Merging Uniform Inductive Learners

Sandra Zilles

Fachbereich Informatik
Universität Kaiserslautern
Postfach 3049
D - 67653 Kaiserslautern
zilles@informatik.uni-kl.de

Abstract. The fundamental learning model considered here is identification of recursive functions in the limit as introduced by Gold [8], but the concept is investigated on a meta-level. A set of classes of recursive functions is *uniformly* learnable under an inference criterion I, if there is a single learner, which synthesizes a learner for each of these classes from a corresponding description of the class. The particular question discussed here is how unions of uniformly learnable sets of such classes can still be identified uniformly. Especially unions of classes leading to strong separations of inference criteria in the uniform model are considered. The main result is that for any pair (I, I') of different inference criteria considered here there exists a fixed set of descriptions of learning problems from I, such that its union with any uniformly I-learnable collection is uniformly I'-learnable, but no longer uniformly I-learnable.

1 Introduction

Inductive Inference is concerned with algorithmic learning of programs for recursive functions. In the fundamental model of identification in the limit, cf. [8], a learner successful for a class of recursive functions must eventually find a single correct program for any function in the class from a gradually growing sequence of its values. By modification of the constraints the learner has to satisfy several inference criteria have been defined and compared with respect to the resulting learning power, see for example [2, 6, 7].

It is a quite natural thought to search for properties that learners in the given abstract model have in common and thus try to find uniform learning methods adequate for the solution of not only one but perhaps infinitely many learning problems. An example for a uniform method is identification by enumeration (cf. [8]). Though this principle does not yield strategies for all learnable classes of recursive functions, it is appropriate for the identification of any recursively enumerable class of recursive functions, if a corresponding enumeration is known. That means, from each enumeration of a set of total recursive functions a program of a learner for the set of functions enumerated can be derived. Another uniform learning method successful in some special hypothesis spaces is temporarily conform identification as defined in [7]. In general uniform learning can be explained as some kind of meta-learning. Instead of considering special strategies each solving a specific learning problem, the aim is to find meta-strategies, which synthesize these special learners from a description of the corresponding learning problem. Such meta-learners reveal a common method for learning the classes of functions and thus a common structure in these classes. They might be helpful in explaining the nature of inductive learning, because they represent some kind of general identification method. For literature on uniform learning the reader is especially referred to [1, 9, 11, 10].

In [9] it is verified, that there are classes of rather simple sets of recursive functions which do not allow the synthesis of identification strategies. Also modifications of identification in the limit are considered. Further results of this nature – moreover concerning language learning – can be found in [11] and [10]. Various modifications of uniform language identification in the limit are presented and compared in [1]. They lead to several separations and nice characterizations and thus give insight into general methods of language learning.

The uniform learning model considered here is equivalent to the one defined in [9], but is investigated in three slightly different variants in order to express the specific difficulties in uniform learning in connection with the choice of hypothesis spaces. The scope now is to investigate unions of collections of learning problems. Given two collections which can be identified uniformly under some inference criterion, is the union still uniformly identifiable? If not, is it identifiable with respect to some weaker inference criterion? What properties of learning problems enable the uniform learnability of unions of such classes? Questions of this style are the main concern throughout the following pages. But why are such questions interesting? The idea is the following: suppose you are given a learnable class. If you want to understand the nature of learning and the structure of learnable classes, it seems quite natural to add more and more objects to the class, until you observe that the resulting class can no longer be identified by a single learner. That means you form the union of two classes and try to find out what unions are still learnable and what unions are not. For example the set of classes identifiable in the limit is not closed under union, as has been verified in [2]. Now the same idea is considered in the uniform learning model. A particular issue in the investigation of learnability of unions is the strong separation of learning classes (see below). On the one hand, some simple properties enabling the uniform learnability of the union of two learnable collections are presented here; on the other hand examples are given to show that many collections of very "simple" learning problems do not allow learnability in the union with other collections.

As in the non-uniform inductive inference model, different inference classes have been compared with respect to their learning power also in the context of meta-learning. Most of the hierarchies verified in the classical model (see [2, 6, 7]) are maintained in the uniform model, see [14]. Hence it might be reasonable to give up certain constraints in the learning model in order to increase the learning power. But still, these separations of inference criteria do not explain, whether any collection of learning problems uniformly identifiable with respect to a criterion I can be *increased* to a superset learnable with respect to some other criterion, but no longer learnable with respect to I. In particular it might give more insight into the structure of inference criteria in uniform learning to find examples of such supersets. Thus an aim might be to transfer *strong separation* results like those in [5] to the meta-learning model. In [5] a strong separation of two inference classes I and I' means that for any object class $U \in I$ there exists some object class $V \in I'$ such that $U \cup V \in I' \setminus I$ (note that this is not a direct consequence of the "weak" separation $I \subset I'$). Still, in the meta-learning model, there might exist a collection of learning problems which is uniformly I-learnable, but no superclass of this collection witnesses to a separation of uniform I'-learning from uniform I-learning. If it were for some reason important to learn at least this collection, then loosening the I-constraints would no longer increase the learning power. Now a strong separation of I and I' ensures that this cannot happen, that means for any collection of learning problems uniformly learnable under the criterion I there exists some collection, such that the union of both is identifiable under I' but not under I. Indeed, such strong separations can be presented for several identification criteria – and even more: for any pair (I, I') of strongly separated criteria there is a *fixed* collection of learning problems which – added to an arbitrary collection uniformly learnable under I – yields a collection appropriate for uniform I'-learning but no longer appropriate for uniform I-learning. Note that this result is much stronger than required before. The proofs of the strong separations moreover provide a method for changing a uniform I-learner into an I'-learner making use of the possible increase of learning power, cf. [5] for similar methods in the non-uniform model.

2 Preliminaries

For notions used here without explicit definition the reader is referred to [12]. $\mathbb{N}$ is used to denote the set of non-negative integers. $\subseteq$ and $\subset$ stand for inclusion and proper inclusion of sets. The cardinality of a set X is denoted by $\operatorname{card} X$. Via some bijective total computable function finite tuples α of integers are identified with elements of $\mathbb{N}$. Given any variable n ranging over $\mathbb{N}$ the quantifier $\forall^\infty n$ expresses that the statement quantified is true for all but finitely many $n \in \mathbb{N}$.

Any total or partial function maps elements of $\mathbb{N}$ again to elements of $\mathbb{N}$. Recursion theory in general concentrates on partial-recursive functions, the set of which is denoted by $\mathcal{P}$. Often the subclass $\mathcal{R}$ of all total functions – called recursive functions – is studied. Both notions may also occur with a superscript indicating the number of input values of the functions considered. $\mathcal{P}_{01}$ stands for the set of partial-recursive functions returning only values from $\{0, 1\}$. Given $f \in \mathcal{P}$ and $n \in \mathbb{N}$, the notion $f(n)\downarrow$ expresses that f is defined on input n, the opposite is denoted by $f(n)\uparrow$. Furthermore $f[n]$ is an abbreviation for the tuple $(f(0), \dots, f(n))$, if all these values are defined. If $f, g \in \mathcal{P}$ and $n \in \mathbb{N}$, then $f =_n g$ means that $\{(x, f(x)) \mid x \leq n \text{ and } f(x)\downarrow\} = \{(x, g(x)) \mid x \leq n \text{ and } g(x)\downarrow\}$. $f =^* g$ expresses that for all but finitely many $n \in \mathbb{N}$ either $f(n)$ and $g(n)$ are both undefined or $f(n) = g(n)$. As functions f, g can be identified with the sequences of their output values, simplified notions might occur, like $f = 0^\infty$ for the function constantly zero or $g = 0^5 \uparrow^\infty$ for the function which equals zero for all inputs less than 5 and is undefined otherwise. If α, β are tuples of integers, $\alpha \subseteq f$ ($\alpha \subseteq \beta$) expresses that α is an initial segment of the function f (of β resp.), and $\alpha \uparrow^\infty$ is called an initial function. If $n \in \mathbb{N}$, then $\overline{n} = 1$ in case $n = 0$ and $\overline{n} = 0$ otherwise. If $\psi \in \mathcal{P}^2$ and $i \in \mathbb{N}$, then ψ_i is the function achieved by fixing the first input parameter of ψ by i. Thus ψ enumerates the set $\{\psi_i \mid i \in \mathbb{N}\} \subseteq \mathcal{P}$. From now on $\varphi \in \mathcal{P}^3$ and $\tau \in \mathcal{P}^2$ denote fixed acceptable numberings. Defining $\varphi^b(x, y) := \varphi(b, x, y)$ for all $b, x, y \in \mathbb{N}$ yields an enumeration $(\varphi^b)_{b \in \mathbb{N}}$ of all two-place partial-recursive functions – concomitant with an enumeration of sets $\mathcal{P}_b := \{\varphi_i^b \mid i \in \mathbb{N}\}$. Given $b \in \mathbb{N}$, the subset $\mathcal{R}_b := \mathcal{P}_b \cap \mathcal{R}$ of recursive functions is called the recursive core of φ^b. Assume for example that, for some fixed $b \in \mathbb{N}$ and all $x, y \in \mathbb{N}$, $\varphi(b, x, y) = 0$ if $x = 0$ or $y < x$, $\varphi(b, x, y) \uparrow$, otherwise. Then $\mathcal{P}_b = \{0^\infty\} \cup \{0^i \uparrow^\infty \mid i \geq 1\}$ and $\mathcal{R}_b = \{0^\infty\}$. A Blum complexity measure as in [4] suggests the notion $f(n) \downarrow_{\leq x}$ to indicate that the computation of $f(n)$ terminates within up to x steps (for $f \in \mathcal{P}$, $n, x \in \mathbb{N}$). If the computation does not terminate or takes more than x steps, then $f(n) \uparrow_{\leq x}$.

The basic learning criterion investigated here is identification in the limit, which has first been studied in [8].

Definition 1. *A set $U \subseteq \mathcal{R}$ of recursive functions is called identifiable in the limit iff there is some $\psi \in \mathcal{P}^2$ and a function $S \in \mathcal{P}$ such that for any $f \in U$:*

1. *$S(f[n])$ is defined for all $n \in \mathbb{N}$ ($S(f[n])$ is called hypothesis on $f[n]$),*
2. *there is some $j \in \mathbb{N}$ such that $\psi_j = f$ and $S(f[n]) = j$ for all but finitely many $n \in \mathbb{N}$.*

EX *is the class of all sets U identifiable in the limit. The notion $U \in \mathrm{EX}_\psi(S)$ shall indicate that S and ψ are known to fulfil the conditions above.*

Weakening the constraints concerning convergence yields the model of behaviourally correct learning as defined in [2]. Here the learner may change its output infinitely often, as long as eventually the hypotheses are correct.

Definition 2. *A set $U \subseteq \mathcal{R}$ is* BC*-identifiable iff there is some $\psi \in \mathcal{P}^2$ and some $S \in \mathcal{P}$ such that any $f \in U$ fulfils $\psi_{S(f[n])} = f$ for all but finitely many n; furthermore $S(f[n])$ is defined for all $f \in U$ and $n \in \mathbb{N}$.* BC *and* $\mathrm{BC}_\psi(S)$ *are notions used as explained in Definition 1.*

Defining several learning models always calls for a comparison of the resulting identification power. In [2] EX is proved to be a proper subclass of BC, i.e. there are classes of recursive functions which are BC-identifiable but not EX-identifiable. Case and Smith [6] propose a variant of BC-learning, called BC-learning with finitely many anomalies, in which correct hypotheses are no longer required. The only restriction here is, that eventually all functions suggested by the learner must have an "almost" correct input-output-behaviour.

Definition 3. *A set $U \subseteq \mathcal{R}$ is* BC^**-identifiable iff, for some $\psi \in \mathcal{P}^2$ and some $S \in \mathcal{P}$, any $f \in U$ satisfies $\psi_{S(f[n])} =^* f$ for all but finitely many n; additionally $S(f[n])$ is defined for all $f \in U$ and $n \in \mathbb{N}$.* BC^* *and* $\mathrm{BC}^*_\psi(S)$ *are defined as usual.*

[6] gives a proof for $\mathrm{BC} \subset \mathrm{BC}^*$ as well as the verification of $\mathcal{R} \in \mathrm{BC}^*$ (proposed by L. Harrington). In any learning process according to the criteria EX, BC, BC^* there is a certain time, after which all hypotheses returned must fulfil certain conditions. But there are no further constraints as to when this time is reached. A restricted learning model with bounds on the number of mind changes is introduced in [6]. In this model the learner is allowed only a certain number of changes in its sequence of hypotheses; in particular, whenever this capacity of mind changes is exhausted, the actual hypothesis must be correct.

Definition 4. *Let $m \in \mathbb{N}$. A set $U \subseteq \mathcal{R}$ is EX_m-learnable iff some $\psi \in \mathcal{P}^2$ and some $S \in \mathcal{P}$ fulfil the following conditions.*

1. *$U \in \mathrm{EX}_\psi(S)$ (where S is additionally permitted to return the sign "?"),*
2. *for all $f \in U$ there is an $n_f \in \mathbb{N}$ such that $S(f[x]) =?$ iff $x < n_f$,*
3. *$\mathrm{card}\{n \in \mathbb{N} \mid ? \neq S(f[n]) \neq S(f[n+1])\} \leq m$ for all $f \in U$.*

The sets $(\mathrm{EX}_m)_\psi(S)$ and EX_m are defined as usual. U is identifiable with a bounded number of mind changes iff there is an $m \in \mathbb{N}$ such that $U \in \mathrm{EX}_m$.

By returning "?" in the initial phase the strategy signals its being hesitant in order not to waste a mind change. For all bounds m the inclusion $\mathrm{EX}_m \subset \mathrm{EX}_{m+1}$ is verified in [6]. Of course in the study of learning processes not only the mind change complexity is interesting, but also whether the quality of the intermediate hypotheses can be improved in some sense. A quite natural motivation is to demand that any hypothesis returned by the learner has to be consistent with the data seen so far, i. e. it should agree with the known part of the input-output-behaviour of the function to be learned, see for example [8, 3, 13].

Definition 5. *A class $U \subseteq \mathcal{R}$ is CONS-identifiable iff it is in $\mathrm{EX}_\psi(S)$ for some $\psi \in \mathcal{P}^2$ and some $S \in \mathcal{P}$ satisfying $\psi_{S(f[n])} =_n f$ for all $f \in U$ and $n \in \mathbb{N}$. The corresponding classes CONS and $\mathrm{CONS}_\psi(S)$ are defined in the usual way.*

The demand for consistency results in a loss of learning power, that means $\mathrm{CONS} \subset \mathrm{EX}$, cf. [3]. Moreover $\mathrm{EX}_0 \subset \mathrm{CONS}$, but EX_m and CONS are incomparable for all $m \geq 1$ (partly verified in [7], the rest follows with similar methods). Note that for all inference criteria defined here – except for CONS – identifiability implies identifiability by a total learner, i. e. whenever there exists a strategy for a class to be learned, then there also exists a strategy $S \in \mathcal{R}$ learning the class. Here consistent learning is an exception, as has been verified in [13].

Throughout the following sections $\mathcal{I}$ denotes the set of inference classes defined above; $\mathcal{I} = \{\mathrm{EX}, \mathrm{BC}, \mathrm{BC}^*, \mathrm{CONS}\} \cup \{\mathrm{EX}_m \mid m \in \mathbb{N}\}$. Moreover, if $n \in \mathbb{N}$, let $J^n := \{U \subseteq \mathcal{R} \mid \mathrm{card}\, U \leq n\}$. Obviously $J^n \in I$ for all $n \in \mathbb{N}$ and all $I \in \mathcal{I}$.

3 The Uniform Learning Model

Uniform learning is concerned with methods for deriving successful learners from a description of a learning problem, which first of all requires a tool for describing learning problems. A quite simple way to describe a class U of recursive functions to be identified is to present some index b of a partial-recursive numbering the recursive core of which equals U. Thus any integer $b \in \mathbb{N}$ corresponds to the class $\mathcal{R}_b = \{\varphi_i^b \mid i \in \mathbb{N}\} \cap \mathcal{R}$ of recursive functions, which is interpreted as a class of objects to be learned. Furthermore any set B of integers corresponds to a collection of classes $U \subseteq \mathcal{R}$ to be learned – each one described by some $b \in B$.

Definition 6. *Let $I \in \mathcal{I}$, $J \subseteq I$. A set $B \subseteq \mathbb{N}$ is suitable for uniform learning with respect to (J, I) iff*

1. *$\mathcal{R}_b \in J$ for all $b \in B$,*
2. *there is a learner $S \in \mathcal{P}^2$ such that for any description $b \in B$ there exists some $\psi \in \mathcal{P}^2$ satisfying $\mathcal{R}_b \in I_\psi(\lambda x.S(b,x))$.*

$\mathrm{suit}(J, I)$ *denotes the set of all sets B suitable in that sense. The notion $B \in \mathrm{suit}(J,I)(S)$ is used to indicate that S is a learner witnessing to $B \in \mathrm{suit}(J,I)$.*

So a set of descriptions is identified uniformly by S under (J, I), if each recursive core described by some $b \in B$ belongs to the class J and is identified by the learner $\lambda x.S(b,x)$ synthesized by S from b. This definition involves some lack of practicability, because only the synthesis of learners but not the synthesis of adequate hypothesis spaces is required. Taking account of this deficiency it is advisable also to consider the following variants of Definition 6.

Definition 7. *Let $I \in \mathcal{I}$, $J \subseteq I$. Then $\mathrm{suit}_\tau(J,I)$ consists of all sets $B \in \mathrm{suit}(J,I)$, for which there is some learner $S \in \mathcal{P}^2$ such that $\mathcal{R}_b \in I_\tau(\lambda x.S(b,x))$ for all $b \in B$. Furthermore $\mathrm{suit}_\varphi(J,I)$ is the set of all $B \in \mathrm{suit}(J,I)$, for which there exists some $S \in \mathcal{P}^2$ which fulfils $\mathcal{R}_b \in I_{\varphi^b}(\lambda x.S(b,x))$ for all $b \in B$. The notions $\mathrm{suit}_\tau(J,I)(S)$ and $\mathrm{suit}_\varphi(J,I)(S)$ are used by analogy with Definition 6.*

Obviously $\mathrm{suit}_\varphi(J,I) \subseteq \mathrm{suit}_\tau(J,I) \subseteq \mathrm{suit}(J,I)$ for all criteria $I \in \mathcal{I}$, but in general equality of these classes does not hold ($\mathrm{suit}_\tau(J, \mathrm{BC}^*) = \mathrm{suit}(J, \mathrm{BC}^*)$ is an exception for any $J \subseteq \mathrm{BC}^*$, see also [14]). As in the non-uniform model, identifiability implies the existence of total strategies, if consistency is not required.

Proposition 1. *Let $I \in \mathcal{I}\backslash\{\mathrm{CONS}\}$, $J \subseteq I$, $B \subseteq \mathbb{N}$. Assume $B \in \mathrm{suit}(J,I)$ ($\mathrm{suit}_\tau(J,I)$, $\mathrm{suit}_\varphi(J,I)$). Then there is a total recursive function S such that $B \in \mathrm{suit}(J,I)(S)$ ($\mathrm{suit}_\tau(J,I)(S)$, $\mathrm{suit}_\varphi(J,I)(S)$ resp.). Moreover, S can be constructed uniformly from a program of a corresponding partial-recursive learner.*

The following example shows that CONS-learning again yields an exception. There even exists a description set suitable for CONS-learning in the most restricted model (suit_φ-model) but not suitable for uniform learning by a total strategy in the least restricted model (suit-model). The proof is omitted.

Example 1. The description set $B := \{b \in \mathbb{N} \mid \varphi^b \in \mathcal{R}^2\}$ is an element of $\mathrm{suit}_\varphi(\mathrm{CONS}, \mathrm{CONS})$, but $B \notin \mathrm{suit}(\mathrm{CONS}, \mathrm{CONS})(S)$ for any learner $S \in \mathcal{R}^2$.

4 Results on Unions of Suitable Description Sets

Assume two description sets $B_1, B_2 \in \mathrm{suit}(I, I)$ for a criterion $I \in \mathcal{I}$ are given. The results below concern the question, what properties concerning B_1, B_2 make the union of the two sets suitable for uniform learning with respect to I, i. e. what properties are sufficient for the satisfaction of $B_1 \cup B_2 \in \mathrm{suit}(I, I)$. The same question is considered for suit_τ or suit_φ instead of suit. $\mathrm{suit}(\mathrm{BC}^*, \mathrm{BC}^*)$ and $\mathrm{suit}_\tau(\mathrm{BC}^*, \mathrm{BC}^*)$ are closed under union, as $\mathcal{R} \in \mathrm{BC}^*$, but in general some further condition concerning B_1 and B_2 is needed, as the following example shows.

Example 2. Consider the description sets $B_1 = \{b \in \mathbb{N} \mid \mathcal{R}_b = \{\varphi_0^b\} = \{0^\infty\}\}$ and $B_2 = \{b \in \mathbb{N} \mid \exists m \in \mathbb{N}\ [\mathcal{R}_b = \{\varphi_1^b\} = \{0^m 1^\infty\}]\}$. Both B_1 and B_2 are elements of $\mathrm{suit}_\varphi(J^1, \mathrm{EX}_0)$, but $B_1 \cup B_2 \notin \mathrm{suit}_\tau(J^1, \mathrm{EX}_0)$.

Proof. The learner constantly $i - 1$ yields $B_i \in \mathrm{suit}_\varphi(J^1, \mathrm{EX}_0)$ for $i \in \{1, 2\}$. Assuming $B_1 \cup B_2 \in \mathrm{suit}_\tau(J^1, \mathrm{EX}_0)$ provides a strategy $S \in \mathcal{R}^2$ which fulfils $\mathcal{R}_b \in (\mathrm{EX}_0)_\tau(\lambda x.S(b, x))$ for all $b \in B_1 \cup B_2$. To reveal a contradiction a description $b_0 \in B_1 \cup B_2$ satisfying $\mathcal{R}_{b_0} \notin (\mathrm{EX}_0)_\tau(\lambda x.S(b_0, x))$ is constructed. For that purpose it is adequate to define numberings ψ^b uniformly in $b \in \mathbb{N}$ and choose b_0 as a fixed point value according to the recursion theorem, such that $\varphi^{b_0} = \psi^{b_0}$.
Construction of b_0. For each $b \in \mathbb{N}$ define a function $\psi^b \in \mathcal{P}^2$ as follows: let $\psi_0^b(0) := 0$. Moreover, let $\psi_0^b(x+1)$ equal 0, if $S(b, 0^{x+1}) =?$ or all $n \leq x + 1$ such that $S(b, 0^n) \in \mathbb{N}$ fulfil $\tau_{S(b,0^n)}(n) \uparrow_{\leq x+1}$ or $\tau_{S(b,0^n)}(n) \neq 0$. Otherwise, $\psi_0^b(x+1)$ is undefined. If $\psi_0^b = 0^k \uparrow^\infty$, there is some minimal integer $n \leq k$, such that $S(b, 0^n) \in \mathbb{N}$ and $\tau_{S(b,0^n)}(n) \downarrow_{\leq k}$, $\tau_{S(b,0^n)}(n) = 0$. In this case let $\psi_1^b := 0^n 1^\infty$. Otherwise, i. e. if $\psi_0^b = 0^\infty$, let $\psi_1^b := \uparrow^\infty$. In any case, all functions ψ_i^b with $i > 1$ shall be empty.

Since ψ^b is defined uniformly in b, there exists some recursive function g, such that $\varphi^{g(b)} = \psi^b$ for all $b \in \mathbb{N}$. Let $b_0 \in \mathbb{N}$ be a fixed point value of that function g, i. e. $\varphi^{b_0} = \varphi^{g(b_0)} = \psi^{b_0}$. *End construction of* b_0.

Note that either $\mathcal{R}_{b_0} = \{\varphi_0^{b_0}\} = \{0^\infty\}$ or $\mathcal{R}_{b_0} = \{\varphi_1^{b_0}\} = \{0^m 1^\infty\}$ for some $m \in \mathbb{N}$. Thus $b_0 \in B_1 \cup B_2$. But if $\mathcal{R}_{b_0} = \{0^\infty\}$, then $S(b_0, 0^n) =?$ or $\tau_{S(b_0,0^n)} \neq 0^\infty$ for all $n \in \mathbb{N}$; if $\mathcal{R}_{b_0} = \{0^m 1^\infty\}$ for some $m \in \mathbb{N}$, then $S(b, 0^m) \in \mathbb{N}$ and $\tau_{S(b,0^m)}(m) = 0 \neq 1 = \varphi_1^{b_0}(m)$. Therefore $\mathcal{R}_{b_0} \notin (\mathrm{EX}_0)_\tau(\lambda x.S(b_0, x))$, which contradicts the choice of S. Hence $B_1 \cup B_2 \notin \mathrm{suit}_\tau(J^1, \mathrm{EX}_0)$. □

For other learning classes in $\mathcal{I}$ similar negative results are obtained with recursive cores consisting of one or two elements.

Example 3. Consider the description sets $C = \{b \in \mathbb{N} \mid \mathcal{R}_b = \{\varphi_0^b\}\}$, as well as $B_1 = \{b \in \mathbb{N} \mid \mathrm{card}\,\mathcal{R}_b = \mathrm{card}\{i \mid \varphi_i^b \in \mathcal{R}\} = 2 \ \wedge\ \forall^\infty j\ [\varphi_j^b = \uparrow^\infty]\}$ and $B_2 = \{b \in \mathbb{N} \mid \mathrm{card}\{i \mid \varphi_i^b \in \mathcal{R}\} = 1 \wedge \forall^\infty j\ [\varphi_j^b = \uparrow^\infty]\}$. Then $C \in \mathrm{suit}_\varphi(J^1, \mathrm{EX}_0)$, $B_1 \in \mathrm{suit}_\varphi(J^2, \mathrm{CONS})$, $B_2 \in \mathrm{suit}_\varphi(J^1, \mathrm{CONS})$, but $C \cup B_1 \notin \mathrm{suit}(J^2, \mathrm{EX})$ and $C \cup B_2 \notin \mathrm{suit}_\tau(J^1, \mathrm{EX})$.

Example 4. Define two sets $B_1 := \{b \in \mathbb{N} \mid \{i \mid \varphi_i^b \text{ not initial}\} = \{0\}\}$ and $B_2 := \{b \in \mathbb{N} \mid \exists k\ [\{i \mid \varphi_i^b \text{ not initial}\} = \{k+1\} \ \wedge\ \forall y > k+1\ [\varphi_y^b = \uparrow^\infty]]\}$. Then

$B_1 \in \text{suit}_\varphi(J^1, \text{EX}_0)$, $B_2 \in \text{suit}_\varphi(J^1, \text{CONS})$, but $B_1 \cup B_2 \notin \text{suit}_\tau(J^1, \text{BC})$ and $B_1 \cup B_2 \notin \text{suit}_\varphi(J^1, \text{BC}^*)$.

The proofs use ideas similar to those in the proof of Example 2. As the reasons for the failure of all strategies for the unions in these examples are not obvious, it might be helpful first to collect some simple properties of description sets enabling the learnability of the corresponding unions.

Theorem 1. *Let $I \in \mathcal{I}$ and $B_1, B_2 \in \text{suit}(I, I)$ (or $\text{suit}_\tau(I, I)$, $\text{suit}_\varphi(I, I)$ resp.).*

1. *If B_1 is recursive, then $B_1 \cup B_2 \in \text{suit}(I, I)$ ($\text{suit}_\tau(I, I)$, $\text{suit}_\varphi(I, I)$ resp.).*
2. *If B_1, B_2 are r. e., then $B_1 \cup B_2 \in \text{suit}(I, I)$ ($\text{suit}_\tau(I, I)$, $\text{suit}_\varphi(I, I)$ resp.).*
3. *If $I \in \{\text{EX}, \text{BC}, \text{BC}^*\}$ and at least one of the sets $B_1, \mathbb{N} \setminus B_2$ is r. e., then $B_1 \cup B_2 \in \text{suit}(I, I)$ ($\text{suit}_\tau(I, I)$, $\text{suit}_\varphi(I, I)$ resp.).*

Proof (sketch). Fix $I \in \mathcal{I}$, $B_1, B_2 \in \text{suit}(I, I)$. For suit_τ, suit_φ all proofs proceed analogously. Choose $T_1, T_2 \in \mathcal{P}^2$ such that $B_i \in \text{suit}(I, I)(T_i)$ for $i \in \{1, 2\}$.

For the proof of the first claim let a new learner on input $(b, f[n])$ return $T_1(b, f[n])$ if $b \in B_1$, $T_2(b, f[n])$ otherwise. The second claim is shown with a strategy searching for the given b in an enumeration of B_1 and in parallel in an enumeration of B_2. As soon as $b \in B_i$ is verified for some $i \in \{1, 2\}$, the learner simulates T_i. For the third claim, let the new learner simulate T_2 until the given description is proved to belong to B_1 (or $\mathbb{N} \setminus B_2$), then simulate T_1. □

The second and third claim in Theorem 1 already suggest, that for some learning classes in $\mathcal{I}$ just one r. e. description set is not sufficient for the uniform identifiability of the union of two suitable description sets. Indeed this is true, as Theorem 4 will prove, but first one more special case in which any union of a suitable set with a suitable r. e. set again yields uniform learnability is regarded.

Theorem 2. *Assume $B_1 \in \text{suit}(\text{CONS}, \text{CONS})$ and $B_2 \in \text{suit}_\tau(\text{CONS}, \text{CONS})$. If B_1 (or $\mathbb{N} \setminus B_2$) is r. e., then $B_1 \cup B_2 \in \text{suit}(\text{CONS}, \text{CONS})$. If B_1 (or $\mathbb{N} \setminus B_2$) is r. e. and $B_1 \in \text{suit}_\tau(\text{CONS}, \text{CONS})$, then also $B_1 \cup B_2 \in \text{suit}_\tau(\text{CONS}, \text{CONS})$.*

Proof (sketch). Let $B_1 \in \text{suit}(\text{CONS}, \text{CONS})$, $B_2 \in \text{suit}_\tau(\text{CONS}, \text{CONS})$, moreover B_1 (or $\mathbb{N} \setminus B_2$) r. e. Then there exist $T_1, T_2 \in \mathcal{P}$ and numberings $\psi^{[b]}$ for $b \in B_1$ with $\mathcal{R}_b \in \text{CONS}_{\psi^{[b]}}(\lambda x.T_1(b, x))$ for $b \in B_1$ and $\mathcal{R}_b \in \text{CONS}_\tau(\lambda x.T_2(b, x))$ for $b \in B_2$. Define a new hypothesis space $\eta^{[b]} \in \mathcal{P}^2$ for each $b \in \mathbb{N}$ as follows.

$$\eta^{[b]}_{2i} = \psi^{[b]}_i \text{ if } b \in B_1,\ \eta^{[b]}_{2i} = \tau_i \text{ otherwise; } \eta^{[b]}_{2i+1} = \tau_i \text{ for all } i \in \mathbb{N}\ .$$

If $b \in B_2$, $f \in \mathcal{R}_b$, $n \in \mathbb{N}$, then $T_2(b, f[n])$ is consistent for $f[n]$ with respect to τ. So on input $(b, f[n])$ a meta-strategy for $B_1 \cup B_2$ can test by dove-tailed computation, whether $b \in B_1$ ($\mathbb{N} \setminus B_2$, resp.) or $T_2(b, f[n])$ is defined and consistent for $f[n]$ with respect to τ. As long as the latter property is verified first, the consistent index $2T_2(b, f[n]) + 1$ is returned; as soon as $b \in B_1$ ($b \notin B_2$, resp.) has once been verified, the hypotheses returned by $\lambda x.2T_1(b, x)$ are used. These are consistent with respect to τ, if $f \in \mathcal{R}_b$. Convergence follows from the choice of T_1, T_2. Proving the second claim does not require new hypothesis spaces. □

The following result gives an example of how weakening the learning intention can lead to positive results without further demands concerning the two description sets. The proof is omitted.

Theorem 3. *Let* $m, k_1, k_2 \in \mathbb{N}$, $m \le k_1$, $m \le k_2$. *If* $B_1 \in \text{suit}(\text{EX}_m, \text{EX}_{k_1})$ *and* $B_2 \in \text{suit}(\text{EX}_m, \text{EX}_{k_2})$, *then* $B_1 \cup B_2 \in \text{suit}(\text{EX}_m, \text{EX}_{k_1+k_2+1})$.

Thus the union of description sets suitable for learning with a bounded number of mind changes is still suitable, if a certain amount of further mind changes is permitted. But unfortunately, if an increasing mind change complexity is forbidden, negative results are obtained, sometimes even if one of the description sets is r. e. (in contrast to Theorems 1 and 2).

Theorem 4. *Let* $m \in \mathbb{N}$. *There exist two description sets* $B_1, B_2 \subseteq \mathbb{N}$ *such that*

1. $B_1, B_2 \in \text{suit}_\varphi(J^2, \text{EX}_m)$ *(or* $B_1, B_2 \in \text{suit}_\varphi(J^1, \text{EX}_m)$*),*
2. B_1 *is r. e., but*
3. $B_1 \cup B_2 \notin \text{suit}(J^2, \text{EX}_m)$ *(*$B_1 \cup B_2 \notin \text{suit}_\tau(J^1, \text{EX}_m)$*, resp.).*

Proof. Fix $m \in \mathbb{N}$. First B_1, B_2 are defined by constructing functions $\psi \in \mathcal{P}$ and $\text{fp} \in \mathcal{R}$ such that $\varphi^{\text{fp}(i)} = \psi^{(i,\text{fp}(i))}$ for all $i \in \mathbb{N}$. The function fp assigns to each integer i some fixed point value according to the recursion theorem. $B_1 \cup B_2$ will be the value set of fp. Below just the statements concerning descriptions of recursive cores in J^2 are verified, the proof for descriptions of singleton sets works with the same method combined with the ideas in the proof below Example 2.

Definition of B_1, B_2. By Proposition 1 there is a function $\text{rec} \in \mathcal{R}$, such that $\varphi^{\text{rec}(i)} \in \mathcal{R}$ and $\text{suit}(J^2, \text{EX}_m)(\varphi^i) \subseteq \text{suit}(J^2, \text{EX}_m)(\varphi^{\text{rec}(i)})$ for all $i \in \mathbb{N}$. From now on the notion S_i is used instead of $\varphi^{\text{rec}(i)}$. Next the function $\psi \in \mathcal{P}^4$ is defined by constructing a numbering $\psi^i \in \mathcal{P}^3$ for any integer i as follows.

Fix $b \in \mathbb{N}$. Go to stage 0.

Stage 0. $\psi_0^{(i,b)}(0) := 0$. For $x \in \mathbb{N}$ define

$$\psi_0^{(i,b)}(x+1) := \begin{cases} 0 & \text{if } S_i(b, \psi_0^{(i,b)}[y]) =? \text{ for } 0 \le y \le x \\ m+2 & \text{if } x \text{ is minimal, such that } S_i(b, \psi_0^{(i,b)}[x]) \in \mathbb{N} \\ \uparrow & \text{otherwise} \end{cases}$$

Then $\psi_0^{(i,b)} = 0^\infty$ iff $S_i(b, 0^n) =?$ for all $n \in \mathbb{N}$. Otherwise, if t is minimal with $S_i(b, 0^t) \in \mathbb{N}$, note $\psi_0^{(i,b)} = 0^t(m+2)\uparrow^\infty$. In the latter case let $\alpha_0^{(i,b)} := 0^t$ and go to stage 1. In the first case $\alpha_0^{(i,b)}\uparrow$ and stage 1 is not reached. *End stage 0.*

Stage k *(*$1 \le k \le m$*).* If l denotes the length of $\alpha_{k-1}^{(i,b)}$, let $\psi_{2k-1}^{(i,b)}[l-1] = \psi_{2k}^{(i,b)}[l-1] = \alpha_{k-1}^{(i,b)}$. For all $x \ge l$ define

$$\psi_{2k-1}^{(i,b)}(x) := \begin{cases} k-1 & \text{if } S_i(b, \alpha_{k-1}^{(i,b)} r^y) = S_i(b, \alpha_{k-1}^{(i,b)}) \\ & \text{for } 0 \le y \le x-l \text{ and for all } r \in \{k-1, k\} \\ m+2 & \text{if } x \text{ is minimal, such that } S_i(b, \alpha_{k-1}^{(i,b)} r^y) \ne S_i(b, \alpha_{k-1}^{(i,b)}) \\ & \text{for } y = x-l \text{ and some } r \in \{k-1, k\} \\ \uparrow & \text{otherwise} \end{cases}$$

$$\psi_{2k}^{(i,b)}(x) := \begin{cases} k & \text{if } \psi_{2k-1}^{(i,b)}(x) = k-1 \\ m+2 & \text{if } \psi_{2k-1}^{(i,b)}(x) = m+2 \\ \uparrow & \text{otherwise} \end{cases}$$

Note that $\psi_{2k-1}^{(i,b)} = \alpha_{k-1}^{(i,b)}(k-1)^\infty$ iff $\psi_{2k}^{(i,b)} = \alpha_{k-1}^{(i,b)} k^\infty$ iff

$$S_i(b, \alpha_{k-1}^{(i,b)}) = S_i(b, \alpha_{k-1}^{(i,b)}(k-1)^n) = S_i(b, \alpha_{k-1}^{(i,b)} k^n) \text{ for all } n \in \mathbb{N} \ .$$

Otherwise, if t is minimal, such that $S_i(b, \alpha_{k-1}^{(i,b)}) \neq S_i(b, \alpha_{k-1}^{(i,b)} r^t)$ for some $r \in \{k-1, k\}$, then $\psi_{2k-1}^{(i,b)} = \alpha_{k-1}^{(i,b)}(k-1)^t (m+2) \uparrow^\infty$ and $\psi_{2k}^{(i,b)} = \alpha_{k-1}^{(i,b)} k^t (m+2) \uparrow^\infty$. In the latter case let $\alpha_k^{(i,b)} := \alpha_{k-1}^{(i,b)} r^t$ and go to stage $k+1$. In the first case $\alpha_k^{(i,b)}$ is undefined and stage $k+1$ is not reached. *End stage k.*

Stage $m+1$. $\psi_{2m+1}^{(i,b)} = \alpha_m^{(i,b)} m^\infty$, $\psi_{2m+2}^{(i,b)} = \alpha_m^{(i,b)}(m+1)^\infty$, $\psi_x^{(i,b)} = \uparrow^\infty$ for $x > 2m+2$. *End stage $m+1$.*

Next let $g \in \mathcal{R}$ be a function satisfying $\varphi^{g(i,b)} = \psi^{(i,b)}$ for all $i, b \in \mathbb{N}$. By the recursion theorem there is a function $\mathrm{fp} \in \mathcal{R}$, such that $\varphi^{\mathrm{fp}(i)} = \varphi^{g(i,\mathrm{fp}(i))} = \psi^{(i,\mathrm{fp}(i))}$ for all $i \in \mathbb{N}$. Define $B := \{\mathrm{fp}(i) \mid i \in \mathbb{N}\}$ and finally let

$$B_1 := \{\mathrm{fp}(i) \mid i \in \mathbb{N} \ \wedge \ \exists x \in \mathbb{N}\, [\varphi_0^{\mathrm{fp}(i)}(x) = m+2]\}, \ B_2 := B \setminus B_1 \ .$$

End definition of B_1, B_2.

It remains to verify the three properties stated in Theorem 4.

Property 1. $B_1, B_2 \in \mathrm{suit}_\varphi(J^2, \mathrm{EX}_m)$.
$\mathrm{card}\,\mathcal{R}_b \leq 2$ for all $b \in B_1 \cup B_2$ follows by construction. The strategy constantly 0 shows $B_2 \in \mathrm{suit}_\varphi(J^2, \mathrm{EX}_m)$. Let a meta-learner for B_1 return "?" as long as the initial segment presented consists of zeros only. Afterwards the minimal consistent index in the numbering described is returned. This yields convergence to a correct program with no more than m mind changes on all relevant inputs.

Property 2. B_1 is r. e.
This property follows obviously by definition of B and B_1.

Property 3. $B_1 \cup B_2 \notin \mathrm{suit}(J^2, \mathrm{EX}_m)$.
Assume $B_1 \cup B_2 \in \mathrm{suit}(J^2, \mathrm{EX}_m)$. Then there exists some $i \in \mathbb{N}$, such that $B \in \mathrm{suit}(J^2, \mathrm{EX}_m)(S_i)$, i. e. for all $b \in B$ there is a numbering $\eta^{[b]} \in \mathcal{P}^2$ satisfying $\mathcal{R}_b \in (\mathrm{EX}_m)_{\eta^{[b]}}(\lambda x.S_i(b,x))$. Now let $b := \mathrm{fp}(i)$. The construction of ψ then implies $\mathcal{R}_b \notin (\mathrm{EX}_m)_{\eta^{[b]}}(\lambda x.S_i(b,x))$ (details are omitted) – a contradiction. □

Corollary 1. *Let $C := \{b \in \mathbb{N} \mid \mathcal{R}_b = \{\varphi_0^b\}\}$. Then $C \in \mathrm{suit}_\varphi(J^1, \mathrm{EX}_0)$ and*

1. *if $I \in \mathcal{I} \setminus \{\mathrm{BC}, \mathrm{BC}^*\}$, there is $B \in \mathrm{suit}_\varphi(J^2, I)$ such that $B \cup C \notin \mathrm{suit}(J^2, I)$,*
2. *if $I \in \mathcal{I} \setminus \{\mathrm{BC}^*\}$, there is $B \in \mathrm{suit}_\varphi(J^1, I)$ such that $B \cup C \notin \mathrm{suit}_\tau(J^1, I)$,*
3. *for all $I \in \mathcal{I}$ there is some $B \in \mathrm{suit}_\varphi(J^1, I)$ such that $B \cup C \notin \mathrm{suit}_\varphi(J^1, I)$.*

Proof. See Examples 2, 3, 4 and the proof of Theorem 4. □

Theorem 1 indicates that regarding the suitability of a union of two description sets it is very promising, if there is a method for separating the two sets algorithmically, that means if a learner can somehow determine which of the two sets a given description belongs to. Since in general such a separability is not attainable, it might be useful to state some properties of appropriate unions more carefully. To merge two meta-learners for two description sets into a successful meta-learner for the union set, it is not really required to find a way of separating the two sets themselves. It is already sufficient to recognize a description, *together* with some initial segment of a function to be learned, as some feasible input corresponding to one determinate description set of the two which are possible. This property can later on be used to prove the strong separations.

Proposition 2. *Let* $I \in \mathcal{I}$ *and* $B_1, B_2 \in \mathrm{suit}(I, I)$ *(or* $\mathrm{suit}_\tau(I, I)$*,* $\mathrm{suit}_\varphi(I, I)$ *resp.). Assume there is a function* $d \in \mathcal{P}_{01}$ *satisfying the following two conditions.*

1. $d(b, f(0)) = 1$ *for all descriptions* $b \in B_1$ *and all functions* $f \in \mathcal{R}_b$*,*
2. $d(b, f(0)) = 0$ *for all descriptions* $b \in B_2 \setminus B_1$ *and all functions* $f \in \mathcal{R}_b$*.*

Then $B_1 \cup B_2 \in \mathrm{suit}(I, I)$ *(*$\mathrm{suit}_\tau(I, I)$*,* $\mathrm{suit}_\varphi(I, I)$ *resp.).*[1]

Proof (sketch). Let $I \in \mathcal{I}$, $B_1, B_2 \in \mathrm{suit}(I, I)$ (the idea is the same for $\mathrm{suit}_\tau(I, I)$ and $\mathrm{suit}_\varphi(I, I)$). Fix two strategies $T_1, T_2 \in \mathcal{P}^2$, such that $B_1 \in \mathrm{suit}(I, I)(T_1)$ and $B_2 \in \mathrm{suit}(I, I)(T_2)$. On input $(b, f[n])$ a learner witnessing to $B_1 \cup B_2 \in \mathrm{suit}(I, I)$ returns $T_1(b, f[n])$ if $d(b, f(0)) = 1$, $T_2(b, f[n])$ if $d(b, f(0)) = 0$. □

Without special requirements concerning the intermediate hypotheses or the mind change complexity, the demands of Proposition 2 can be weakened. For EX- or BC-learning it is not necessary to recognize feasible inputs corresponding to a description set at once. It is enough to determine one of the two sets in the limit from gradually growing feasible information on a function to be learned.

Proposition 3. *Let* $I \in \{\mathrm{EX}, \mathrm{BC}, \mathrm{BC}^*\}$ *and* $B_1, B_2 \in \mathrm{suit}(I, I)$ *(or* $\mathrm{suit}_\tau(I, I)$*,* $\mathrm{suit}_\varphi(I, I)$ *resp.). Assume some function* $d \in \mathcal{P}_{01}$ *fulfils the following conditions.*

1. $\forall^\infty n\ [d(b, f[n]) = 1]$ *for all descriptions* $b \in B_1$ *and all functions* $f \in \mathcal{R}_b$*,*
2. $\forall^\infty n\ [d(b, f[n]) = 0]$ *for all descriptions* $b \in B_2 \setminus B_1$ *and all functions* $f \in \mathcal{R}_b$*.*

Then $B_1 \cup B_2 \in \mathrm{suit}(I, I)$ *(*$\mathrm{suit}_\tau(I, I)$*,* $\mathrm{suit}_\varphi(I, I)$ *resp.).*

For a proof a similar strategy as in the verification of Proposition 2 is adequate. Example 2 witnesses to the fact that Proposition 3 does not hold for $I = \mathrm{EX}_0$. Define a function $d \in \mathcal{P}_{01}$ on input $(b, f[n])$ to equal 1, if $f[n] = 0^{n+1}$, and to equal 0, if $f[n] \neq 0^{n+1}$. This yields the following properties.

- If $b \in B_1$ and $f \in \mathcal{R}_b$ (i.e. $f = 0^\infty$), then $d(b, f[n]) = 1$ for all $n \in \mathbb{N}$.
- If $b \in B_2 \setminus B_1$, $f \in \mathcal{R}_b$ (i.e. $f = 0^m 1^\infty$ for some $m \in \mathbb{N}$), then $d(b, f[n]) = 0$ for all but finitely many $n \in \mathbb{N}$.
- $B_1, B_2 \in \mathrm{suit}_\varphi(\mathrm{EX}_0, \mathrm{EX}_0)$.

If Proposition 3 was also true for the criterion $I = \mathrm{EX}_0$, then $B_1 \cup B_2 \in \mathrm{suit}_\varphi(\mathrm{EX}_0, \mathrm{EX}_0)$ would hold – a contradiction to Example 2.

[1] That the decision of the function d, for any $b \in B_1 \cup B_2$ and any $f \in \mathcal{R}_b$, does not depend on the values $f(n)$ for $n > 0$, may seem rather odd. Indeed a stronger version of Proposition 2 holds, yet for the purpose of this paper the actual version suffices.

5 Strong Separations of Criteria for Uniform Learning

In the uniform model now a strong separation of I and I' is expressed in terms of description sets as follows:

$$\forall B_1 \in \text{suit}(I,I)\ \exists B_2 \in \text{suit}(I,I')\ [B_1 \cup B_2 \in \text{suit}(I,I') \setminus \text{suit}(I,I)] \qquad (1)$$

or analogously with suit_τ or suit_φ. Fortunately such strong separations hold for any pair of criteria in which the weak separations have been verified. And it is possible to prove even more; the quantifiers in (1) can be switched. That means for any strongly separated pair (I, I') there is one fixed description set $B \in \text{suit}(I,I')$ such that for any set $B' \in \text{suit}(I,I)$ the union $B \cup B'$ belongs to $\text{suit}(I,I')$, but no longer to $\text{suit}(I,I)$. So there are special fixed description sets, which are in some sense typical for a set not suitable for uniform I-learning, but on the other hand restrictive enough to be suitable for uniform I'-identification in combination with any description set in $\text{suit}(I,I)$. The proof below Theorem 5 just gives an example for the pair (EX, BC), because other results of this style can be achieved by similar methods, even for the suit_τ- and suit_φ-models.

Theorem 5. *Let* (I, I') *be a pair in* $\{(\text{EX}_0, \text{CONS}), (\text{CONS}, \text{EX}), (\text{EX}, \text{BC})\} \cup \{(\text{EX}_m, \text{EX}_{m+1}) \mid m \in \mathbb{N}\}$. *Then there is a set* $B \subseteq \mathbb{N}$ *satisfying*

1. *B is recursively enumerable,*
2. $\text{card}\,\mathcal{R}_b \leq 2$ *for all* $b \in B$,
3. *if* $B' \in \text{suit}(I,I)$, *then* $B \cup B' \in \text{suit}(I,I') \setminus \text{suit}(I,I)$.

Moreover, this statement holds for suit_τ *and* suit_φ *instead of* suit, *even if the second condition is strengthened to* $\text{card}\,\mathcal{R}_b = 1$ *for all* $b \in B$.

Proof for $(I, I') = (\text{EX}, \text{BC})$. First the set B is defined via the construction of a partial-recursive function ψ and a recursive function fp such that $\varphi^{\text{fp}(i)} = \psi^{(i,\text{fp}(i))}$ for all $i \in \mathbb{N}$. The function fp assigns to each integer i some fixed point value according to the recursion theorem. B will be the value set of fp.

Definition of B. By Proposition 1 there is some rec $\in \mathcal{R}$, such that $\varphi^{\text{rec}(i)} \in \mathcal{R}$ and $\text{suit}(\text{EX},\text{EX})(\varphi^i) \subseteq \text{suit}(\text{EX},\text{EX})(\varphi^{\text{rec}(i)})$ for all $i \in \mathbb{N}$. From now on the notion S_i is used instead of $\varphi^{\text{rec}(i)}$. Next the function $\psi \in \mathcal{P}^4$ is defined by constructing a numbering $\psi^i \in \mathcal{P}^3$ for any integer i as follows.

Fix $b \in \mathbb{N}$. Let $\alpha_0^{(i,b)} := (i)$, $\psi_0^{(i,b)}(0) = i$ and go to stage 0. In general, the construction in stage k ($k \in \mathbb{N}$) proceeds in the following way:
Look for the minimal integer $m_k^{(i,b)} \in \mathbb{N}$ which fulfils

$$S_i(b, \alpha_k^{(i,b)} 0 0^{m_k^{(i,b)}}) \neq S_i(b, \alpha_k^{(i,b)}) \text{ or } S_i(b, \alpha_k^{(i,b)} 1 0^{m_k^{(i,b)}}) \neq S_i(b, \alpha_k^{(i,b)})\ .$$

If such an integer does not exist, $m_k^{(i,b)}$ is undefined. Then let

$$\alpha_{k+1}^{(i,b)} := \begin{cases} \alpha_k^{(i,b)} 0 0^{m_k^{(i,b)}} 2 & \text{if } m_k^{(i,b)} \downarrow \text{ and } S_i(b, \alpha_k^{(i,b)} 0 0^{m_k^{(i,b)}}) \neq S_i(b, \alpha_k^{(i,b)}) \\ \alpha_k^{(i,b)} 1 0^{m_k^{(i,b)}} 2 & \text{if } m_k^{(i,b)} \downarrow \text{ and } S_i(b, \alpha_k^{(i,b)} 0 0^{m_k^{(i,b)}}) = S_i(b, \alpha_k^{(i,b)}) \\ \uparrow & \text{if } m_k^{(i,b)} \uparrow \end{cases}$$

In addition $\alpha_{k+1}^{(i,b)}$ becomes an initial segment of $\psi_0^{(i,b)}$, i. e. $\psi_0^{(i,b)}[l-1] := \alpha_{k+1}^{(i,b)}$, if $\alpha_{k+1}^{(i,b)}$ is defined and of length l; $\psi_0^{(i,b)} := \alpha_k^{(i,b)}\uparrow^\infty$ otherwise. Furthermore define

$$\psi_{2k+1}^{(i,b)} := \begin{cases} \alpha_k^{(i,b)}00^\infty & \text{if } m_k^{(i,b)}\uparrow \\ \psi_0^{(i,b)} & \text{if } m_k^{(i,b)}\downarrow \text{ and } S_i(b,\alpha_k^{(i,b)}00^{m_k^{(i,b)}}) \neq S_i(b,\alpha_k^{(i,b)}) \\ \alpha_k^{(i,b)}00^{m_k^{(i,b)}}\uparrow^\infty & \text{if } m_k^{(i,b)}\downarrow \text{ and } S_i(b,\alpha_k^{(i,b)}00^{m_k^{(i,b)}}) = S_i(b,\alpha_k^{(i,b)}) \end{cases}$$

$$\psi_{2k+2}^{(i,b)} := \begin{cases} \alpha_k^{(i,b)}10^\infty & \text{if } m_k^{(i,b)}\uparrow \\ \psi_0^{(i,b)} & \text{if } m_k^{(i,b)}\downarrow \text{ and } S_i(b,\alpha_k^{(i,b)}00^{m_k^{(i,b)}}) = S_i(b,\alpha_k^{(i,b)}) \\ \alpha_k^{(i,b)}10^{m_k^{(i,b)}}\uparrow^\infty & \text{if } m_k^{(i,b)}\downarrow \text{ and } S_i(b,\alpha_k^{(i,b)}00^{m_k^{(i,b)}}) \neq S_i(b,\alpha_k^{(i,b)}) \end{cases}$$

If $m_k^{(i,b)}$ exists, then go to stage $k+1$. *End stage k.*

Let $g \in \mathcal{R}$ satisfy $\varphi^{g(i,b)} = \psi^{(i,b)}$ for all $i,b \in \mathbb{N}$. By the recursion theorem there is some fp $\in \mathcal{R}$, such that $\varphi^{\text{fp}(i)} = \varphi^{g(i,\text{fp}(i))} = \psi^{(i,\text{fp}(i))}$ for all $i \in \mathbb{N}$. Finally let $B := \{\text{fp}(i) \mid i \in \mathbb{N}\}$. *End definition of B.*

Claim. Let $b \in B$ and $i \in \mathbb{N}$ such that $\text{fp}(i) = b$. Then

1. $\mathcal{R}_b = \{\varphi_0^b\} \iff m_k^{(i,b)}$ is defined for all $k \in \mathbb{N} \iff$ for all $k \in \mathbb{N}$ some $t_k \in \{0,1\}$ fulfils $\alpha_k^{(i,b)}t_k \subseteq \varphi_0^b$ and $\{\varphi_{2k+1}^b, \varphi_{2k+2}^b\} = \{\varphi_0^b, \alpha_k^{(i,b)}\overline{t_k}0^{m_k^{(i,b)}}\uparrow^\infty\}$.
2. $\mathcal{R}_b = \{\varphi_{2k+1}^b, \varphi_{2k+2}^b\} = \{\alpha_k^{(i,b)}00^\infty, \alpha_k^{(i,b)}10^\infty\} \iff k$ is the minimal integer such that $m_k^{(i,b)}$ is not defined,
3. if $\alpha_k^{(i,b)}$ is defined for some $k \in \mathbb{N}$, then $\text{card}\{r \geq 1 \mid \alpha_k^{(i,b)}(r) = 2\} = k$.

These claims follow immediately from the construction; hence it remains to verify the properties 1 up to 3 stated in the Theorem.

Property 1. B is recursively enumerable.
B is the value set of a recursive function and thus r. e.

Property 2. $\text{card}\,\mathcal{R}_b \leq 2$ for all $b \in B$.
By Claims 1 and 2 either $\mathcal{R}_b = \{\varphi_0^b\}$ or $\mathcal{R}_b = \{\varphi_{2k+1}^b, \varphi_{2k+2}^b\}$ for some $k \in \mathbb{N}$.

Property 3. If $B' \in \text{suit}(\text{EX},\text{EX})$, then $B \cup B' \in \text{suit}(\text{EX},\text{BC}) \setminus \text{suit}(\text{EX},\text{EX})$.
Fix some description set $B' \in \text{suit}(\text{EX},\text{EX})$. First $B \cup B' \notin \text{suit}(\text{EX},\text{EX})$ will be verified, afterwards the proof of $B \cup B' \in \text{suit}(\text{EX},\text{BC})$ follows.

Proof of $B \cup B' \notin \text{suit}(\text{EX},\text{EX})$. Assume that $B \cup B' \in \text{suit}(\text{EX},\text{EX})$. Then in particular there exists some total recursive uniform strategy appropriate for EX-identification of the set B; that means, there is some $i \in \mathbb{N}$ such that

$$\text{for all } b \in B \text{ some } \eta^{[b]} \in \mathcal{P}^2 \text{ fulfils } \mathcal{R}_b \in \text{EX}_{\eta^{[b]}}(\lambda x.S_i(b,x)) \ . \qquad (2)$$

Let $b := \text{fp}(i)$, i. e. $\varphi^b = \psi^{(i,b)}$. By Claims 1 and 2 it suffices to consider two cases.

Case (i). $\mathcal{R}_b = \{\varphi_0^b\}$. Then for all $k \in \mathbb{N}$, by Claim 1, $m_k^{(i,b)}$ is defined and the construction yields $\alpha_k^{(i,b)}t_k0^{m_k^{(i,b)}} \subseteq \varphi_0^b$ and $S_i(b,\alpha_k^{(i,b)}) \neq S_i(b,\alpha_k^{(i,b)}t_k0^{m_k^{(i,b)}})$ for some $t_k \in \{0,1\}$. Thus the sequence of hypotheses returned by $\lambda x.S_i(b,x)$ on φ_0^b diverges, which implies $\mathcal{R}_b \notin \mathrm{EX}_{\eta^{[b]}}(\lambda x.S_i(b,x))$ in contradiction to (2).

Case (ii). $\mathcal{R}_b = \{\varphi^b_{2k+1}, \varphi^b_{2k+2}\} = \{\alpha_k^{(i,b)}00^\infty, \alpha_k^{(i,b)}10^\infty\}$ for some $k \in \mathbb{N}$. Then by Claim 2 the value $m_k^{(i,b)}$ is not defined. This implies

$$S_i(b, \alpha_k^{(i,b)}00^m) = S_i(b, \alpha_k^{(i,b)}) = S_i(b, \alpha_k^{(i,b)}10^m) \text{ for all } m \in \mathbb{N} \ .$$

Thus $S_i(b, \varphi^b_{2k+1}[n]) = S_i(b, \varphi^b_{2k+2}[n])$ for all but finitely many $n \in \mathbb{N}$, although $\varphi^b_{2k+1} \neq \varphi^b_{2k+2}$. Therefore – without influence of the hypothesis space regarded – the sequence of hypotheses returned by $\lambda x.S_i(b,x)$ must converge to an incorrect program for at least one of the two functions in $\mathcal{R}_b$. Hence $\mathcal{R}_b \notin \mathrm{EX}_{\eta^{[b]}}(\lambda x.S_i(b,x))$; again a contradiction to (2).

Both cases lead to a contradiction, so $B \cup B' \notin \mathrm{suit}(\mathrm{EX}, \mathrm{EX})$ is obtained.

Proof of $B \cup B' \in \mathrm{suit}(\mathrm{EX}, \mathrm{BC})$. By definition $B' \in \mathrm{suit}(\mathrm{EX}, \mathrm{BC})$. It is possible to show that $B \in \mathrm{suit}(\mathrm{EX}, \mathrm{BC})$ and afterwards apply Proposition 2. First a strategy $T \in \mathcal{R}$ satisfying $\mathcal{R}_b \in \mathrm{BC}_{\varphi^b}(T)$ for all $b \in B$ is defined as follows: let $f \in \mathcal{R}$. Then $T(f[0]) := 0$ and for any $n > 0$ with $f(n) = 2$, let $T(f[n]) := T(f[n-1])$. If $n > 0$ and $f(n) \neq 2$, let $k := \mathrm{card}\{r \mid 1 \leq r \leq n \wedge f(r) = 2\}$ and define $T(f[n]) := 2k+1$ in case $f[n] = \alpha 20^x$ for some segment α and some $x > 0$. Otherwise $T(f[n]) := 2k+2$. $\mathcal{R}_b \in \mathrm{BC}_{\varphi^b}(T)$ for all $b \in B$ can be verified with the help of Claims 1–3. Details are omitted here. So $B \in \mathrm{suit}(\mathrm{EX}, \mathrm{BC})$.

In order to apply Proposition 2 it is necessary to find some $d \in \mathcal{P}_{01}$ satisfying:

1. $d(b, f(0)) = 1$ for all descriptions $b \in B$ and all functions $f \in \mathcal{R}_b$,
2. $d(b, f(0)) = 0$ for all descriptions $b \in B' \setminus B$ and all functions $f \in \mathcal{R}_b$.

For that purpose let $d(b,n) := 1$, if $\mathrm{fp}(n) = b$ and $d(b,n) := 0$, if $\mathrm{fp}(n) \neq b$ ($b, n \in \mathbb{N}$). If $b \in B$, $f \in \mathcal{R}_b$, the construction of ψ implies $\mathrm{fp}(f(0)) = b$, so $d(b, f(0)) = 1$. If $b \notin B = \{\mathrm{fp}(i) \mid i \in \mathbb{N}\}$, then $\mathrm{fp}(n) \neq b$ is obtained for any $n \in \mathbb{N}$; in particular $d(b,n) = 0$. So $B \cup B' \in \mathrm{suit}(\mathrm{EX}, \mathrm{BC})$ by Proposition 2.[2] □

According to Theorem 5 it might in some cases be reasonable to give up certain constraints concerning the inference criterion I, because thus an increase of learning power can be achieved, no matter what description set in $\mathrm{suit}(I,I)$ has to be learned. The proof (together with the proof of Proposition 2) indicates, *how* to modify a uniform I-learner for an arbitrary set in $\mathrm{suit}(I,I)$ into an adequate $\mathrm{suit}(I,I')$-learner for a proper superset, which does no longer belong to $\mathrm{suit}(I,I)$. Hence this result does not only give the advice to sometimes loosen the learning criteria, but furthermore provides a method for designing more powerful learners. The fact that there is a fixed description set, which witnesses to the strong separation for *any* description set in $\mathrm{suit}(I,I)$, indicates that there exists some kind of structure for a somehow characteristic description set corresponding to the class $\mathrm{suit}(I,I') \setminus \mathrm{suit}(I,I)$. The structure of such a characteristic set is on the one hand complex enough to disallow uniform I-learning, but on the other hand

[2] Actually Proposition 2 is not needed: since B is r.e., Theorem 1.3 suffices. But in order to provide also the idea for the omitted parts, a universal method is chosen.

simple enough to enable uniform I'-identification of its union with any arbitrary description set in $\mathrm{suit}(I, I)$.

Theorem 6. *There is a description set $B \subseteq \mathbb{N}$ satisfying*

1. *B is recursively enumerable,*
2. $\mathrm{card}\,\mathcal{R}_b = 1$ *for all $b \in B$,*
3. *if $B' \in \mathrm{suit}_\tau(\mathrm{BC}, \mathrm{BC})$, then $B \cup B' \in \mathrm{suit}_\tau(\mathrm{BC}, \mathrm{BC}^*) \setminus \mathrm{suit}_\tau(\mathrm{BC}, \mathrm{BC})$ (or analogously for suit_φ instead of suit_τ).*

The proof is omitted. Theorem 6 is formulated separately from Theorem 5, because its statement does not hold for the uniform learning model given by $\mathrm{suit}(\mathrm{BC}, \mathrm{BC})$ (i. e. without specification of the hypothesis spaces). The reason is that $\mathrm{suit}(\mathrm{BC}, \mathrm{BC}) = \{B \subseteq \mathbb{N} \mid \mathcal{R}_b \in \mathrm{BC} \text{ for all } b \in B\}$, which implies $\mathrm{suit}(\mathrm{BC}, \mathrm{BC}^*) = \mathrm{suit}(\mathrm{BC}, \mathrm{BC})$ (the proof uses a simple construction of appropriate hypothesis spaces). Note that Theorems 5 and 6 yield strong separations for all pairs (I, I') of learning classes from $\mathcal{I}$, for which $I \subset I'$ holds.

References

1. Baliga, G.; Case, J.; Jain, S.; *The Synthesis of Language Learners*, Information and Computation 152, 16–43 (1999).
2. Barzdin, J.; *Two Theorems on the Limiting Synthesis of Functions*, Theory of Algorithms and Programs, Latvian State University, Riga 210, 82–88 (1974) (in Russian).
3. Barzdin, J.; *Inductive Inference of Automata, Functions and Programs*, in Proc. International Congress of Math., Vancouver, 455–460 (1974).
4. Blum, M.; *A Machine-Independent Theory of the Complexity of Recursive Functions*, Journal of the ACM 14 (2), 322–336 (1967).
5. Case, J.; Chen, K.; Jain, S.; *Strong Separation of Learning Classes*, Journal of Experimental and Theoretical Artificial Intelligence 4, 281–293 (1992).
6. Case, J.; Smith, C.; *Comparison of Identification Criteria for Machine Inductive Inference*, Theoretical Computer Science 25, 193–220 (1983).
7. Freivalds, R.; Kinber, E. B.; Wiehagen, R. (1995); *How Inductive Inference Strategies Discover Their Errors*, Information and Computation 118, 208–226.
8. Gold, E. M.; *Language Identification in the Limit*, Information and Control 10, 447–474 (1967).
9. Jantke, K. P.; *Natural Properties of Strategies Identifying Recursive Functions*, Elektronische Informationsverarbeitung und Kybernetik 15, 487–496 (1979).
10. Kapur, S.; Bilardi, G.; *On uniform learnability of language families*, Information Processing Letters 44, 35–38 (1992).
11. Osherson, D. N.; Stob, M.; Weinstein, S.; *Synthesizing Inductive Expertise*, Information and Computation 77, 138–161 (1988).
12. Rogers, H.; *Theory of Recursive Functions and Effective Computability*, MIT Press, Cambridge, Massachusetts (1987).
13. Wiehagen, R.; Zeugmann, T.; *Learning and Consistency*, in K. P. Jantke and S. Lange (Eds.): Algorithmic Learning for Knowledge-Based Systems, LNAI 961, 1–24, Springer-Verlag (1995).
14. Zilles, S.; *On the Comparison of Inductive Inference Criteria for Uniform Learning of Finite Classes*, in N. Abe, R. Khardon, and T. Zeugmann (Eds.): ALT 2001, LNAI 2225, 251–266, Springer-Verlag (2001).