Towards a Map for Incremental Learning in the Limit from Positive and Negative Information

Ardalan Khazraei[1], Timo Kötzing[2], and Karen Seidel[2]

[1] University of Potsdam, Germany
[2] Hasso-Plattner-Institute, Potsdam, Germany

Abstract. In order to model an efficient learning paradigm, iterative learning algorithms access data one by one, updating the current hypothesis without regress to past data. Prior research investigating the impact of additional requirements on iterative learners left many questions open, especially in learning from informant, where the input is binary labeled.
We first compare learning from positive information (text) with learning from informant. We provide different concept classes learnable from text but not by an iterative learner from informant. Further, we show that totality restricts iterative learning from informant.
Towards a map of iterative learning from informant, we prove that strongly non-U-shaped learning is restrictive and that iterative learners from informant can be assumed canny for a wide range of learning criteria. Finally, we compare two syntactic learning requirements.

Keywords: Learning in the Limit, Map for Iterative Learners from Informant, (Strongly) Non-U-Shaped Learning

1 Introduction

We are interested in the problem of algorithmically learning a description for a formal language (a computably enumerable subset of the set of natural numbers) when presented successively all information about that language; this is sometimes called *inductive inference*, a branch of (algorithmic) learning theory.
Many criteria for deciding whether a learner M is *successful* on a language L have been proposed in the literature. Gold, in his seminal paper [Gol67], gave a first, simple learning criterion, **Ex**-*learning*[3], where a learner is *successful* iff, on every complete information about L it eventually stops changing its conjectures, and its final conjecture is a correct description for the input sequence. Trivially, each single, describable language L has a suitable constant function as a **Ex**-learner (this learner constantly outputs a description for L). As we want algorithms for more than a single learning task, we are interested in analyzing for which *classes of languages* $\mathcal{L}$ there is a *single learner* M learning *each* member of $\mathcal{L}$. This framework is also sometimes known as *language learning in*

[3] **Ex** stands for *explanatory*.

the limit and has been studied using a wide range of learning criteria in the flavor of **Ex**-learning (see, for example, the textbook [JORS99]).
One major criticism of the model suggested by Gold, see for example [CM08], is its excessive use of memory: for each new hypothesis the entire history of past data is available. Iterative learning [Wie76], is the most common variant of learning in the limit which addresses memory constraints: the memory of the learner on past data is just its current hypothesis. Due to the padding lemma, this memory is still not void, but finitely many data can be memorized in the hypothesis.
Prior work on iterative learning [CK10,CM08,JKMS16,JMZ13,JORS99] focused on learning from text, that is, from positive data only. Hence, in **TxtEx**-learning the complete information is a listing of all and only the elements of L. In this paper we are mainly interested in the paradigm of learning from both positive and negative information. For example, when learning half-spaces, one could see data declaring that $\langle 1,1\rangle$ is in the target half-space, further is $\langle 3,2\rangle$, but $\langle 1,7\rangle$ is not, and so on. This setting is called *learning from informant* (in contrast to learning from *text*).

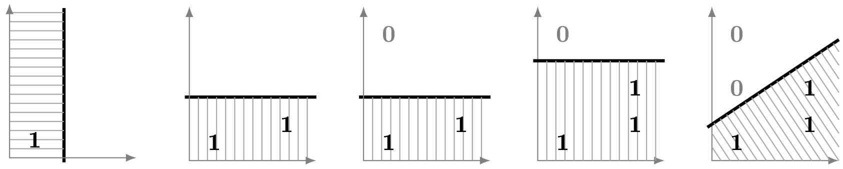

Fig. 1. Example Learning Process with binary labeled data and half-spaces as hypotheses.

Iterative learning from informant was analyzed by [JLZ07], where various natural restrictions have been considered and the authors focused on the case of learning indexable families (classes of languages which are uniformly decidable). In this paper we are looking at other established restrictions and also consider learning of arbitrary classes of computably enumerable languages.
In Section 3 we consider the two aforementioned restrictions on learning from informant: learning from text and learning iteratively. Both restrictions render fewer classes of languages learnable; in fact, the two restrictions yield two incomparable sets of language classes being learnable, which also shows that learning iteratively from text is weaker than supposing just one of the two restrictions.
Towards a better understanding of iterative learners we analyze which normal forms can be assumed in Section 4. First we show that, analogously to the case of learning from text (as analyzed in [CM09]), we cannot assume learners to be total (i.e. always giving an output).
However, from [CM08] we know that we can assume iterative text learners to be *canny* (also defined in Section 4); we adapt this normal form for the case of iterative learning from informant and show that it can be assumed to hold for iterative learners generally.
Many works in inductive inference, see for example [JKMS16], [KP16], [KS16], [KSS17], focus on relating different additional learning requirements for a fixed learning model. In particular, [JKMS16] mapped out

all pairwise relations for an established choice of learning restrictions for iterative learning from text. The complete map of all pairwise relations between for full-information learners from informant can be found in [AKS18]. A similar map for the case of iterative learning from informant is not known. Canniness is central in investigating the learning power of iterative learning from texts. Hence, it is an important stepping stone to understand iterative learners better and determine such pairwise relations. We argue in Lemma 3 that the normal form of canniness still can be assumed in case we pose additional semantic learning requirements.
In Section 5 we collect all previously known results for such a map, see [LZ92], [JLZ07]. We observe that it decreases learning power to require the learner to never change its hypothesis, once it is correct. The proof for separating this notion, called strong non-U-shapedness, relies on the ORT recursion theorem [Cas74]. We close this section by comparing two syntactic learning requirements for iterative learners from informant that proved important to derive the equivalence of all syntactic requirements for iterative learners from text.
We continue this paper with some mathematical preliminaries in Section 2 before discussing our results in more detail.

2 Iterative Learning from Informant

Notation and terminology on the learning theoretic side follow [OSW86], [JORS99] and [LZZ08], whereas on the computability theoretic side we refer to [Odi99] and [Rog67]. For both we also recommend [Köt09].
A *language* L is a recursively enumerable subset of $\mathbb{N}$. We denote the characteristic function for $L \subseteq \mathbb{N}$ by $f_L : \mathbb{N} \to \{0,1\}$.
Gold in his seminal paper [Gol67], distinguished two major different kinds of information presentation. A function $I : \mathbb{N} \to \mathbb{N} \times \{0,1\}$ is an *informant for language* L, if there is a surjection $n : \mathbb{N} \to \mathbb{N}$ such that $I(t) = (n(t), f_L(n(t)))$ holds for every $t \in \mathbb{N}$. Moreover, for an informant I let

$$\mathrm{pos}(I) := \{y \in \mathbb{N} \mid \exists x \in \mathbb{N}\colon \mathrm{pr}_1(I(x)) = y \wedge \mathrm{pr}_2(I(x)) = 1\} \text{ and}$$
$$\mathrm{neg}(I) := \{y \in \mathbb{N} \mid \exists x \in \mathbb{N}\colon \mathrm{pr}_1(I(x)) = y \wedge \mathrm{pr}_2(I(x)) = 0\}$$

denote the sets of all natural numbers, about which I gives some positive or negative information, respectively. A *text for language* L is a function $T : \mathbb{N} \to \mathbb{N} \cup \{\#\}$ with range L after removing #. The symbol # is interpreted as pause symbol.
Therefore, when learning from informant, the *set of admissible inputs to the learning algorithm* $\mathbb{S}$ is the set of all finite sequences

$$\sigma = ((n_0, y_0), \ldots, (n_{|\sigma|-1}, y_{|\sigma|-1}))$$

of *consistently* binary labeled natural numbers. When learning from text (positive data only), we encounter inputs to the learning algorithm from the set $\mathbb{T}$ of finite sequences $\tau = (n_0, \ldots, n_{|\tau|-1})$ of natural numbers and the pause symbol #. The initial subsequence relation is denoted by $\sqsubseteq$.
A set $\mathcal{L} = \{L_i \mid i \in \mathbb{N}\}$ of languages is called *indexable family* if there is a computer program that on input $(i, n) \in \mathbb{N}^2$ returns 1 if $n \in L_i$ and 0

otherwise. Examples are **Fin** and **CoFin**, the set of all finite subsets of $\mathbb{N}$ and the set of all complements of finite subsets of $\mathbb{N}$, respectively.
Let $\mathcal{L}$ be a collection of languages we seek a provably correct learning algorithm for. We will refer to $\mathcal{L}$ as the *concept class* which will often be an indexable family. Further, let $\mathcal{H} = \{L_i \mid i \in \mathbb{N}\}$ with $\mathcal{L} \subseteq \mathcal{H}$ be a collection of languages called the *hypothesis space*. In general we do *not* assume that for every $L \in \mathcal{L}$ there is a unique index $i \in \mathbb{N}$ with $L_i = L$. Indeed, ambiguity in the hypothesis space helps memory-resticted learners to remember data.
A *learner M from informant (text)* is a computable function

$$M : \mathbb{S} \to \mathbb{N} \cup \{?\} \qquad (M : \mathbb{T} \to \mathbb{N} \cup \{?\})$$

with the output i interpreted with respect to $\mathcal{H} = \{L_i \mid i \in \mathbb{N}\}$, a prefixed hypothesis space. The output ? often serves as initial hypothesis or is interpreted as no new hypothesis. Often $\mathcal{H}$ is an indexable class or the established W-hypothesis space defined in Subsection 4.
Let I be an informant (T be a text) for L and $\mathcal{H} = \{L_i \mid i \in \mathbb{N}\}$ a hypothesis space. A learner $M : \mathbb{S} \to \mathbb{N} \cup \{?\}$ ($M : \mathbb{T} \to \mathbb{N} \cup \{?\}$) is *successful on I (on T)* if it eventually settles on $i \in \mathbb{N}$ with $L_i = L$. This means that when receiving increasingly long finite initial segments of I (of T) as inputs, it will from some time on be correct and not change the output on longer initial segments of I (of T). *M learns L wrt $\mathcal{H}$* if it is successful on every informant I (on every text T) for L. *M learns $\mathcal{L}$* if there is a hypothesis space $\mathcal{H}$ such that M learns every $L \in \mathcal{L}$ wrt $\mathcal{H}$. We denote the collection of all $\mathcal{L}$ learnable from informant (text) by **[InfEx]** (**[TxtEx]**). If we fix the hypothesis space, we denote this by a subscript for **Ex**.
According to [Wie76], [LZ96], [CJLZ99] a learner M is *iterative* if its output on $\sigma \in \mathbb{S}$ ($\tau \in \mathbb{T}$) only depends on the last input $\mathrm{last}(\sigma)$ and the hypothesis $M(\sigma^-)$ after observing σ without its last element $\mathrm{last}(\sigma)$. The collection of all concept classes $\mathcal{L}$ learnable by an iterative learner from informant (text) is denoted by **[ItInfEx]** (**[ItTxtEx]**).
The s-m-n theorem gives finite and infinite recursion theorems, see [Cas94], [Odi99]. We will refer to Case's Operator Recursion Theorem ORT in its 1-1-form, see [Cas74], [JORS99], [Köt09].

3 Comparison with Learning from Text

By ignoring negative information every informant incorporates a text for the language presented and we gain **[ItTxtEx]** $\subseteq$ **[ItInfEx]**.
It has been observed in [OSW86] that the superfinite language class **Fin** $\cup \{\mathbb{N}\}$ is in **[InfEx]**$\setminus$**[ItInfEx]**. With $L_k = 2\mathbb{N} \cup \{2k+1\}$ and $L'_k = L_k \setminus \{2k\}$ the indexable family $\mathcal{L} = \{2\mathbb{N}\} \cup \{L_k, L'_k \mid k \in \mathbb{N}\}$ lies in **[TxtEx]**$\cap$**[ItInfEx]** but not in **[ItTxtEx]**. In [JORS99] the separations are witnessed by the indexable family $\{\mathbb{N}\setminus\{0\}\} \cup \{D \cup \{0\} : D \in$ **Fin**$\}$.
It can easily be verified that **CoFin** $\in$ **[ItInfEx]**$\setminus$**[TxtEx]** and with the next result **[ItInfEx]** and **[TxtEx]** are incomparable by inclusion.

Lemma 1. *There is an indexable family in* **[TxtEx]**$\setminus$**[ItInfEx]**.

Summing up, we know $[\mathbf{ItTxtEx}] \subsetneq [\mathbf{TxtEx}] \perp [\mathbf{ItInfEx}] \subsetneq [\mathbf{InfEx}]$, where $\perp$ stands for incomparability with respect to set inclusion, meaning (1) there is a concept class learnable from text but not by an iterative learner from informant and (2) there is a concept class learnable by an iterative learner from informant but not from text.

Moreover, with a Boolean function we can show that every concept class separating $[\mathbf{ItInfEx}]$ and $[\mathbf{InfEx}]$ yields a separating class for $[\mathbf{ItInfEx}]$ and $[\mathbf{TxtEx}]$. We generalize this further in the full version.

4 Total and Canny Learners

For the rest of the paper, without further notation, all results are understood with respect to the W-hypothesis space defined in the following. We fix a programming system φ as introduced in [RC94]. Briefly, in the φ-system, for a natural number p, we denote by φ_p the partial computable function with program code p. We also call p an *index* for W_p defined as $\mathrm{dom}(\varphi_p)$.
We show that totality, denoted by $\mathcal{R}$, restricts iterative learning from informant.The proof uses an easy ORT argument.

Theorem 1. $[\mathbf{ItInfEx}]\backslash[\mathcal{R}\mathbf{ItInfEx}] \neq \varnothing$.

Proof. Let o be an index for $\varnothing$ and define the iterative learner M for all $\xi \in \mathbb{N} \times \{0, 1\}$ by

$$\begin{aligned} M(\varnothing) &= o; \\ h_M(h, \xi) &= \begin{cases} \varphi_{\mathrm{pr}_1(\xi)}(0), & \text{else if } \mathrm{pr}_2(\xi) = 1 \text{ and } h \notin \mathrm{ran}(\mathrm{ind}); \\ h, & \text{otherwise.} \end{cases} \end{aligned}$$

We argue that $\mathcal{L} := \{ L \subseteq \mathbb{N} \mid L \in \mathbf{ItInfEx}(M) \}$ is not learnable by a total learner from informants. Assume towards a contradiction M' is such a learner. For a finite informant sequence σ we denote by $\overline{\sigma}$ the corresponding canonical finite informant sequence, ending with σ's datum with highest first coordinate. Then by padded ORT there are $e \in \mathbb{N}$ and a strictly increasing computable function $a : \mathbb{N}^{<\omega} \to \mathbb{N}$, such that for all $\sigma \in \mathbb{N}^{<\omega}$ and all $i \in \mathbb{N}$

$$\begin{aligned} \sigma_0 &= \varnothing; \\ \sigma_{i+1} &= \sigma_i{}^\frown \begin{cases} (a(\sigma_i), 1), & \text{if } M'(\overline{\sigma_i{}^\frown (a(\sigma_i), 1)}) \neq M'(\overline{\sigma_i}); \\ \varnothing, & \text{otherwise;} \end{cases} \\ W_e &= \bigcup_{i \in \mathbb{N}} \mathrm{pos}(\overline{\sigma_i}); \\ \varphi_{a(\sigma)}(x) &= \begin{cases} e, & \text{if } M'(\overline{\sigma^\frown (a(\sigma), 1)}) \neq M'(\overline{\sigma}); \\ \mathrm{ind}_{\mathrm{pos}(\sigma) \cup \{a(\sigma)\}}, & \text{otherwise;} \end{cases} \end{aligned} \tag{1}$$

Clearly, we have $W_e \in \mathcal{L}$ and thus M' also $\mathbf{InfEx}$-learns W_e. By the $\mathbf{Ex}$-convergence there are $e', t_0 \in \mathbb{N}$, where t_0 is minimal, such that $W_{e'} = W_e$

and for all $t \geqslant t_0$ we have $M'(\bigcup_{i\in\mathbb{N}} \overline{\sigma_i}[t]) = e'$ and hence by (1) for all i with $|\overline{\sigma_i}| \geqslant t_0$

$$M'(\overline{\sigma_i{}^\frown(a(\sigma_i),1)}) = M'(\overline{\sigma_i}) = M'(\overline{\sigma_i{}^\frown(a(\sigma_i),0)}).$$

It is easy to see, that $W_e = \mathrm{pos}(\sigma_i)$ and $W_e \cup \{a(\sigma_i)\} \in \mathcal{L}$. Moreover, M' is iterative and hence does not learn W_e and $W_e \cup \{a(\sigma_i)\}$. □

We transfer the notion of canny learners to learning from informant.

Definition 1. *A learner M from informant is called* canny *in case for every finite informant sequence σ holds*

1. *if $M(\sigma)$ is defined then $M(\sigma) \in \mathbb{N}$;*
2. *for every $x \in \mathbb{N}\backslash(\mathrm{pos}(\sigma) \cup \mathrm{neg}(\sigma))$ and $i \in \{0,1\}$ a mind change $M(\sigma^\frown(x,i)) \neq M(\sigma)$ implies for all finite informant sequences τ with $\sigma^\frown(x,i) \sqsubseteq \tau$ that $M(\tau^\frown(x,i)) = M(\tau)$.*

Hence, the learner is canny in case it always outputs a hypotheses and no datum twice causes a mind change of the learner. Also for learning from informant the learner can be assumed canny by a simulation argument.

Lemma 2. *For every iterative learner M, there exists a canny iterative learner N such that* $\mathbf{InfEx}(M) \subseteq \mathbf{InfEx}(N)$.

Proof. Let f be a computable 1-1 function mapping every finite informant sequence σ to a natural number encoding a program with $W_{f(\sigma)} = W_{M(\sigma)}$ if $M(\sigma) \in \mathbb{N}$ and $W_{f(\sigma)} = \varnothing$ otherwise. Clearly, σ can be reconstructed from $f(\sigma)$. We define the canny learner M' by letting

$$M'(\varnothing) = f(\varnothing)$$

$$h_{M'}(f(\sigma),(x,i)) = \begin{cases} f(\sigma^\frown(x,i)), & \text{if } x \notin \mathrm{pos}(\sigma) \cup \mathrm{neg}(\sigma) \wedge \\ & \quad M(\sigma^\frown(x,i))\downarrow \neq M(\sigma)\downarrow; \\ f(\sigma), & \text{if } M(\sigma^\frown(x,i))\downarrow = M(\sigma)\downarrow \vee \\ & \quad x \in \mathrm{content}(\sigma); \\ \uparrow, & \text{otherwise.} \end{cases}$$

M' mimics M via f on a possibly finite informant subsequence of the originally presented informant with ignoring data not causing mind changes of M or that has already caused a mind change.
Let $L \in \mathbf{InfEx}(M)$ and $I' \in \mathbf{Inf}(L)$. As M has to learn L from every informant for it, M' will always be defined. Further, let $\sigma_0 = \varnothing$ and

$$\sigma_{t+1} = \begin{cases} \sigma_t{}^\frown I'(t), & \text{if } I'(t) \notin \mathrm{ran}(\sigma_t) \wedge M(\sigma_t{}^\frown I'(t))\downarrow \neq M(\sigma_t)\downarrow; \\ \sigma_t, & \text{otherwise.} \end{cases}$$

Then by induction for all $t \in \mathbb{N}$ holds $M'(I'[t]) = f(\sigma_t)$.
The following function translates between the two settings

$$\mathfrak{r}(0) = 0;$$
$$\mathfrak{r}(t+1) = \min\{r > \mathfrak{r}(t) \mid I'(r-1) \notin \mathrm{ran}(\sigma_{\mathfrak{r}(t)})\}.$$

Intuitively, the infinite range of $\mathfrak{r}$ captures all points in time r at which a datum that has not caused a mind change so far, is seen and a mind-change of M' is possible. Thus the mind change condition is of interest in order to decide whether $\sigma_{\mathfrak{r}(t+1)} \neq \sigma_{\mathfrak{r}(t)}$. Note that $\sigma_r = \sigma_{\mathfrak{r}(t)}$ for all r with $\mathfrak{r}(t) \leqslant r < \mathfrak{r}(t+1)$.
Let $I(t) = I'(\mathfrak{r}(t+1)-1)$ for all $t \in \mathbb{N}$. Since only already observed data is ommited, I is an informant for L.
We next argue that $M(I[t]) = M(\sigma_{\mathfrak{r}(t)})$ for all $t \in \mathbb{N}$. As $I[0] = \varnothing = \sigma_0$, the claim holds for $t = 0$. Now we assume $M(I[t]) = M(\sigma_{\mathfrak{r}(t)})$ and obtain $M(I[t+1]) = M(I[t]^\frown I(t)) = M(\sigma_{\mathfrak{r}(t)}{}^\frown I(t)) = M(\sigma_{\mathfrak{r}(t+1)})$.
As by the definitions of I and $\mathfrak{r}$ we have $I(t) = I'(\mathfrak{r}(t+1)-1) \notin \mathrm{ran}(\sigma_{\mathfrak{r}(t)})$ there are two cases:

1. If $M(\sigma_{\mathfrak{r}(t)}{}^\frown I(t)) = M(\sigma_{\mathfrak{r}(t)})$, then from $\sigma_{\mathfrak{r}(t+1)-1} = \sigma_{\mathfrak{r}(t)}$ and the definition of M' we obtain $\sigma_{\mathfrak{r}(t+1)} = \sigma_{\mathfrak{r}(t)}$. Putting both together the claimed equality $M(\sigma_{\mathfrak{r}(t)}{}^\frown I(t)) = M(\sigma_{\mathfrak{r}(t+1)})$ follows.
2. If $M(\sigma_{\mathfrak{r}(t)}{}^\frown I(t)) \neq M(\sigma_{\mathfrak{r}(t)})$, the definition of M' yields $\sigma_{\mathfrak{r}(t+1)} = \sigma_{\mathfrak{r}(t)}{}^\frown I(t)$. Hence the claimed equality also holds in this case.

We now argue that M' explanatory learns L from I'. In order to see this, first observe $\sigma_{\mathfrak{r}(t+1)} = \sigma_{\mathfrak{r}(t)}$ if and only if $M(I'[t+1]) = M(I'[t])$ for every $t \in \mathbb{N}$. This is because

$$\begin{aligned}\sigma_{\mathfrak{r}(t+1)} = \sigma_{\mathfrak{r}(t)} &\Leftrightarrow M(\,\sigma_{\mathfrak{r}(t)}{}^\frown I(t)\,) = M(\sigma_{\mathfrak{r}(t)})\\ &\Leftrightarrow M(I[t]^\frown I(t)) = M(I[t])\\ &\Leftrightarrow M(I[t+1]) = M(I[t]).\end{aligned}$$

As I is an informant for L, the learner M explanatory learns L from I. Hence there exists some t_0 such that $W_{M(I[t_0])} = L$ and for all $t \geqslant t_0$ holds $M(I[t]) = M(I[t_0])$. With this follows $\sigma_{\mathfrak{r}(t)} = \sigma_{\mathfrak{r}(t_0)}$ for all $t \geqslant t_0$. As for every r there exists some t with $\mathfrak{r}(t) \leqslant r$ and $\sigma_r = \sigma_{\mathfrak{r}(t)}$, we obtain $\sigma_r = \sigma_{\mathfrak{r}(t_0)}$ for all $r \geqslant \mathfrak{r}(t_0)$. We conclude $M'(I'[t]) = f(\sigma_t) = f(\sigma_{\mathfrak{r}(t_0)})$ for all $t \geqslant \mathfrak{r}(t_0)$ and by the definition of f finally $W_{f(\sigma_{\mathfrak{r}(t_0)})} = W_{M(\sigma_{\mathfrak{r}(t_0)})} = W_{M(I[t_0])} = L$. □

5 Additional Requirements

In the following we review additional properties one might require the learning process to have in order to consider it successful. For this, we employ the following notion of consistency when learning from informant. As in [LZZ08] according to [BB75] and [Bār77] for $A \subseteq \mathbb{N}$ we define

$$\mathbf{Cons}(f, A) \quad :\Leftrightarrow \quad \mathrm{pos}(f) \subseteq A \;\wedge\; \mathrm{neg}(f) \subseteq \mathbb{N}\backslash A$$

and say f *is consistent with* A or f *is compatible with* A.
Learning restrictions incorporate certain desired properties of the learners' behavior relative to the information being presented. We state the definitions for learning from informant here.

Definition 2. *Let M be a learner and I an informant. We denote by $h_t = M(I[t])$ the hypothesis of M after observing $I[t]$ and write*

1. **Conv**(M, I) *([Ang80]), if M is* conservative on I*, i.e., for all s, t with $s \leqslant t$ holds* $\mathbf{Cons}(I[t], W_{h_s}) \Rightarrow h_s = h_t$.
2. **Dec**(M, I) *([OSW82]), if M is* decisive on I*, i.e., for all r, s, t with $r \leqslant s \leqslant t$ holds* $W_{h_r} = W_{h_t} \Rightarrow W_{h_r} = W_{h_s}$.
3. **Caut**(M, I) *([OSW86]), if M is* cautious on I*, i.e., for all s, t with $s \leqslant t$ holds* $\neg W_{h_t} \subsetneq W_{h_s}$.
4. **WMon**(M, I) *([Jan91],[Wie91]), if M is* weakly monotonic on I*, i.e., for all s, t with $s \leqslant t$ holds* $\mathbf{Cons}(I[t], W_{h_s}) \Rightarrow W_{h_s} \subseteq W_{h_t}$.
5. **Mon**(M, I) *([Jan91],[Wie91]), if M is* monotonic on I*, i.e., for all s, t with $s \leqslant t$ holds* $W_{h_s} \cap \mathrm{pos}(I) \subseteq W_{h_t} \cap \mathrm{pos}(I)$.
6. **SMon**(M, I) *([Jan91],[Wie91]), if M is* strongly monotonic on I*, i.e., for all s, t with $s \leqslant t$ holds* $W_{h_s} \subseteq W_{h_t}$.
7. **NU**(M, I) *([BCM⁺08]), if M is* non-U-shaped on I*, i.e., for all r, s, t with $r \leqslant s \leqslant t$ holds* $W_{h_r} = W_{h_t} = \mathrm{pos}(I) \Rightarrow W_{h_r} = W_{h_s}$.
8. **SNU**(M, I) *([CM11]), if M is* strongly non-U-shaped on I*, i.e., for all r, s, t with $r \leqslant s \leqslant t$ holds* $W_{h_r} = W_{h_t} = \mathrm{pos}(I) \Rightarrow h_r = h_s$.
9. **SDec**(M, I) *([KP16]), if M is* strongly decisive on I*, i.e., for all r, s, t with $r \leqslant s \leqslant t$ holds* $W_{h_r} = W_{h_t} \Rightarrow h_r = h_s$.

When additional requirements apply to the definition of learning success, we write them between **Inf** and **Ex**. For example, Theorem 1 proves $[\mathbf{ItInfConvSDecSMonEx}] \backslash [\mathcal{R}\mathbf{ItInfEx}] \neq \varnothing$ because the non-total learner acts conservatively, strongly decisively and strongly monotonically when learning $\mathcal{L}$.

The text variants can be found in [JKMS16] where all pairwise relations $=$, $\subsetneq$ or $\perp$ between the sets $[\mathbf{ItTxt}\delta\mathbf{Ex}]$ (iterative learners from text) for $\delta \in \Delta$, where $\Delta = \{\mathbf{Conv}, \mathbf{Dec}, \mathbf{Caut}, \mathbf{WMon}, \mathbf{Mon}, \mathbf{SMon}, \mathbf{NU}, \mathbf{SNU}, \mathbf{SDec}\}$, are depicted. We sum up the current status regarding the map for iterative learning from informant in the following.

For all $\delta \in \Delta \backslash \{\mathbf{SMon}\}$ with a locking sequence argument we can observe $[\mathbf{ItInfSMonEx}] \subsetneq [\mathbf{ItInf}\delta\mathbf{Ex}]$. If we denote by $\mathbf{Inf}_{\mathrm{can}}$ the set of all informants labelling the natural numbers according to their canonical order, which corresponds to the characterisic function of the respective language, we obtain $\mathbf{Fin} \cup \{\mathbb{N}\} \in [\mathcal{R}\mathbf{ItInf}_{\mathrm{can}}\mathbf{ConsConvSDecMonEx}]$ and thus it holds in contrast to full-information learning from informant $[\mathbf{ItInf}_{\mathrm{can}}\mathbf{Ex}] \neq [\mathbf{ItInfEx}]$, see [AKS18]. [LZ92] observed that requiring a monotonic behavior of the learner is restrictive, i.e. there exists an indexable family in $[\mathbf{ItInfMonEx}] \subsetneq [\mathbf{ItInfEx}]$. The indexable family $\{\mathbb{N}\} \cup \{\mathbb{N} \backslash \{x\} \mid x \in \mathbb{N}\}$ is clearly not cautiously learnable but conservatively, strongly decisively and monotonically learnable by a total iterative learner from informant. Hence, $[\mathbf{ItInfCautEx}] \perp [\mathbf{ItInfMonEx}]$. Moreover, [JLZ07] observed that requiring a conservative learning behavior is also restrictive. Indeed, they provide an indexable family in $[\mathbf{ItInfCautWMonNUDecEx}] \backslash [\mathbf{ItInfConvEx}]$ and another indexable family in $[\mathcal{R}\mathbf{ItTxtCautConvSDecEx}] \backslash [\mathbf{ItInfMonEx}]$. Hence, the map on iterative learning from informant differs from the map on iterative learning from text in [JKMS16] as **Caut** is restrictive and also from the map of full-information learning in [AKS18] from informant as **Conv** is restrictive too. It has been open how **WMon**, **Dec**, **NU**, **SDec** and **SNU** relate to each other and the other requirements. We show that also **SNU** restricts **ItInfEx** with an intricate **ORT**-argument.

Theorem 2. **[ItInfSNUEx]** $\subsetneq$ **[ItInfEx]**

In the following we provide a lemma that might help to investigate **WMon**, **Dec** and **NU**.

Definition 3. *Denote the set of all unbounded and non-decreasing functions by* $\mathfrak{S}$, *i.e.,*

$$\mathfrak{S} := \{\, \mathfrak{s} : \mathbb{N} \to \mathbb{N} \mid \forall x \in \mathbb{N}\, \exists t \in \mathbb{N} \colon \mathfrak{s}(t) \geqslant x \textit{ and } \forall t \in \mathbb{N} \colon \mathfrak{s}(t+1) \geqslant \mathfrak{s}(t) \,\}.$$

Then every $\mathfrak{s} \in \mathfrak{S}$ *is a so called* admissible simulating function.
A predicate $\beta \subseteq \mathfrak{P} \times \mathcal{I}$*, where* $\mathfrak{P}$ *stands for the set of all learners, is* semantically delayable*, if for all* $\mathfrak{s} \in \mathfrak{S}$*, all* $I, I' \in \mathcal{I}$ *and all learners* $M, M' \in \mathfrak{P}$ *holds: Whenever we have* $\mathrm{pos}(I'[t]) \supseteq \mathrm{pos}(I[\mathfrak{s}(t)])$, $\mathrm{neg}(I'[t]) \supseteq \mathrm{neg}(I[\mathfrak{s}(t)])$ *and* $W_{M'(I'[t])} = W_{M(I[\mathfrak{s}(t)])}$ *for all* $t \in \mathbb{N}$*, from* $\beta(M, I)$ *we can conclude* $\beta(M', I')$*.*

It is easy to see that every $\delta \in \{\mathbf{Caut}, \mathbf{Dec}, \mathbf{WMon}, \mathbf{Mon}, \mathbf{SMon}, \mathbf{NU}\}$ is semantically delayable and Lemma 2 can be restated as follows.

Lemma 3. *For every iterative learner* M *and every semantically delayable learning restriction* δ*, there exists a canny iterative learner* N *such that* $\mathbf{Inf}\delta\mathbf{Ex}(M) \subseteq \mathbf{Inf}\delta\mathbf{Ex}(N)$*.*

Proof. Add any semantically delayable δ in front of **Ex** in the proof of Lemma 2. We define a simulating function (Definition 3) by $\mathfrak{s}(t) = \max\{s \in \mathbb{N} \mid \mathfrak{r}(s) \leqslant t\}$. It is easy to check that $\mathfrak{s}$ is unbounded and clearly it is non-decreasing. Then by the definitions of I and $\mathfrak{s}$ we have $\mathrm{pos}(I[\mathfrak{s}(t)]) \subseteq \mathrm{pos}(I'[\mathfrak{r}(\mathfrak{s}(t))]) \subseteq \mathrm{pos}(I'[t])$ and similarly $\mathrm{neg}(I[\mathfrak{s}(t)]) \subseteq \mathrm{neg}(I'[t])$ for all $t \in \mathbb{N}$. As $M'(I'[t]) = f(\sigma_t)$ and $M(\sigma_{\mathfrak{r}(\mathfrak{s}(t))}) = M(I[\mathfrak{s}(t)])$ for all $t \in \mathbb{N}$, in order to obtain $W_{M'(I'[t])} = W_{M(I[\mathfrak{s}(t)])}$ it suffices to show $W_{f(\sigma_t)} = W_{M(\sigma_{\mathfrak{r}(\mathfrak{s}(t))})}$. Since $W_{f(\sigma_t)} = W_{M(\sigma_t)}$ for all $t \in \mathbb{N}$, this can be concluded from $\sigma_t = \sigma_{\mathfrak{r}(\mathfrak{s}(t))}$. But this obviously holds because $\mathfrak{r}(\mathfrak{s}(t)) \leqslant t < \mathfrak{r}(\mathfrak{s}(t)+1)$ follows from the definition of $\mathfrak{s}$.
Finally, from $\delta(M, I)$ we conclude $\delta(M', I')$. □

Two other learning restrictions that might be helpful to understand the syntactic learning criteria **SNU**, **SDec** and **Conv** better are the following.

Definition 4. *Let* M *be a learner and* I *an informant. We denote by* $h_t = M(I[t])$ *the hypothesis of* M *after observing* $I[t]$ *and write*

1. $\mathbf{LocConv}(M, I)$ *([JLZ07]), if* M *is* locally conservative *on* I*, i.e., for all* t *holds* $h_t \neq h_{t+1} \Rightarrow \neg\mathbf{Cons}(I(t), W_{h_t})$*.*
2. $\mathbf{Wb}(M, I)$ *([KS16]), if* M *is* witness-based *on* I*, i.e., for all* r, s, t *with* $r < s \leqslant t$ *the mind-change* $h_r \neq h_s$ *implies* $\mathrm{pos}(I[s]) \cap W_{h_t} \backslash W_{h_r} \neq \varnothing \vee \mathrm{neg}(I[s]) \cap W_{h_r} \backslash W_{h_t} \neq \varnothing$*.*

Hence, in a locally conservative learning process every mind-change is justified by the datum just seen. Moreover, a in witness-based learning process each mind-change is witnessed by some false negative or false positive datum. Obviously, **LocConv** $\Rightarrow$ **Conv** and **Wb** $\Rightarrow$ **Conv**.
As for learning from text, see [JKMS16], we gain that every concept class locally conservatively learnable by an iterative learner from informant is also learnable in a witness-based fashion by an iterative learner.

Theorem 3. $[\mathbf{ItInfLocConvEx}] \subseteq [\mathbf{ItInfWbEx}]$

Proof. Let $\mathcal{L}$ be a concept class learned by the iterative learner M in a locally conservative manner. As we are interested in a witness-based learner N, we always enlarge the guess of M by all data witnessing a mind-change in the past. As we want N to be iterative, this is done via padding the set of witnesses to the hypothesis and a total computable function g adding this information to the hypothesis of M as follows:

$$W_{g(\mathrm{pad}(h,\langle MC\rangle))} = (W_h \cup \mathrm{pos}[MC]) \setminus \mathrm{neg}[MC];$$
$$N(\varnothing) = g(\mathrm{pad}(M(\varnothing), \langle\varnothing\rangle));$$
$$h_N(g(\mathrm{pad}(h,\langle MC\rangle)), \xi) = \begin{cases} g(\mathrm{pad}(h,\langle MC\rangle)), & \text{if } h_M(h,\xi) = h \vee \\ & \quad \xi \in MC; \\ g(\mathrm{pad}(h_M(h,\xi), & \\ \quad \langle MC \cup \{\xi\}\rangle)), & \text{otherwise.} \end{cases}$$

Clearly, N is iterative. Further, whenever M is locked on h and $W_h = L$, since MC is consistent with L, we also have $W_{g(\mathrm{pad}(f(h),\langle MC\rangle))} = L$. As N simulates M on an informant omitting all data that already caused a mind-change beforehand, N does explanatory learn $\mathcal{L}$. As M learns locally conservatively and by employing g, the learner N acts witness-based. □

6 Suggestions for Future Research

Future work should address the complete map for iterative learners from informant. In particular, **WMon**, **Dec** and **NU** seem to be challenging as the proofs in related settings fail without an obvious fix. We hope that Lemma 3 is a helping hand in this endeavour. Also the equivalence of the syntactic criteria **SNU**, **SDec** and **Conv** does not trivially hold. Theorem 3 might be helpful regarding the latter.

Maps for other models of memory-limited learning, such as **BMS**, see [CCJS07], or **Bem**, see [FJO94], [LZ96] and [CJLZ99], would help to rate models.

Last but not least we encourage to investigate the learnability of indexable classes motivated by grammatical inference or machine learning research. The pattern languages often serve as a helpful example to refer to and we hope for even more examples of this kind. As a starting point, in the full version, we prove the learnability of half-spaces.

Acknowledgements

This work was supported by DFG Grant Number KO 4635/1-1. We are grateful to the people supporting us.

References

AKS18. M. Aschenbach, T. Kötzing, and K. Seidel. Learning from informants: Relations between learning success criteria. *arXiv preprint arXiv:1801.10502*, 2018.

Ang80. D. Angluin. Inductive inference of formal languages from positive data. *Information and control*, 45(2):117–135, 1980.

Bār77. J. Bārzdiņš. Inductive inference of automata, functions and programs. In *Amer. Math. Soc. Transl.*, pages 107–122, 1977.

BB75. L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.

BCM+08. G. Baliga, J. Case, W. Merkle, F. Stephan, and R. Wiehagen. When unlearning helps. *Information and Computation*, 206:694–709, 2008.

Cas74. J. Case. Periodicity in generations of automata. *Mathematical Systems Theory*, 8(1):15–32, 1974.

Cas94. J. Case. Infinitary self-reference in learning theory. *Journal of Experimental and Theoretical Artificial Intelligence*, 6:3–16, 1994.

CCJS07. L. Carlucci, J Case, S. Jain, and F. Stephan. Results on memory-limited U-shaped learning. *Information and Computation*, 205:1551–1573, 2007.

CJLZ99. J. Case, S. Jain, S. Lange, and T. Zeugmann. Incremental concept learning for bounded data mining. *Information and Computation*, 152:74–110, 1999.

CK10. J. Case and T. Kötzing. Strongly non-U-shaped learning results by general techniques. In Adam Tauman Kalai and Mehryar Mohri, editors, *COLT 2010*, pages 181–193, 2010.

CM08. J. Case and S. E. Moelius. U-shaped, iterative, and iterative-with-counter learning. *Machine Learning*, 72:63–88, 2008.

CM09. J. Case and S. Moelius. Parallelism increases iterative learning power. *Theoretical Computer Science*, 410(19):1863 – 1875, 2009.

CM11. J. Case and S. Moelius. Optimal language learning from positive data. *Information and Computation*, 209:1293–1311, 2011.

FJO94. M. Fulk, S. Jain, and D. Osherson. Open problems in Systems That Learn. *Journal of Computer and System Sciences*, 49(3):589–604, December 1994.

Gol67. E. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

Jan91. K. P. Jantke. Monotonic and nonmonotonic inductive inference of functions and patterns. In *Nonmonotonic and Inductive Logic, 1st International Workshop, Proc.*, pages 161–177, 1991.

JKMS16. S. Jain, T. Kötzing, J. Ma, and F. Stephan. On the role of update constraints and text-types in iterative learning. *Information and Computation*, 247:152–168, 2016.

JLZ07. S. Jain, S. Lange, and S. Zilles. Some natural conditions on incremental learning. *Information and Computation*, 205:1671–1684, 2007.

JMZ13. S. Jain, S. Moelius, and S. Zilles. Learning without coding. *Theoretical Computer Science*, 473:124–148, 2013.

JORS99. S. Jain, D. Osherson, J. Royer, and A. Sharma. *Systems that Learn: An Introduction to Learning Theory.* MIT Press, Cambridge, Massachusetts, second edition, 1999.

Köt09. T. Kötzing. *Abstraction and Complexity in Computational Learning in the Limit.* PhD thesis, University of Delaware, 2009.

KP16. T. Kötzing and R. Palenta. A map of update constraints in inductive inference. *Theoretical Computer Science*, 650:4–24, 2016.

KS16. T. Kötzing and M. Schirneck. Towards an atlas of computational learning theory. In *33rd Symposium on Theoretical Aspects of Computer Science*, 2016.

KSS17. T. Kötzing, M. Schirneck, and K. Seidel. Normal forms in semantic language identification. In *Proc. of Algorithmic Learning Theory*, pages 493–516. PMLR, 2017.

LZ92. S. Lange and T. Zeugmann. Types of monotonic language learning and their characterization. In *Proc. 5th Annual ACM Workshop on Comput. Learning Theory*, pages 377–390, New York, NY, 1992.

LZ96. S. Lange and T. Zeugmann. Incremental learning from positive data. *Journal of Computer and System Sciences*, 53:88–103, 1996.

LZZ08. S. Lange, T. Zeugmann, and S. Zilles. Learning indexed families of recursive languages from positive data: A survey. *Theoretical Computer Science*, 397(1):194–232, 2008.

Odi99. P. Odifreddi. *Classical Recursion Theory*, volume II. Elsivier, Amsterdam, 1999.

OSW82. D. Osherson, M. Stob, and S. Weinstein. Learning strategies. *Information and Control*, 53:32–51, 1982.

OSW86. D. Osherson, M. Stob, and S. Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists.* MIT Press, Cambridge, Mass., 1986.

RC94. J. Royer and J. Case. *Subrecursive Programming Systems: Complexity and Succinctness.* Research monograph in *Progress in Theoretical Computer Science.* Birkhäuser Boston, 1994.

Rog67. H. Rogers. *Theory of Recursive Functions and Effective Computability.* McGraw Hill, New York, 1967. Reprinted, MIT Press, 1987.

Wie76. R. Wiehagen. Limes-erkennung rekursiver funktionen durch spezielle strategien. *J. Inf. Process. Cybern.*, 12 (1-2):93–99, 1976.

Wie91. R. Wiehagen. A thesis in inductive inference. In *Nonmonotonic and Inductive Logic, 1st International Workshop, Proc.*, pages 184–207, 1991.