, H_1, H_2, \ldots\}$ be the uniformly recursive hypothesis space for $\{L_0, L_1, L_2, \ldots\}$ from Theorem 13. Now define

$$\begin{aligned} G_0 &= \mathbb{N}; \\ G_{2a+1} &= H_a; \\ G_{2\langle a,b\rangle+2} &= \begin{cases} \mathbb{N} & \text{if there are } c, e, t \text{ with } t > b+c+(e+7)^2 \\ & \text{and } F_{a,t} \ne \emptyset \text{ and either } F_{a,t} = H_e \cap \{0,1,\ldots,t\} \\ & \text{or } \{\langle c,0\rangle, \langle c,1\rangle, \ldots, \langle c,b\rangle\} \subseteq F_{a,t} \subseteq \{\langle c,0\rangle, \langle c,1\rangle, \ldots\}; \\ F_a & \text{otherwise.} \end{cases} \end{aligned}$$

Note that the bound $(e+7)^2$ is used in the formula above to ensure the condition $\max(H_{\langle a,2b+1\rangle}) \le (\langle a, 2b+1\rangle + 7)^2$ for all a, b which is used implicitly in Case (e) below.

Clearly $\{L_0, L_1, L_2, \ldots\} \cup \{\mathbb{N}\} \subseteq \{G_0, G_1, G_2, \ldots\}$. Furthermore, if $\emptyset \in \{F_0, F_1, F_2, \ldots\}$ then $\emptyset \in \{G_0, G_1, G_2, \ldots\}$.

Assume now that F_a is not in $\{L_0, L_1, L_2, \ldots\} \cup \{\emptyset, \mathbb{N}\}$. Let c be the least number such that there is some d with $\langle c,d\rangle \in F_a$; fix this d as well. There are five cases.

Case (a): F_a contains two elements $\langle c',d'\rangle, \langle c'',d''\rangle$ with $c' > c, d'' \ge d'$ and $\langle c',d'\rangle \ne \langle c'',d''\rangle$. Then let b be so large that $\langle c,d\rangle, \langle c',d'\rangle, \langle c'',d''\rangle \in F_{a,b}$. Now it follows that $F_{a,b} \not\subseteq H_e$ for all e and $G_{2\langle a,b\rangle+2} = F_a$.

Case (b): F_a is a union of $\{\langle c', d'+1\rangle\}$ with a subset of $\{\langle c,0\rangle, \langle c,1\rangle, \ldots, \langle c,d'\rangle\}$ for some c', d' with $c' > c$. It is easy to verify that $F_a \ne H_e \cap \{0, 1, \ldots, \langle c', d'+1\rangle\}$ for all e. Let b be so large that $F_{a,b} = F_a$ and $b \ge \langle c', d'+1\rangle$. Then, for all $t > b$, $F_{a,t} \ne H_e \cap \{0, 1, \ldots, t\}$. Hence $G_{2\langle a,b\rangle+2} = F_a$.

Case (c): $F_a \subset \{\langle c, 0\rangle, \langle c, 1\rangle, \ldots\}$ and there are d', d'' with $d' < d''$, $\langle c, d'\rangle \notin F_a$ and $\langle c, d''\rangle \in F_a$. Let b be so large that $\langle c, d''\rangle \in F_{a,b}$ and $b > \langle c, d''\rangle$. Then, for all e and all $t > b$, $F_{a,t} \neq H_e \cap \{0, 1, \ldots, t\}$. Furthermore, the condition $\{\langle c, 0\rangle, \langle c, 1\rangle, \ldots, \langle c, b\rangle\} \subseteq F_{a,t} \subseteq \{\langle c, 0\rangle, \langle c, 1\rangle, \ldots\}$ does not hold. Hence $G_{2\langle a,b\rangle+2} = F_a$.

Case (d): $F_a = \{\langle c, 0\rangle, \langle c, 1\rangle, \ldots, \langle c, d'\rangle\}$ and $d' > |W_c|$. Then there is no H_e with $F_a = H_e \cap \{0, 1, \ldots, \langle c, d'+1\rangle\}$. Taking b so large that $F_{a,b} = F_a$ and $b \geq \langle c, d'+1\rangle$, the condition $\{\langle c, 0\rangle, \langle c, 1\rangle, \ldots, \langle c, b\rangle\} \subseteq F_{a,t} \subseteq \{\langle c, 0\rangle, \langle c, 1\rangle, \ldots\}$ becomes false and thus $G_{2\langle a,b\rangle+2} = F_a$.

Case (e): $F_a = \{\langle c, 0\rangle, \langle c, 1\rangle, \ldots, \langle c, d'\rangle\}$ and $d' < |W_c|$. Let s be so large that $d'+1 \leq |W_{c,s}|$ and take b so large that $\langle c+1, |W_{c,s}|+s+1\rangle \leq b$, $\langle c, d'+1\rangle \leq b$ and $F_{a,b} = F_a$. Then $F_{a,t} \neq H_e \cap \{0, 1, \ldots, t\}$ for all e and $t > b + (e+7)^2$. Furthermore, the condition $\{\langle c, 0\rangle, \langle c, 1\rangle, \ldots, \langle c, b\rangle\} \subseteq F_{a,t} \subseteq \{\langle c, 0\rangle, \langle c, 1\rangle, \ldots\}$ does not hold. Hence $G_{2\langle a,b\rangle+2} = F_a$.

Note that the sets in $\{H_0, H_1, H_2, \ldots\}$ have only odd indices in the numbering $\{G_0, G_1, G_2, \ldots\}$. Hence, given a behaviourally correct learner M for $\{F_0, F_1, F_2, \ldots\}$ using the hypothesis space $\{G_0, G_1, G_2, \ldots\}$, one can build the following new learner N for $\{H_0, H_1, H_2, \ldots\}$ using the hypothesis space $\{H_0, H_1, H_2, \ldots\}$ itself:

$$N(\sigma) = \begin{cases} \frac{1}{2}(M(\sigma) - 1) & \text{if } M(\sigma) \text{ is odd;} \\ ? & \text{if } M(\sigma) \text{ is even or ?.} \end{cases}$$

This contradicts Theorem 13 which showed that such a learner does not exist. □

For the next result, let $I_n = \{2^n - 1, 2^n, 2^n + 1, \ldots, 2^{n+1} - 3, 2^{n+1} - 2\}$ form a partition of the natural numbers into intervals of length 2^n and let C denote the plain Kolmogorov complexity [18]: $C(x) = \min(\{n : \exists y \in I_n\, [U(y) = x]\})$, where U is a fixed universal machine such that the Kolmogorov complexity does not improve by more than an additive constant when U is replaced by some other partial-recursive function. Furthermore, let

$$A = \{m : \exists n\, [m \in I_n \wedge C(m) < 0.4n]\} \text{ and}$$
$$B = \{m : \exists n\, [m \in I_n \wedge C(m) > 0.8n]\}$$

be the sets of numbers of small and large Kolmogorov complexity, respectively.

Theorem 15. *Let A and B be the sets of numbers of small and large Kolmogorov complexity as above. Then the class consisting of $\mathbb{N}$, A and all sets $A \cup \{b\}$ with $b \in B$ is uniformly r.e. and is class-comprisingly but not class-preservingly behaviourally correct learnable.*

Proof. Note that A is recursively enumerable and B is co-r.e.; an indexing of the class is now given by fixing one index $a \in A$ and then letting $L_a = A$, $L_b = A \cup \{b\}$ for all $b \in B$ and $L_b = \mathbb{N}$ for all $b \in \mathbb{N} - B - \{a\}$.

Assume without loss of generality that $C(0) = 0$. Thus, $0 \notin A \cup B$. Hence $\mathbb{N}$ is the only member of $\{L_0, L_1, L_2, \ldots\}$ containing 0. Furthermore, let $D_0, D_1, \ldots$ be a canonical enumeration of all finite sets. Now let

$$H_b = \begin{cases} \mathbb{N} & \text{if } 0 \in D_b; \\ D_b \cup A & \text{if } 0 \notin D_b. \end{cases}$$

Furthermore, one can build a behaviourally correct learner using the hypothesis space $\{H_0, H_1, H_2, \ldots\}$ by conjecturing H_b for the unique b with $D_b = \text{content}(\sigma)$ on input σ. It is easy to verify that this learner succeeds on all languages in $\{H_0, H_1, H_2, \ldots\}$. Therefore $\{L_0, L_1, L_2, \ldots\}$ is class-comprisingly behaviourally correct learnable.

Now assume that M is a class-preserving behaviourally correct learner for $\{L_0, L_1, L_2, \ldots\}$. There is a family $T_0, T_1, \ldots$ of texts and an n such that:

- $T_x[n]$ is a fixed semantic locking sequence for M on A;
- $T_x(n) = x$;
- for all x, the subsequence $T_x(n+1), T_x(n+2), T_x(n+3), \ldots$ of T_x is the same recursive enumeration of A.

Now one defines two sets X and Y according to the behaviour of M on T_x.

- X is the set of all x such that, for some $m > n$, $M(T_x[m])$ conjectures a set containing x;
- Y is the set of all x such that, for some $m > n$, $M(T_x[m])$ conjectures a set containing 0.

Both sets are recursively enumerable. The set Y is disjoint to A as, for all $x \in A$ and all $m > n$, $M(T_x[m])$ is an index of A. As A is a simple set [18], Y is finite. As $A \cup B \subseteq X \subseteq A \cup B \cup Y$, the set $A \cup B$ is recursively enumerable. For each sufficiently large n, at least half of the elements of I_n are in $A \cup B$. Now let J_n be the first $2^{0.6n}$ elements of I_n to be enumerated into $A \cup B$. The J_n are uniformly r.e. and there is a constant c with $C(x) \leq 0.6n + 2\log(n) + c$ for all $x \in J_n$. It follows from the definition of B that $J_n \cap B = \emptyset$ for all sufficiently large n. Hence $J_n \subseteq A \cap I_n$ in contradiction to the fact that $|A \cap I_n| \leq 2^{0.4n}$. This shows that the learner M cannot exist and $\{L_0, L_1, L_2, \ldots\}$ is not class-preservingly behaviourally correct learnable. □

Theorem 16. *There exists an r.e. class $\mathcal{L}$ which is class-comprisingly but not class-preservingly vacillatorily learnable.*

Proof. In the following, let $\langle x, y, z\rangle$ denote $\langle x, \langle y, z\rangle\rangle$. The class $\mathcal{L}$ will be a suitable subclass of the following:

$$L_{\langle e,2a\rangle} = \{\langle e, a, b\rangle : b \in \mathbb{N}\}$$
$$L_{\langle e,2a+1\rangle} = \{\langle e, a, b\rangle : b \leq |W_a|\}.$$

The proof of Theorem 12 can be adapted to show that $\mathcal{L}$ is class-comprisingly vacillatorily learnable. Let

$$G_a^e = \{x : \langle a, x\rangle \in W_e\}$$

be the a-th set in the e-th recursively enumerable hypothesis space. Now define a limit-recursive predicate P as follows:

$$P(e, a, b) = \begin{cases} 1 & \text{if } \forall c < b\,[G_c^e \neq \{\langle e, a, d\rangle : d \leq b\}]; \\ 0 & \text{if } \exists c < b\,[G_c^e = \{\langle e, a, d\rangle : d \leq b\}]. \end{cases}$$

and let

$$\mathcal{L} = \{L_{\langle e,2a\rangle} : e, a \in \mathbb{N}\} \cup \{L_{\langle e,2a+1\rangle} : e, a \in \mathbb{N} \wedge |W_a| < \infty \wedge P(e, a, |W_a|)\}.$$

It is easy to verify that $\mathcal{L}$ is an r.e. class. Suppose by way of contradiction that $\{H_0, H_1, H_2, \ldots\} = \mathcal{L}$ and a learner M vacillatorily learns $\mathcal{L}$ using $\{H_0, H_1, H_2, \ldots\}$ as the hypothesis space. Let e be such that $W_e = \{\langle c, x\rangle : x \in H_c\}$. So $G_c^e = H_c$ for all c. Note that $P(e, a, |W_a|) = 1$ for all a where W_a is finite; the reason is the following chain of implications: $P(e, a, |W_a|) = 0 \Rightarrow \exists c < |W_a|\,[G_c^e = L_{\langle e,2a+1\rangle}] \Rightarrow L_{\langle e,2a+1\rangle} \in \mathcal{L} \Rightarrow P(e, a, |W_a|) = 1$. Thus $L_{\langle e,2a+1\rangle} \in \mathcal{L}$ whenever W_a is finite. Furthermore, $L_{\langle e,2a+1\rangle} \in \mathcal{L}$ whenever W_a is infinite, as then $L_{\langle e,2a+1\rangle} = L_{\langle e,2a\rangle}$. Thus M learns $L_{\langle e,2a+1\rangle}$ for all a.

Let T_a be a text for $L_{\langle e,2a+1\rangle}$ uniformly recursive in the parameter a. Then M on T_a outputs only finitely many indices; let $g(a)$ be the greatest among these indices. It follows that $g(a) \geq |W_a|$ whenever W_a is finite; the reason is that $P(e, a, |W_a|) = 1$, thus no H_c with $c < |W_a|$ equals $L_{\langle e,2a+1\rangle}$. This gives that W_a is finite iff $|W_a| \leq g(a)$. As $g \leq_T K$, $\{a : |W_a| < \infty\} \leq_T K$, a contradiction. □

4. Prudence for behaviourally correct learning

Osherson, Stob and Weinstein [21] were interested in the question whether every learnable class is prudently learnable. Fulk [10] showed that every explanatory learnable class is prudently explanatory learnable. Jain and Sharma [13] showed the corresponding result for vacillatory learning. The next theorem shows this result for behaviourally correct learning. Furthermore, the construction of the prudent learner in the next theorem is effective in the original learner. It is still open whether prudence for explanatory and vacillatory learning can be effectivized.

Theorem 17. *If $\mathcal{L}$ is a (not necessarily uniformly r.e.) behaviourally correct learnable class then $\mathcal{L}$ is a subclass of an r.e. class which is class-preservingly behaviourally correct learnable.*

Proof. For any set A, let T_A be the ascending text which is given by $T_A(x) = x$ for all $x \in A$ and $T_A(x) = \#$ for all $x \notin A$. Furthermore, let $\delta_\emptyset$ be the empty string and $\delta_A = T_A[\max(A) + 1]$ for all finite non-empty sets A. For example, $\delta_{\{0,2,3\}} = 0\,\#\,2\,3$.

There is a behaviourally correct learner for the class $\mathcal{L}$ using the acceptable numbering $\{W_0, W_1, W_2, \ldots\}$ as the hypothesis space and satisfying the following constraints:

- M is consistent, that is, $\text{content}(\sigma) \subseteq W_{M(\sigma)}$ for all σ;
- M is rearrangement-independent, that is, $W_{M(\sigma)} = W_{M(\tau)}$ whenever σ, τ have the same content and length;
- $W_{M(\sigma)}$ is finite whenever σ is not a semantical locking sequence for M on $W_{M(\sigma)}$.

Kurtz and Royer [15] showed that the first two conditions can be satisfied and such a learner can be found effectively from any given learner. The third condition can also be effectively added since the complement of the set of semantical locking sequences is K-r.e.; that is, σ is not a semantical locking sequence iff there are a τ in $(W_{M(\sigma)} \cup \{\#\})^*$ and an $x \in \mathbb{N}$ with $x \in W_{M(\sigma\tau)} \Leftrightarrow x \notin W_{M(\sigma)}$. For that reason, M is a behaviourally correct learner for all infinite sets for which some index is output by M. So, to prove the theorem, one has to mainly take care of finite sets.

Now the following new learner N is constructed. N is defined by mapping σ to a hypothesis H_σ; thus the hypothesis space is given directly instead of N. H_σ takes the first case which applies.

Intuitively, Case (2) below handles learnability of all infinite sets behaviourally learned by M, besides ensuring some nice properties of H_{δ_D} (the main one being that H_{δ_D} does not contain any element in $\{0, 1, \ldots, \max(D)\} - D$ or it follows H_{δ_F} for some appropriate proper subset F of D). During this process, Case (2) might introduce some finite sets into the hypothesis space. Case (3) ensures learnability of all the finite sets learned by M as well as those introduced by Case (2) in the hypothesis space — for all other finite sets D, Case (3) would mimic Case (2). Case (4) just maps the remaining sequences to one of Case (2) or (3). We now formally define H_σ.

Case (1): $H_{\#^s} = \emptyset$ for all s.
Case (2): H_{δ_D} first enumerates all elements of D.
Let $D' = \{0, 1, \ldots, \max(D)\} - D$. Let $S = \{s : W_{M(\delta_D\#^{\max(D)}),s} \cap D' = \emptyset\}$.
For all $s \in S$, enumerate all elements of $W_{M(\delta_D\#^{\max(D)}),s}$ into H_{δ_D}.
If $W_{M(\delta_D\#^{\max(D)})} \cap D'$ is not empty, let $s = \max(S)$,
let $E = D \cup W_{M(\delta_D\#^{\max(D)}),s}$, let $x = \min(W_{M(\delta_D\#^{\max(D)}),s+1} \cap D')$ and let $F = D \cap \{0, 1, \ldots, x\}$.
Now, if $H_{\delta_F} \supseteq E$ then $H_{\delta_D} = H_{\delta_F}$ else $H_{\delta_D} = E$.
Case (3): $H_{\delta_D\#^s}$ with $s > 0$ is defined as follows. If there is an x such that $H_{\delta_{E_x},s} = H_{\delta_{E_x}} = D$ for the set $E_x = D \cap \{0, 1, \ldots, x\}$ or if $W_{M(\delta_D\#^t)} = D$ for all $t \geq s$ then $H_{\delta_D\#^s} = D$ else $H_{\delta_D\#^s} = H_{\delta_D}$.
Case (4): $H_\sigma = H_{\delta_D\#^s}$ if H_σ is not defined by Cases (1), (2), (3), $s = \max(\{|\sigma| - \max(D) - 1, 0\})$ and $D = \text{content}(\sigma)$.

Note that the only infinite sets in the hypothesis space are the ones which are conjectured by M. So M learns all the infinite sets in the hypothesis space.

Furthermore, for any A in the hypothesis space, if $E_x = \{0, 1, \ldots, x\} \cap A$ and $\delta_{E_x}\#^{\max(E_x)}$ is a semantic locking sequence for M on A, then for all finite D such that $E_x \subseteq D \subseteq A$, $H_{\delta_D} = A$. This can be easily seen by induction on cardinality of $D - E_x$, as in Case (2), either H_{δ_D} is made equal to A or H_{δ_D} would simulate H_{δ_F} for some F such that $E_x \subseteq F \subset D$.

It will be shown first that the hypothesis space covers all sets learned by M and then it will be shown that all sets in the hypothesis space are learned by N.

Clearly if M learns a finite set D then $H_{\delta_D\#^s} = D$ for almost all s. Now consider an infinite set A learned by M. Let $E_x = A \cap \{0, 1, 2, \ldots, x\}$ for all x. As M learns A, there is a semantic locking sequence τ for M on A. Now let $x \in A$ be such that $x > |\tau| + \max(\text{content}(\tau))$. Then, for the sequence $\delta_{E_x}\#^{\max(E_x)}$, there is an $\eta \in (E_x \cup \{\#\})^*$ such that $|\tau\eta| = |\delta_{E_x}\#^{\max(E_x)}|$ and $\text{content}(\tau\eta) = \text{content}(\delta_{E_x}\#^{\max(E_x)}) = E_x$. As M is rearrangement-independent, one has that $W_{M(\delta_{E_x}\#^{\max(E_x)})} = A$. Hence $H_{\delta_{E_x}} = A$. This completes the first part of the verification.

For the second part of the verification consider any set A occurring in the hypothesis space of N. There are three cases, those where A is empty, where A is finite but not empty and where A is infinite.

Case (a): $A = \emptyset$. N learns A as $H_{\#^s} = \emptyset$ for all s by Case (1) in the algorithm to enumerate the hypothesis space.

Case (b): A is finite but not empty. Let D be the smallest set such that $H_{\delta_D\#^s} = A$ for some s. By Case (1) in the algorithm for H_σ, D is not empty.

Assume the subcase $A = H_{\delta_D\#^s} \subset H_{\delta_D}$. By Case (3) and D being the smallest set such that $H_{\delta_D\#^t} = A$ for some t, this can happen only if $A = D$ and $W_{M(\delta_D\#^t)} = D$ for all $t \geq s$. So $H_{\delta_D\#^t} = D$ for all $t \geq s$ and hence N learns A in this subcase as well.

Assume the subcase $A = H_{\delta_D\#^s} = H_{\delta_D}$. Hence, by Case (2) it follows that there is no element in $A - D$ below $\max(D)$ since otherwise $H_{\delta_F} = A$ for some $F \subset D$. Thus, $D = A \cap \{0, 1, \ldots, \max(D)\}$. It follows that $H_{\delta_A\#^t} = A$ for almost all t. Therefore, N learns A.

Case (c): A is infinite. Again, let $E_z = A \cap \{0, 1, \ldots, z\}$ for all z. As M is rearrangement-independent, there is a semantic locking sequence for M on A of the form $\delta_{E_x}\#^{\max(E_x)}$. Hence only finitely many sets $H_{\delta_{E_z}}$ are finite. So there is a $y \in A$ such that $y > x$ and y is greater than all elements of these finite sets $H_{\delta_{E_z}}$. Let F be any finite set with $E_y \subseteq F \subseteq A$. Let $G_z = F \cap \{0, 1, \ldots, z\}$. If $z \geq y$ then $H_{\delta_{G_z}} = A$ (as $E_x \subseteq G_z \subseteq A$) and $H_{\delta_{G_z}} \neq F$. If $z < y$ then $G_z = E_z$ and $H_{\delta_{G_z}} \neq F$ again. Furthermore, M does not learn F. Hence $H_{\delta_F\#^s} = H_{\delta_F} = A$ for all s. So δ_{E_y} is a semantic locking sequence for N on A. It follows that N learns A. This completes the verification that N is a behaviourally correct learner for all the languages in its hypothesis space. □

Acknowledgements

The authors would like to thank James Royer and Sandra Zilles for correspondence. They also thank the anonymous referees of this journal (Theoretical Computer Science) and of the conference Algorithmic Learning Theory (ALT) 2007 in Sendai for comments which helped in improving the paper.

The first and second authors were supported in part by NUS grant numbers R252-000-212-112 and R252-000-308-112.

References

[1] Dana Angluin, Inductive inference of formal languages from positive data, Information and Control 45 (1980) 117–135.
[2] D. Angluin, C. Smith, Inductive inference: Theory and methods, Computing Surveys 15 (1983) 237–289.
[3] Ganesh Baliga, John Case, Wolfgang Merkle, Frank Stephan, Rolf Wiehagen, When unlearning helps, Information and Computation 206 (2008) 694–709.
[4] Janis Bārzdiņš, Two theorems on the limiting synthesis of functions, in: Theory of Algorithms and Programs, Volume 1, Latvian State University, Riga, 210, 1974, pp. 82–88.
[5] Lenore Blum, Manuel Blum, Toward a mathematical theory of inductive inference, Information and Control 28 (1975) 125–155.
[6] John Case, The power of vacillation in language learning, SIAM Journal on Computing 28 (1999) 1941–1969.
[7] John Case, Christopher Lynes, Inductive inference and language identification, in: M. Nielsen, E.M. Schmidt (Eds.), Ninth International Colloquium on Automata, Languages and Programming, ICALP, Aarhus, Denmark, 1982, Proceedings, in: Lecture Notes in Computer Science, vol. 140, Springer, 1982, pp. 107–115.
[8] Dick de Jongh, Makoto Kanazawa, Angluin's theorem for indexed families of r.e. sets and applications, in: Proceedings of the Ninth Annual Conference on Computational Learning Theory, ACM Press, 1996, pp. 193–204.
[9] Richard Friedberg, Three theorems on recursive enumeration, Journal of Symbolic Logic 23 (1958) 309–316.
[10] Mark Fulk, Prudence and other conditions on formal language learning, Information and Computation 85 (1990) 1–11.
[11] E. Mark Gold, Language identification in the limit, Information and Control 10 (1967) 447–474.
[12] Sanjay Jain, Daniel Osherson, James S. Royer, Arun Sharma, Systems That Learn: An Introduction to Learning Theory, MIT-Press, Boston, 1999.
[13] Sanjay Jain, Arun Sharma, Prudence in vacillatory language identification, Mathematical Systems Theory 28 (1995) 267–279.
[14] Sanjay Jain, Frank Stephan, Learning in Friedberg numberings, Information and Computation 206 (2008) 776–790.
[15] Stuart Kurtz, James S. Royer, Prudence in language learning, in: Proceedings of the First Annual Workshop on Computational Learning Theory, MIT, Cambridge, United States, 1988, pp. 143–156.
[16] Steffen Lange, Thomas Zeugmann, Language learning in dependence on the space of hypotheses, in: Proceedings of the Sixth Annual Conference on Computational Learning Theory, Santa Cruz, California, United States, 1993, pp. 127–136.
[17] Steffen Lange, Thomas Zeugmann, Shyam Kapur, Monotonic and dual monotonic language learning, Theoretical Computer Science 155 (1996) 365–410.
[18] Ming Li, Paul Vitányi, An Introduction to Kolmogorov Complexity and Its Applications, Springer, Heidelberg, 1993.
[19] Piergiorgio Odifreddi, Classical Recursion Theory, North-Holland, Amsterdam, 1989.
[20] Daniel Osherson, Michael Stob, Scott Weinstein, Learning strategies, Information and Control 53 (1982) 32–51.

[21] Daniel Osherson, Michael Stob, Scott Weinstein, Systems That Learn, An Introduction to Learning Theory for Cognitive and Computer Scientists, Bradford — The MIT Press, Cambridge, Massachusetts, 1986.
[22] Emil Post, Recursively enumerable sets of positive integers and their decision problems, Bulletin of the American Mathematical Society 50 (1944) 284–316.
[23] Robert Soare, Recursively Enumerable Sets and Degrees, Springer, Heidelberg, 1987.
[24] Rolf Wiehagen, A thesis in inductive inference, in: J. Dix, K. Jantke, P. Schmitt (Eds.), Proceedings First International Workshop on Nonmonotonic and Inductive Logic, Karlsruhe, Germany, Proceedings, in: Lecture Notes in Artificial Intelligence, vol. 543, Springer, 1990, pp. 184–207.
[25] Thomas Zeugmann, Algorithmisches Lernen von Funktionen und Sprachen, Habilitationsschrift, Technische Hochschule Darmstadt, 1993.
[26] Thomas Zeugmann, Steffen Lange, A guided tour across the boundaries of learning recursive languages, in: K.P. Jantke, S. Lange (Eds.), Algorithmic Learning for Knowledge-Based Systems, GOSLER final report, in: Lecture Notes in Artificial Intelligence, vol. 961, Springer, 1995, pp. 193–262.
[27] Thomas Zeugmann, Steffen Lange, Shyam Kapur, Characterizations of monotonic and dual monotonic language learning, Information and Computation 120 (1995) 155–173.
[28] Sandra Zilles, Separation of uniform learning classes, Theoretical Computer Science 313 (2004) 229–265.
[29] Sandra Zilles, Increasing the power of uniform inductive learners, Journal of Computer and System Sciences 70 (2005) 510–538.