Learning Recursive Concepts with Anomalies

Gunter Grieser[1], Steffen Lange[2], and Thomas Zeugmann[3]

[1] Technische Universität Darmstadt, Fachbereich Informatik, Alexanderstr. 10
64283 Darmstadt, Germany, e-mail: grieser@informatik.tu-darmstadt.de

[2] Universität Leipzig, Institut für Informatik, Augustusplatz 10–11
04109 Leipzig, Germany, e-mail: slange@informatik.uni-leipzig.de

[3] Medizinische Universität Lübeck, Institut für Theoretische Informatik, Wallstr. 40
23560 Lübeck, Germany, e-mail: thomas@tcs.mu-luebeck.de

Abstract. This paper provides a systematic study of inductive inference of indexable concept classes in learning scenarios in which the learner is successful if its final hypothesis describes a finite variant of the target concept – henceforth called learning with anomalies. As usual, we distinguish between learning from only positive data and learning from positive and negative data.
We investigate the following learning models: finite identification, conservative inference, set-driven learning, and behaviorally correct learning. In general, we focus our attention on the case that the number of allowed anomalies is finite but not *a priori* bounded. However, we also present a few sample results that affect the special case of learning with an *a priori* bounded number of anomalies. We provide characterizations of the corresponding models of learning with anomalies in terms of finite tell-tale sets. The varieties in the degree of recursiveness of the relevant tell-tale sets observed are already sufficient to quantify the differences in the corresponding models of learning with anomalies.
In addition, we study variants of incremental learning and derive a complete picture concerning the relation of all models of learning with and without anomalies mentioned above.

1 Introduction

Induction constitutes an important feature of learning. The corresponding theory is called inductive inference. Inductive inference may be characterized as the study of systems that map evidence on a target concept into hypotheses about it. The investigation of scenarios in which the sequence of hypotheses stabilizes to an accurate and finite description of the target concept is of some particular interest. The precise definitions of the notions evidence, stabilization, and accuracy go back to Gold [10] who introduced the model of learning in the limit.

The present paper deals with inductive inference of indexable classes of recursive concepts (indexable classes, for short). A concept class is said to be an indexable class if it possesses an effective enumeration with uniformly decidable membership. Angluin [2] started the systematic study of learning indexable concept classes. [2] and succeeding publications (cf., e.g., [20], for an overview)

found a lot of interest, since most natural concept classes form indexable classes. For example, the class of all context sensitive, context free, regular, and pattern languages as well as the set of all boolean formulas expressible as monomial, k-CNF, k-DNF, and k-decision list constitute indexable classes.

As usual, we distinguish learning from positive data and learning from positive and negative data, synonymously called learning from text and informant, respectively. A text for a target concept c is an infinite sequence of elements of c such that every element from c eventually appears. Alternatively, an informant is an infinite sequence of elements exhausting the underlying learning domain that are classified with respect to their membership to the target concept.

An algorithmic learner takes as input larger and larger initial segments of a text (an informant) and outputs, from time to time, a hypothesis about the target concept. The set of all admissible hypotheses is called hypothesis space. When learning of indexable classes is considered, it is natural to require that the hypothesis space is an effective enumeration of a (possibly larger) indexable concept class. This assumption underlies almost all studies (cf., e.g., [2, 20]).

Gold's [10] original model requires the sequence of hypotheses to converge to a hypothesis correctly describing the target concept. However, from a viewpoint of potential applications, it suffices in most cases that the final hypothesis approximates the target concept sufficiently well. Blum and Blum [5] introduced a quite natural refinement of Gold's model that captures this aspect. In their setting of learning recursive functions with anomalies, it is admissible that the learner's final hypothesis may differ from the target function at finitely many data points. Case and Lynes [6] adapted this model to language learning.

Learning with anomalies has been studied intensively in the context of learning recursive functions and recursively enumerable languages (cf., e.g., [11]). Preliminary results concerning the learnability of indexable classes with anomalies can be found in Tabe and Zeugmann [17]. Note that Baliga *et al.* [3] studied the learnability of indexable classes with anomalies, too. However, unlike all other work on learning indexable classes, [3] allows the use of *arbitrary* hypothesis spaces (including those not having a decidable membership problem). Therefore, the results from [3] do not directly translate into our setting.

The present paper provides a systematic study of learning indexable concept classes with anomalies. We investigate the following variants of Gold-style concept learning: finite identification, conservative inference, set-driven inference, behaviorally correct learning, and incremental learning. We relate the resulting models of learning with anomalies to one another as well as to the corresponding versions of learning without anomalies. In general, we focus our attention to the case that the number of allowed anomalies is finite but not *a priori* bounded. However, we also present a few sample results that affect the special case that the number of allowed anomalies is *a priori* bounded.

Next, we mention some prototypical results. In the setting of learning with anomalies, the learning power of set-driven learners, conservative learners, and unconstrained IIMs does coincide. In contrast, when anomaly-free learning is considered, conservative learners and set-driven learners are strictly less power-

ful. Moreover, a further difference to learning without anomalies is established by showing that behaviorally correct learning with anomalies is strictly more powerful than learning in the limit with anomalies. Furthermore, in case the number of allowed anomalies is finite but not *a priori* bounded, it is proved that there is no need to use arbitrary hypothesis spaces in order to design superior behaviorally correct learners, thus refining the corresponding results from [3]. However, if the number of anomalies is *a priori* bounded, it is advantageous to use arbitrary hypothesis spaces. In order to establish these results, we provide characterizations of the corresponding models of learning with anomalies in terms of finite tell-tale sets (cf. [2]). As it turns out, the observed varieties in the degree of recursiveness of the relevant tell-tale sets are already sufficient to quantify the differences in the corresponding models of learning with anomalies.

Moreover, we derive a complete picture concerning the relation of the different models of incremental learning with and without anomalies.

2 Preliminaries

2.1 Basic notions

Let $\mathbb{N} = \{0, 1, 2, \ldots\}$ be the set of all natural numbers. By $\langle ., . \rangle \colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ we denote Cantor's pairing function. Let A and B be sets. As usual, $A \triangle B$ denotes the symmetrical difference of A and B, i.e., $A \triangle B = (A \setminus B) \cup (B \setminus A)$. We write $A \# B$ to indicate that $A \triangle B \neq \emptyset$. For all $a \in \mathbb{N}$, $A =^a B$ iff $card(A \triangle B) \leq a$, while $A =^* B$ iff $card(A \triangle B) < \infty$. We let $\sigma \diamond \tau$ denote the concatenation of two possibly infinite sequences σ and τ.

Any recursively enumerable set $\mathcal{X}$ is called a *learning domain*. By $\wp(\mathcal{X})$ we denote the power set of $\mathcal{X}$. Let $\mathcal{C} \subseteq \wp(\mathcal{X})$ and let $c \in \mathcal{C}$. We refer to $\mathcal{C}$ and c as to a *concept class* and a *concept*, respectively. Sometimes, we will identify a concept c with its characteristic function, i.e., we let $c(x) = +$, if $x \in c$, and $c(x) = -$, otherwise. What is actually meant will become clear from the context.

We deal with the learnability of indexable concept classes with uniformly decidable membership defined as follows (cf. [2]). A class of non-empty concepts $\mathcal{C}$ is said to be an *indexable concept class with uniformly decidable membership* if there are an effective enumeration $(c_j)_{j \in \mathbb{N}}$ of all and only the concepts in $\mathcal{C}$ and a recursive function f such that, for all $j \in \mathbb{N}$ and all $x \in \mathcal{X}$, it holds $f(j, x) = +$, if $x \in c_j$, and $f(j, x) = -$, otherwise. We refer to indexable concept classes with uniformly decidable membership as to *indexable classes*, for short, and let $\mathcal{IC}$ denote the collection of all indexable classes.

2.2 Gold-style concept learning

Let $\mathcal{X}$ be the underlying learning domain, let $c \subseteq \mathcal{X}$ be a concept, and let $t = (x_n)_{n \in \mathbb{N}}$ be an infinite sequence of elements from c such that $\{x_n \mid n \in \mathbb{N}\} = c$. Then, t is said to be a *text* for c. By $Text(c)$ we denote the set of all texts for c. Let t be a text and let y be a number. Then, t_y denotes the initial segment of t of length $y + 1$. Furthermore, we set $content(t_y) = \{x_n \mid n \leq y\}$.

Let $\mathcal{C}$ be an indexable class. Then, we let $\mathit{Text}(\mathcal{C})$ be the collection of all texts in $\bigcup_{c \in \mathcal{C}} \mathit{Text}(c)$.

As in [10], we define an *inductive inference machine* (abbr. *IIM*) to be an algorithmic mapping from initial segments of texts to $\mathbb{N} \cup \{?\}$. Thus, an IIM either outputs a hypothesis, i.e., a number encoding a certain computer program, or it outputs "?," a special symbol representing the case the machine outputs "no conjecture." Note that an IIM, when learning some target class $\mathcal{C}$, is required to produce an output when processing any *admissible* information sequence, i.e., any initial segment of any text in $\mathit{Text}(\mathcal{C})$.

The numbers output by an IIM are interpreted with respect to a suitably chosen *hypothesis space* $\mathcal{H} = (h_j)_{j \in \mathbb{N}}$. Since we exclusively deal with the learnability of indexable classes $\mathcal{C}$, we always assume that $\mathcal{H}$ is also an indexing of some possibly larger indexable class. Hence, membership is uniformly decidable in $\mathcal{H}$, too. If $\mathcal{C} \subseteq \{h_j \mid j \in \mathbb{N}\}$ ($\mathcal{C} = \{h_j \mid j \in \mathbb{N}\}$), then $\mathcal{H}$ is said to be a *class comprising* (*class preserving*) hypothesis space for $\mathcal{C}$ (cf. [20]). When an IIM outputs some number j, we interpret it to mean that it hypothesizes h_j.

We define convergence of IIMs as usual. Let t be a text and let M be an IIM. The sequence $(M(t_y))_{y \in \mathbb{N}}$ of M's hypotheses *converges* to a number j iff all but finitely many terms of it are equal to j.

Now, we are ready to define *learning in the limit.*

Definition 1 ([6, 10]). *Let $\mathcal{C} \in \mathcal{IC}$, let c be a concept, let $\mathcal{H} = (h_j)_{j \in \mathbb{N}}$ be a hypothesis space, and let $a \in \mathbb{N} \cup \{*\}$.*

An IIM M $\mathit{Lim}^a\mathit{Txt}_{\mathcal{H}}$–identifies c iff, for every $t \in \mathit{Text}(c)$, there is a $j \in \mathbb{N}$ with $h_j =^a c$ such that the sequence $(M(t_y))_{y \in \mathbb{N}}$ converges to j.

M $\mathit{Lim}^a\mathit{Txt}_{\mathcal{H}}$–identifies $\mathcal{C}$ iff, for all $c' \in \mathcal{C}$, M $\mathit{Lim}^a\mathit{Txt}_{\mathcal{H}}$–identifies c'.

$\mathit{Lim}^a\mathit{Txt}$ denotes the collection of all indexable classes $\mathcal{C}'$ for which there are a hypothesis space $\mathcal{H}' = (h'_j)_{j \in \mathbb{N}}$ and an IIM M such that M $\mathit{Lim}^a\mathit{Txt}_{\mathcal{H}'}$–identifies $\mathcal{C}'$.

Subsequently, we write LimTxt instead of $\mathit{Lim}^0\mathit{Txt}$. We adopt this convention to all learning types defined below.

In general, it is not decidable whether or not an IIM has already converged on a text t for the target concept c. Adding this requirement to the above definition results in *finite learning* (cf. [10]). The resulting learning type is denoted by $\mathit{Fin}^a\mathit{Txt}$, where again $a \in \mathbb{N} \cup \{*\}$.

Next, we define *conservative* IIMs. Intuitively speaking, conservative IIMs maintain their actual hypothesis at least as long as they have not seen data contradicting it.

Definition 2 ([2]). *Let $\mathcal{C} \in \mathcal{IC}$, let c be a concept, let $\mathcal{H} = (h_j)_{j \in \mathbb{N}}$ be a hypothesis space, and let $a \in \mathbb{N} \cup \{*\}$.*

An IIM M $\mathit{Consv}^a\mathit{Txt}_{\mathcal{H}}$–identifies c iff M $\mathit{Lim}^a\mathit{Txt}_{\mathcal{H}}$–identifies c and, for every $t \in \mathit{Text}(c)$ and for any two consecutive hypotheses $k = M(t_y)$ and $j = M(t_{y+1})$, if $k \in \mathbb{N}$ and $k \neq j$, then $\mathit{content}(t_{y+1}) \not\subseteq h_k$.

M $\mathit{Consv}^a\mathit{Txt}_{\mathcal{H}}$–identifies $\mathcal{C}$ iff, for all $c' \in \mathcal{C}$, M $\mathit{Consv}^a\mathit{Txt}_{\mathcal{H}}$–identifies c'.

For every $a \in \mathbb{N} \cup \{*\}$, the resulting learning type $\mathit{Consv}^a\mathit{Txt}$ is defined analogously to Definition 1.

Next, we define *set-driven* learning. Intuitively speaking, the output of a set-driven IIM depends exclusively on the content of its input, thereby ignoring the order as well as the frequency in which the examples occur.

Definition 3 ([18]). *Let $\mathcal{C} \in \mathcal{IC}$, let c be a concept, let $\mathcal{H} = (h_j)_{j \in \mathbb{N}}$ be a hypothesis space, and let $a \in \mathbb{N} \cup \{*\}$.*

An IIM M $Sdr^a Txt_{\mathcal{H}}$–identifies c iff M $Lim^a Txt_{\mathcal{H}}$–identifies c and, for every $t, t' \in Text(\mathcal{C})$ and for all $n, m \in \mathbb{N}$, if $content(t_n) = content(t'_m)$ then $M(t_n) = M(t'_m)$.

M $Sdr^a Txt_{\mathcal{H}}$–identifies $\mathcal{C}$ iff, for all $c' \in \mathcal{C}$, M $Sdr^a Txt_{\mathcal{H}}$–identifies c'.

For every $a \in \mathbb{N} \cup \{*\}$, the resulting learning type $Sdr^a Txt$ is defined analogously to Definition 1.

At the end of this subsection, we provide a formal definition of *behaviorally correct learning*.

Definition 4 ([4, 6]). *Let $\mathcal{C} \in \mathcal{IC}$, let c be a concept, let $\mathcal{H} = (h_j)_{j \in \mathbb{N}}$ be a hypothesis space, and let $a \in \mathbb{N} \cup \{*\}$.*

An IIM M $Bc^a Txt_{\mathcal{H}}$–identifies c iff, for every $t \in Text(c)$ and for all but finitely many $y \in \mathbb{N}$, $h_{M(t_y)} =^a c$.

M $Bc^a Txt_{\mathcal{H}}$–identifies $\mathcal{C}$ iff, for all $c' \in \mathcal{C}$, M $Bc^a Txt_{\mathcal{H}}$–identifies c'.

For every $a \in \mathbb{N} \cup \{*\}$, the resulting learning type $Bc^a Txt$ is defined analogously to Definition 1.

2.3 Incremental concept learning

Now, we formally define the different models of incremental learning. An ordinary IIM M has always access to the whole history of the learning process, i.e., it computes its actual guess on the basis of the whole initial segment of the text t seen so far. In contrast, an *iterative* IIM is only allowed to use its last guess and the next element in t. Conceptually, an iterative IIM M defines a sequence $(M_n)_{n \in \mathbb{N}}$ of machines each of which takes as its input the output of its predecessor.

Definition 5 ([19]). *Let $\mathcal{C} \in \mathcal{IC}$, let c be a concept, let $\mathcal{H} = (h_j)_{j \in \mathbb{N}}$ be a hypothesis space, and let $a \in \mathbb{N} \cup \{*\}$.*

An IIM M $It^a Txt_{\mathcal{H}}$–identifies c iff, for every $t = (x_n)_{n \in \mathbb{N}} \in Text(c)$, the following conditions are fulfilled:

(1) *for all $n \in \mathbb{N}$, $M_n(t)$ is defined, where $M_0(t) = M(x_0)$ and $M_{n+1}(t) = M(M_n(t), x_{n+1})$.*
(2) *the sequence $(M_n(t))_{n \in \mathbb{N}}$ converges to a number j with $h_j =^a c$.*

M $It^a Txt_{\mathcal{H}}$–identifies $\mathcal{C}$ iff, for each $c' \in \mathcal{C}$, M $It^a Txt_{\mathcal{H}}$–identifies c'.

For every $a \in \mathbb{N} \cup \{*\}$, the resulting learning type $It^a Txt$ is defined analogously to Definition 1.

Let M be an iterative IIM as defined in Definition 5 and t be a text. Then, $M_*(t_n)$ denotes the last hypothesis output by M when processing t_n, i.e., $M_*(t_n) = M_n(t)$. We adopt this convention to all versions of incremental learners defined below.

Next, we consider a natural relaxation of iterative learning, named *k-bounded example-memory* inference. Now, an IIM M is allowed to memorize at most k of the elements in t which it has already seen, where $k \in \mathbb{N}$ is *a priori* fixed. Again, M defines a sequence $(M_n)_{n\in\mathbb{N}}$ of machines each of which takes as input the output of its predecessor. A k-bounded example-memory IIM outputs a hypothesis along with the set of memorized data elements.

Definition 6 ([15]). *Let $\mathcal{C} \in \mathcal{IC}$, let c be a concept, let $\mathcal{H} = (h_j)_{j\in\mathbb{N}}$ be a hypothesis space, let $a \in \mathbb{N} \cup \{*\}$, and let $k \in \mathbb{N}$.*

An IIM M $Bem_k^a Txt_{\mathcal{H}}$–identifies c iff, for every $t = (x_n)_{n\in\mathbb{N}} \in Text(c)$, the following conditions are satisfied:

(1) *for all $n \in \mathbb{N}$, $M_n(t)$ is defined, where $M_0(t) = M(x_0) = \langle j_0, S_0\rangle$ such that $S_0 \subseteq \{x_0\}$ and $card(S_0) \leq k$ and $M_{n+1}(t) = M(M_n(t), x_{n+1}) = \langle j_{n+1}, S_{n+1}\rangle$ such that $S_{n+1} \subseteq S_n \cup \{x_{n+1}\}$ and $card(S_{n+1}) \leq k$.*

(2) *the j_n in the sequence $(\langle j_n, S_n\rangle)_{n\in\mathbb{N}}$ of M's guesses converge to a number j with $h_j =^a c$.*

M $Bem_k^a Txt_{\mathcal{H}}$–identifies $\mathcal{C}$ iff, for each $c' \in \mathcal{C}$, M $Bem_k^a Txt_{\mathcal{H}}$–identifies c'.

For every $k \in \mathbb{N}$ and every $a \in \mathbb{N} \cup \{*\}$, the resulting learning type $Bem_k^a Txt$ is defined analogously to Definition 1. By definition, $Bem_0^a Txt = It^a Txt$.

Next, we define learning by *feedback* IIMs. Informally speaking, a feedback IIM M is an iterative IIM that is additionally allowed to make a particular type of queries. In each learning stage $n+1$, M has access to the actual input x_{n+1} and its previous guess j_n. Moreover, M computes a query from x_{n+1} and j_n which concerns the history of the learning process. That is, the feedback learner computes a data element x and receives a "YES/NO" answer $A(x)$ such that $A(x) = 1$, if $x \in content(t_n)$, and $A(x) = 0$, otherwise. Hence, M can just ask whether or not the particular data element x has already been presented in previous learning stages.

Definition 7 ([19]). *Let $\mathcal{C} \in \mathcal{IC}$, let c be a concept, let $\mathcal{H} = (h_j)_{j\in\mathbb{N}}$ be a hypothesis space, let $a \in \mathbb{N} \cup \{*\}$, and let $Q\colon \mathbb{N} \times \mathcal{X} \to \mathcal{X}$ be a total computable function. An IIM M, with a computable query asking function Q, $Fb^a Txt_{\mathcal{H}}$–identifies c iff, for every $t = (x_n)_{n\in\mathbb{N}} \in Text(c)$, the following conditions are satisfied:*

(1) *for all $n \in \mathbb{N}$, $M_n(t)$ is defined, where $M_0(t) = M(x_0)$ as well as $M_{n+1}(t) = M(M_n(t), A(Q(M_n(t), x_{n+1})), x_{n+1})$.*

(2) *the sequence $(M_n(t))_{n\in\mathbb{N}}$ converges to a number j with $h_j =^a c$ provided A truthfully answers the questions computed by Q.*

M $Fb^a Txt_{\mathcal{H}}$–identifies $\mathcal{C}$ iff, for each $c' \in \mathcal{C}$, M $Fb^a Txt_{\mathcal{H}}$–identifies c'.

For every $a \in \mathbb{N} \cup \{*\}$, the resulting learning type $Fb^a Txt$ is defined analogously to Definition 1.

3 Learning from positive data only

In this section, we study the power and the limitations of the various models of learning with anomalies. We relate these models to one another as well as to

the different models of anomaly-free learning. We are mainly interested in the case that the number of allowed anomalies is finite but not *a priori* bounded. Nevertheless, in order to give an impression of how the overall picture changes when the number of allowed anomalies is *a priori* bounded, we also present selected results for this case.

3.1 Gold-style learning with anomalies

Proposition 1 summarizes the known relations between the considered models of anomaly-free learning from text.

Proposition 1 ([10, 14, 16]).
$FinTxt \subset ConsvTxt = SdrTxt \subset LimTxt = BcTxt \subset \mathcal{IC}$.

In the setting of learning recursive functions the first observation made when comparing learning in the limit with anomalies to behaviorally correct inference was the *error correcting power* of Bc-learners, i.e., $Ex^* \subseteq Bc$ (cf., e.g., [4, 7]). Interestingly enough, this result did not translate into the setting of learning recursively enumerable languages from positive data (cf. [6]). But still, a certain error correcting power is preserved in this setting, since $Lim^aTxt \subseteq Bc^bTxt$ provided $a \leq 2b$ (cf. [6]).

When comparing learning with and without anomalies in our setting of learning indexable classes, it turns out that even finite learners may become more powerful than Bc-learners.

Theorem 1. $Fin^1Txt \setminus BcTxt \neq \emptyset$.

However, the opposite is also true. For instance, PAT, the well-known class of all pattern languages (cf. [2]), witnesses the even stronger result:

Theorem 2. $ConsvTxt \setminus Fin^*Txt \neq \emptyset$.

As we will see, the relation between the standard learning models changes considerably, if it is no longer required that the learner must almost always output hypotheses that describe the target concept correctly. The following picture displays the established coincidences and differences by relating the models of learning with anomalies to one another and by ranking them in the hierarchy of the models of anomaly-free learning.

$$\begin{array}{ccccccccc}
Fin^*Txt & \subset & Consv^*Txt & = & Sdr^*Txt & = & Lim^*Txt & \subset & Bc^*Txt \subset \mathcal{IC} \\
\cup & & \cup & & \cup & & \cup & & \cup \\
FinTxt & \subset & ConsvTxt & = & SdrTxt & \subset & LimTxt & = & BcTxt
\end{array}$$

To achieve the overall picture, we establish characterizations of all models of learning with a finite but not *a priori* bounded number of anomalies. On the one hand, we present characterizations in terms of finite tell-tale sets. On the other hand, we prove that some of the learning models coincide.

Proposition 2 ([17]). *For all $\mathcal{C} \in \mathcal{IC}$ and all $a \in \mathbb{N} \cup \{*\}$: $\mathcal{C} \in Lim^aTxt$ iff there is an indexing $(c_j)_{j\in\mathbb{N}}$ of $\mathcal{C}$ and a recursively enumerable family $(T_j)_{j\in\mathbb{N}}$ of finite sets such that*

(1) *for all* $j \in \mathbb{N}$, $T_j \subseteq c_j$,
(2) *for all* $j, k \in \mathbb{N}$, *if* $T_j \subseteq c_k \subseteq c_j$, *then* $c_k =^a c_j$.

The characterization of Fin^*Txt is similar to the known characterization of $FinTxt$ (cf. [13]).

Theorem 3. *For all* $\mathcal{C} \in \mathcal{IC}$: $\mathcal{C} \in Fin^*Txt$ *iff there is an indexing* $(c_j)_{j\in\mathbb{N}}$ *of* $\mathcal{C}$ *and a recursively generable family* $(T_j)_{j\in\mathbb{N}}$ *of finite sets such that*

(1) *for all* $j \in \mathbb{N}$, $T_j \subseteq c_j$,
(2) *for all* $j, k \in \mathbb{N}$, *if* $T_j \subseteq c_k$, *then* $c_k =^* c_j$.

In contrast to Proposition 1, when a finite number of errors in the final hypothesis is allowed, conservative IIMs become exactly as powerful as unconstrained IIMs.

Theorem 4. $Lim^*Txt = Consv^*Txt$.

Proof. Let $\mathcal{C} \in Lim^*Txt$, let $\mathcal{H} = (h_j)_{j\in\mathbb{N}}$ be a hypothesis space, and let M be an IIM that $Lim^*Txt_{\mathcal{H}}$–identifies $\mathcal{C}$. Moreover, assume that M never outputs "?." The conservative IIM M' uses the following hypothesis space $\mathcal{H}'$. For all $j \in \mathbb{N}$ and $x \in \mathcal{X}$, we let $h'_{j,x} = h_j \setminus \{x\}$. Moreover, we let $\mathcal{H}'$ be the canonical enumeration of all those concepts $h'_{j,x}$.

Let $c \in \mathcal{C}$, let $t = (x_j)_{j\in\mathbb{N}}$ be a text for c, and let $y \in \mathbb{N}$. On input t_y, M' determines $j = M(t_y)$, and outputs the canonical index of h'_{j,x_0} in $\mathcal{H}'$.

It is straightforward to verify that M is a conservative IIM that witnesses $\mathcal{C} \in Lim^*Txt$. □

As it turns out, when learning with anomalies is considered, set-driven learners become exactly as powerful as unconstrainted IIMs, again nicely contrasting Proposition 1.

Theorem 5. $Sdr^*Txt = Lim^*Txt$.

However, there is a difference between conservative inference and set-driven learning, on the one hand, and learning in the limit, on the other hand, which we want to point out next. While learning in the limit is invariant to the choice of the hypothesis space (cf. [17]), conservative inference and set-driven learning, respectively, is not. Moreover, in order to design a superior conservative and a set-driven learner, respectively, it is sometimes inevitable to select a hypothesis space that contains concepts which are not subject to learning.

Theorem 6.
(1) *There is an indexable class* $\mathcal{C} \in Consv^*Txt$ *such that, for all class preserving hypothesis spaces* $\mathcal{H}$ *for* $\mathcal{C}$, *there is no IIM* M *that* $Consv^*Txt_{\mathcal{H}}$*–identifies* $\mathcal{C}$.
(2) *There is an indexable class* $\mathcal{C} \in Sdr^*Txt$ *such that, for all class preserving hypothesis spaces* $\mathcal{H}$ *for* $\mathcal{C}$, *there is no IIM* M *that* $Sdr^*Txt_{\mathcal{H}}$*–identifies* $\mathcal{C}$.

For conservative learning and set-driven inference without anomalies, the analogue of Theorem 6 holds, as well (cf. [14, 16]).

Next, we study behaviorally correct identification. As we will see, finite telltale sets form a conceptual basis that is also well-suited to characterize the

collection of all Bc^*Txt–identifiable indexable classes. Surprisingly, the existence of the corresponding tell-tale sets is still sufficient.

Theorem 7. *For all $\mathcal{C} \in \mathcal{IC}$: $\mathcal{C} \in Bc^*Txt$ iff there is an indexing $(c_j)_{j\in\mathbb{N}}$ of $\mathcal{C}$ and a family $(T_j)_{j\in\mathbb{N}}$ of finite sets such that*

(1) *for all $j \in \mathbb{N}$, $T_j \subseteq c_j$,*
(2) *for all $j, k \in \mathbb{N}$, if $T_j \subseteq c_k \subseteq c_j$, then $c_k =^* c_j$.*

Proof. Due to the space constraint we sketch the sufficiency part, only. First, we define an appropriate hypothesis space $\mathcal{H} = (h_{\langle j,k\rangle})_{j,k\in\mathbb{N}}$. Let $(F_j)_{j\in\mathbb{N}}$ be an effective enumeration of all finite subsets of $\mathcal{X}$ and let $(w_j)_{j\in\mathbb{N}}$ be the lexicographically ordered enumeration of all elements in $\mathcal{X}$.

We subsequently use the following notions and notations. For all $c \subseteq \mathcal{X}$ and all $z \in \mathbb{N}$, we let $c^z = \{w_r \mid r \leq z,\ w_r \in c\}$. Moreover, for all $j, k, z \in \mathbb{N}$, we let $S_{(j,k,z)}$ be the set of all indices $r \leq k$ that meet (i) $F_j \subseteq c_r$ and (ii), for all $r' < r$ with $c_{r'} \supseteq F_j$, $c_r^z \subset c_{r'}^z$.

Now, we are ready to define the required hypothesis space $\mathcal{H}$. For all $j, k \in \mathbb{N}$ we define the characteristic function of $h_{\langle j,k\rangle}$ as follows. If $S_{(j,k,z)} = \emptyset$, we set $h_{\langle j,k\rangle}(w_z) = -$. If $S_{(j,k,z)} \neq \emptyset$, we let $n = max\, S_{(j,k,z)}$ and set $h_{\langle j,k\rangle}(w_z) = c_n(w_z)$.

Since membership is uniformly decidable in $(c_j)_{j\in\mathbb{N}}$, we know that $\mathcal{H}$ constitutes an admissible hypothesis space.

The required IIM M is defined as follows. Let $c \in \mathcal{C}$, $t \in Text(c)$, and $y \in \mathbb{N}$.

IIM M: "On input t_y proceed as follows:
Determine $j \in \mathbb{N}$ with $F_j = content(t_y)$ and output $\langle j, y\rangle$."

Due to lack of space, the verification of M'c correctness is omitted. □

Note that Baliga *et al.* [3] have recently shown that the same characterizing condition completely describes the collection of all indexable classes that are Bc^*Txt–identifiable with respect to *arbitrary* hypothesis spaces (including hypothesis space not having a decidable membership problem). Hence, our result refines the result from [3] in that it shows that, in order to Bc^*Txt–identify an indexable class, it is always possible to select a hypothesis space with uniformly decidable membership. However, as we see next, it is inevitable to select the actual hypothesis space appropriately.

Theorem 8. *There is an indexable class $\mathcal{C} \in Bc^*Txt$ such that, for all class preserving hypothesis spaces $\mathcal{H}$ for $\mathcal{C}$, there is no IIM M that $Bc^*Txt_{\mathcal{H}}$–learns $\mathcal{C}$.*

In contrast, $BcTxt$ is invariant to the choice of the hypothesis space.

To be complete, note that it is folklore that there are indexable classes which are not Bc^*Txt-identifiable. Furthermore, applying the stated characterizations of the learning types Fin^*Txt, Lim^*Txt, and Bc^*Txt, the following hierarchy can be shown.

Theorem 9. $Fin^*Txt \subset Lim^*Txt \subset Bc^*Txt \subset \mathcal{IC}$.

At the end of this subsection, we turn our attention to the case that the number of allowed anomalies is *a priori* bounded. On the one hand, Case and Lynes' [6] result that $Lim^{2a}Txt \subseteq Bc^aTxt$ nicely translates into our setting.

Surprisingly, the opposite is also true, i.e., every IIM that Bc^aTxt–identifies a target indexable class can be simulated by a learner that $Lim^{2a}Txt$–identifies the same class, as expressed by the following theorem.

Theorem 10. *For all $a \in \mathbb{N}$: $Bc^aTxt = Lim^{2a}Txt$.*

Proof. Let $a \in \mathbb{N}$. As mentioned above, $Lim^{2a}Txt \subseteq Bc^aTxt$ can be shown by adapting the corresponding ideas from [6] (see also [11], for the relevant details).

Next, we verify that $Bc^aTxt \subseteq Lim^{2a}Txt$. Let $\mathcal{C} \in Bc^aTxt$, let $\mathcal{H}$ be a hypothesis space, and let M be an IIM that $Bc^aTxt_{\mathcal{H}}$–identifies $\mathcal{C}$. Since membership is uniformly decidable in $\mathcal{H}$, the set $\{(j,k) \mid h_j \neq^{2a} h_k\}$ is recursively enumerable. Hence, without loss of generality, we may assume that there is a total recursive function $f\colon \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ such that $\{f(n) \mid n \in \mathbb{N}\} = \{(j,k) \mid h_j \neq^{2a} h_k\}$.

The required IIM M' also uses the hypothesis space $\mathcal{H}$. Let $c \in \mathcal{C}$, $t \in Text(c)$, and $y \in \mathbb{N}$.

IIM M': "On input t_y proceed as follows:
If $y = 0$, set $z = 0$, determine $j_0 = M(t_0)$, and output j_0. If $y \geq 1$, determine $j = M'(t_{y-1})$. For all $s = z, \ldots, y$, determine $j_s = M(t_s)$, and test whether or not $(j, j_s) \in \{f(n) \mid n \leq y\}$. In case there is no such pair, then output j. Otherwise, set $z = y$ and output j_y."

Since M $Bc^aTxt_{\mathcal{H}}$–identifies c from t, there has to be a least y such that, for all $y' \geq y$, $h_{M(t_{y'})} =^a c$, and therefore, for all $y', y'' \geq y$, $h_{M(t_{y'})} =^{2a} h_{M(t_{y''})}$. Hence, M' converges on t to a hypothesis j that meets $h_j =^{2a} c$. □

Applying Theorem 2, we may conclude:

Corollary 11. *For all $\mathcal{C} \in \mathcal{IC}$ and all $a \in \mathbb{N}$: $\mathcal{C} \in Bc^aTxt$ iff there is an indexing $(c_j)_{j\in\mathbb{N}}$ of $\mathcal{C}$ and a recursively enumerable family $(T_j)_{j\in\mathbb{N}}$ of finite sets such that*

(1) *for all $j \in \mathbb{N}$, $T_j \subseteq c_j$, and*
(2) *for all $j, k \in \mathbb{N}$, if $T_j \subseteq c_k$ and $c_k \subseteq c_j$, then $c_k =^{2a} c_j$.*

The latter corollary nicely contrasts the results in [3]. When arbitrary hypothesis spaces are admissible (including hypothesis space not having a decidable membership problem), there is no need to add any recursive component, i.e., the existence of the corresponding tell-tale sets is again sufficient.

Moreover, the relation between set-driven learners and conservative inference changes completely, if the number of allowed anomalies is *a priori* bounded.

Theorem 12. $Consv^1Txt \setminus \bigcup_{a\in\mathbb{N}} Sdr^aTxt \neq \emptyset$.

Theorem 13. *For all $a \in \mathbb{N}$: $Sdr^aTxt \subset Consv^aTxt$.*

The relation between conservative learners and unconstrained IIMs is also affected, if the number of allowed anomalies is *a priori* bounded.

Theorem 14. *For all $a \in \mathbb{N}$: $Lim^aTxt \subset Consv^{a+1}Txt \subset Lim^{a+1}Txt$.*

Proof. Let $a \in \mathbb{N}$. By definition, we get $Consv^{a+1}Txt \subseteq Lim^{a+1}Txt$. Moreover, $Consv^{a+1}Txt \setminus Lim^aTxt \neq \emptyset$ follows via Theorem 15 below. Furthermore, $Lim^{a+1}Txt \setminus Consv^{a+1}Txt \neq \emptyset$ can be shown by diagonalization.

It remains to show that $Lim^aTxt \subseteq Consv^{a+1}Txt$. To see this, recall the definition of the conservative IIM M' from the demonstration of Theorem 4. It

is easy to see that the final hypothesis of M' differs at most at one data point from the final hypothesis of the unconstrained IIM M which M' simulates. $\square$

Finally, when learning with an *a priori* bounded number of allowed anomalies is considered, the existence of infinite hierarchies of more and more powerful *Fin*-learners, *Consv*-learners, *Lim*-learners, and *Bc*-learners, parameterized in the number of allowed anomalies, can be shown. The following theorem provides the missing piece to establish these infinite hierarchies.

Theorem 15. *For all* $a \in \mathbb{N}$*:* $Fin^{2a+1}Txt \setminus Bc^{a}Txt \neq \emptyset$.

3.2 Incremental learning with anomalies

Proposition 3 summarizes the known results concerning incremental learning.

Proposition 3 ([15]).
(1) $ItTxt \subset FbTxt$.
(2) $ItTxt \subset Bem_1Txt$.
(3) *For all* $k \in \mathbb{N}$, $Bem_kTxt \subset Bem_{k+1}Txt$.
(4) $Bem_1Txt \setminus FbTxt \neq \emptyset$.
(5) $FbTxt \setminus \bigcup_{k\in\mathbb{N}} Bem_kTxt \neq \emptyset$.

The overall picture remains unchanged for incremental learning with a finite number of allowed anomalies.

More specifically, iterative learners that have the freedom to store *one* additional example may outperform feedback learners that are allowed to make up to finitely many errors in their final hypothesis.

Theorem 16. $Bem_1Txt \setminus Fb^{*}Txt \neq \emptyset$.

Proof. The separating class $\mathcal{C}$ is defined as follows. $\mathcal{C}$ contains $c_0 = \{a\}^+$ and, for all $j \geq 1$, $c_j = \{a^\ell \mid 1 \leq \ell \leq 2j\} \cup \{b\}^+$. Moreover, for all $j, k, m \geq 1$, $\mathcal{C}$ contains the concept $c'_{j,k,m} = \{a^\ell \mid 1 \leq \ell \leq 2j\} \cup \{a^{2\langle j,k\rangle+1}\} \cup \{b^\ell \mid 1 \leq \ell \leq m\}$.

Claim 1. $\mathcal{C} \in Bem_1Txt$.

The required IIM M updates its example-memory as follows. As long as no element from $\{b\}^+$ occurs, M memorizes the maximal element from $\{a\}^+$ seen so far. Otherwise, it memorizes the maximal element from $\{b\}^+$ that has been presented so far. In addition, M updates its hypotheses in accordance with the following cases.

Case 1. *M has never received an element from* $\{b\}^+$.

Then, M guesses c_0.

Case 2. *M receives an element x from* $\{b\}^+$ *for the first time.*

Let $x = b^m$. If M has memorized an element of type a^{2j}, M guesses c_j. If it has memorized an element of type $a^{2\langle j,k\rangle+1}$, M guess $c'_{j,k,m}$. If x is the first element presented at all, M simply guesses c_1.

Case 3. *Otherwise.*

Let x be the new element presented, let c' be M's actual guess, and let b^m be the element memorized by M.

First, if $x \in \{b\}^+$ and c' is of type $c'_{j,k,m}$, M guesses $c'_{j,k,m'}$, where $m' = max\,\{m, |x|\}$. If $x \in \{b\}^+$ and c' is of type c_j, M guesses c'.

Second, if $x \in \{a\}^+$ and $x \in c'$, M guesses c'. If $x \in \{a\}^+$, $x \notin c'$, and x is of type a^{2j}, M guesses c_j. Otherwise, i.e., $x \in \{a\}^+$, $x \notin c'$, and x is of type $a^{2\langle j,k\rangle+1}$, M guesses $c'_{j,k,m}$

The verification of M's correctness is straightforward.

Claim 2. $\mathcal{C} \notin Fb^*Txt$.

Suppose to the contrary that there is a feedback learner M' that witnesses $\mathcal{C} \in Lim^*Txt$. Hence, there is a locking sequence σ for c_0, i.e., σ is a finite sequence with $content(\sigma) \subseteq c_0$ and, for all finite sequences ρ with $content(\rho) \subseteq c_0$, $M'_*(\sigma \diamond \rho) = M'_*(\sigma)$.

Let j be the least index with $content(\sigma) \subseteq c_j$. Consider M when fed the text $t = \sigma \diamond a, \ldots, a^{2j} \diamond b \diamond b, b^2 \diamond b, b^2, b^3 \diamond \cdots \diamond b, b^2, \ldots, b^n \diamond \cdots$ for c_j. Since M' learns c_j, M' converges on t. Hence, there is a y such that (i) the last element in t_y equals b and (ii), for all $r \in \mathbb{N}$, $M'_*(t_y) = M'_*(t_{y+r})$.

Finally, fix τ such that $t_y = \sigma \diamond a, \ldots, a^{2j} \diamond \tau$. Let k, m be the least indices such that $content(t_y) \subseteq c'_{j,k,m}$ and $a^{2\langle j,k\rangle+1}$ is an element from c_0 which M' has never asked for when processing t_y. Consider M' when fed the text $t' = \sigma \diamond a, \ldots, a^{2j} \diamond a^{2\langle j,k\rangle+1} \diamond \tau \diamond b, b, \ldots$ for $c'_{j,k,m}$. By the choice of σ und y, M' converges on t and t' to the same hypothesis. (To see this note that the b's at the end of t' guarantee that M' almost always ask the same question as in case it is fed t_y, thereby, due to choice of $a^{2\langle j,k\rangle+1}$, always receiving the same answer.) Since $c_j \neq^* c'_{j,k,m}$, M' cannot learn both concepts, a contradiction. □

The opposite holds, as well. Feedback queries may compensate the ability of a bounded-example memory learner to memorize any *a priori* fixed number of examples *and* to make finitely many errors in its final hypothesis.

Theorem 17. $FbTxt \setminus \bigcup_{k\in\mathbb{N}} Bem_k^*Txt \neq \emptyset$

Proof. We define the separating class $\mathcal{C}$ as follows. We set $\mathcal{C} = \bigcup_{k\in\mathbb{N}} \mathcal{C}_k$, where, for all $k \in \mathbb{N}$, the subclass $\mathcal{C}_k$ is defined as follows.

Let $(F_j)_{j\in\mathbb{N}}$ be a repetition-free enumeration of all finite sets of natural numbers. By convention, let $F_0 = \emptyset$. Moreover, we let $P_0 = \{b\}^+$ and $P_{j+1} = P_j \setminus \{b^{np_j} \mid n \geq 1\}$, where, for all $j \in \mathbb{N}$, p_j is the $j+1$-st prime number.

Let $k \in \mathbb{N}$. Then, $\mathcal{C}_k$ contains the concept $c_0 = \{a\}^+$ as well as, for all $j, m \geq 1$ and all $l_0, \ldots, l_k$ with $j < l_0 < \cdots < l_k$, the concept $c_{(j,m,l_0,\ldots,l_k)} = \{a^\ell \mid 1 \leq \ell \leq j\} \cup \{a^{l_0}, \ldots, a^{l_k}\} \cup \{b^{j+1} \mid j \in F_m\} \cup P_{\langle l_0,\ldots,l_k\rangle} \cup \{d^j\}$.

By definition, $\mathcal{C}$ contains exclusively infinite concepts, and thus $\mathcal{C} \in FbTxt$ (cf. [8], for the relevant details).

For proving $\mathcal{C} \notin \bigcup_{k\in\mathbb{N}} Bem_k^*Txt$, it suffices to show that, for every $k \in \mathbb{N}$, $\mathcal{C}_k \notin Bem_k^*Txt$. The corresponding verification is part of the demonstration of Theorem 18 below. □

Our next result illustrates the error-correcting power of bounded example-memories. As it turns out, every additional example which an incremental learner can memorize may help to correct up to finitely many errors.

Theorem 18. *For all* $k \in \mathbb{N}$, $Bem_{k+1}Txt \setminus Bem_k^*Txt \neq \emptyset$.

Proof. Let $k \in \mathbb{N}$. We claim that $\mathcal{C}_k$ (cf. the demonstration of Theorem 17 above) separates the learning types $Bem_{k+1}Txt$ and Bem_k^*Txt.

Claim 1. $\mathcal{C}_k \in Bem_{k+1}Txt$.

The required bounded example-memory learner M behaves as follows. As a rule, M memorizes the $k+1$ longest elements from $\{a\}^+$ which it has seen so far. Moreover, M updates its hypotheses in accordance with the following cases.

Case 1. *M has never received an element from $\{d\}^+$.*

Then, M outputs an index for the concept c_0 that allows M to determine all elements from $\{b\}^+$ that have been presented so far.

Case 2. *M receives an element x from $\{d\}^+$ for the first time.*

Let $x = d^j$ and let S' be the set of all elements from $\{b\}^+$ seen so far. M outputs an index for the concept $\{a^\ell \mid 1 \leq \ell \leq j\} \cup \{d^j\} \cup S'$ that allows M to determine the elements in S'.

Case 3. *Otherwise.*

We distinguish the following subcases.

Case 3.1. *M has memorized $k+1$ elements s with $|s| > j$.*

Let x be the new element presented, let $S = \{a^{l_0}, \ldots, a^{l_k}\}$ be the set of elements memorized by M, and let S' be the set of elements from $\{b^+\}$ that are encoded in M's last hypothesis. If $x \in \{b\}^+ \setminus P_{\langle l_0,\ldots,l_k\rangle}$, we let $S' = S' \cup \{x\}$. Otherwise, S' remains unchanged. Moreover, M outputs an index for the concept $\{a^\ell \mid 1 \leq \ell \leq j\} \cup S \cup S' \cup P_{\langle l_0,\ldots,l_k\rangle} \cup \{d^j\}$ that allows M to recompute the elements in S'.

Case 3.2. *Not Case* 3.1.

As above, M outputs an index of the concept $\{a^\ell \mid 1 \leq \ell \leq j\} \cup \{d^j\} \cup S'$ that allows M to determine the elements in S', where S' is again the set of all elements from $\{b\}^+$ seen so far.

The verification of M's correctness is straightforward.

Claim 2. $\mathcal{C}_k \notin Bem_k^*Txt$.

Suppose to the contrary that there is a k-bounded example-memory learner M' that witnesses $\mathcal{C} \in Lim^*Txt$. Hence, there is a locking sequence σ for c_0, i.e., σ is a finite sequence with $content(\sigma) \subseteq c_0$ and, for all finite sequences ρ with $content(\rho) \subseteq c_0$, $\pi_1(M'_*(\sigma \diamond \rho)) = \pi_1(M'_*(\sigma))$.[1] Now let $j = max\,\{|x| \mid x \in content(\sigma)\}$. Similarly as in the demonstration of Theorem 6 in [15], one may use counting arguments to show that there are indices $l_0, l'_0, \ldots, l_k, l'_k$ such that Conditions (a) to (d) are fulfilled, where

(a) $j < l_0 < l_1 < \cdots < l_k$.
(b) $j < l'_0 < l'_1 < \cdots < l'_k$.
(c) $\{l_0, l_1, \ldots, l_k\} \neq \{l'_0, l'_1, \ldots, l'_k\}$.
(d) $M'_*(\sigma \diamond a^{l_0}, \ldots, a^{l_k}) = M'_*(\sigma \diamond a^{l'_0}, \ldots, a^{l'_k})$.

Assume that $\langle l_0, \ldots, l_k\rangle < \langle l'_0, \ldots, l'_k\rangle$. Let t_1 and t'_1 be the lexicographically ordered text for $P_{\langle l_0,\ldots,l_k\rangle}$ and $P_{\langle l'_0,\ldots,l'_k\rangle}$, respectively. Moreover, we set $\sigma' = \sigma \diamond a, a^2, \ldots, a^j$. Since M' infers $c_{(j,0,l_0,\ldots,l_k)}$, there is a finite sequence τ with $content(\tau) \subseteq P_{\langle l_0,\ldots,l_k\rangle}$ such that, for all finite sequences ρ with $content(\rho) \subseteq P_{\langle l_0,\ldots,l_k\rangle}$, $\pi_1(M'_*(\sigma' \diamond a^{l_0}, \ldots, a^{l_k} \diamond d^j \diamond \tau)) = \pi_1(M'_*(\sigma' \diamond a^{l_0}, \ldots, a^{l_k} \diamond d^j \diamond \tau \diamond \rho))$.

[1] Recall that M outputs pairs $\langle j, S\rangle$. By convention, we let $\pi_1(\langle j, S\rangle) = j$.

Now, fix $m' \in \mathbb{N}$ with $F_{m'} = \{\ell \mid b^\ell \in content(\tau)\}$ and consider M' when successively fed the text $t = \sigma' \diamond a^{l_0}, a^{l_1}, \ldots, a^{l_k} \diamond d^j \diamond \tau \diamond t_1$ for $c_{(j,0,l_0,\ldots,l_k)}$ and the text $t' = \sigma' \diamond a^{l'_0}, a^{l'_1}, \ldots, a^{l'_k} \diamond d^j \diamond \tau \diamond t'_1$ for $c_{(j,m',l'_0,\ldots,l'_k)}$, respectively. By the choice of σ and τ and since, by definition, $P_{\langle l'_0,\ldots,l'_k\rangle} \subset P_{\langle l_0,\ldots,l_k\rangle}$, we may conclude that M' converges to the same hypothesis when fed t and t', respectively. Since $c_{(j,0,l_0,\ldots,l_k)} \neq^* c_{(j,m',l'_0,\ldots,l'_k)}$, M' cannot learn both concepts, a contradiction. □

For incremental learning with anomalies, Proposition 3 rewrites as follows.

Corollary 19.
(1) $It^*Txt \subset Fb^*Txt$.
(2) $It^*Txt \subset Bem_1^*Txt$.
(3) *For all* $k \in \mathbb{N}$, $Bem_k^*Txt \subset Bem_{k+1}^*Txt$.
(4) $Bem_1^*Txt \setminus Fb^*Txt \neq \emptyset$.
(5) $Fb^*Txt \setminus \bigcup_{k\in\mathbb{N}} Bem_k^*Txt \neq \emptyset$.

4 Learning from positive and negative data

In the section, we briefly summarize the results that can be obtained when studying learning with anomalies from positive and negative data.

Let $\mathcal{X}$ be the underlying learning domain, let $c \subseteq \mathcal{X}$ be a concept, and let $i = ((x_n, b_n))_{n\in\mathbb{N}}$ be any sequence of elements of $\mathcal{X} \times \{+,-\}$ such that $content(i) = \{x_n \mid n \in \mathbb{N}\} = \mathcal{X}$, $content^+(i) = \{x_n \mid n \in \mathbb{N},\, b_n = +\} = c$ and $content^-(i) = \{x_n \mid n \in \mathbb{N},\, b_n = -\} = \mathcal{X} \setminus c = \overline{c}$. Then, we refer to i as an informant. By $Info(c)$ we denote the set of all informants for c.

For all $a \in \mathbb{N} \cup \{*\}$, the standard learning models Fin^aInf, $Consv^aInf$, Lim^aInf and Bc^aInf are defined analogously as their text counterparts by replacing text by informant. Moreover, we extend the definitions of all variants of iterative learning in the same manner and denote the resulting learning types by It^aInf, Fb^aInf, and Bem_k^aInf, where $k \in \mathbb{N}$.

Since $\mathcal{IC} = ConsvInf$ (cf. [10]), we may easily conclude:

Corollary 20.
For all $a \in \mathbb{N} \cup \{*\}$: $ConsvInf = Consv^aInf = Lim^aInf = Bc^aInf$.

Moreover, one can easily show that the known inclusions $FinTxt \subset FinInf$ and $FinInf \subset ConsvTxt$ (cf. [13]) rewrite as follows:

Theorem 21. $Fin^*Txt \subset Fin^*Inf \subset Consv^*Txt$.

Concerning incremental learning, it has recently be shown that $\mathcal{IC} = FbInf = Bem_1Inf$ (cf. [12]). Clearly this allows for the following corollary.

Corollary 22. *For all* $a \in \mathbb{N} \cup \{*\}$: $ConsvInf = Fb^aInf = Bem_1^aInf$.

Moreover, it is folklore that $\mathcal{IC} = It^*Inf$. In contrast, if the number of allowed anomalies is *a priori* bounded, an infinite hierarchy of more and more powerful iterative learners can be observed.

Theorem 23. $ItInf \subset It^1Inf \subset It^2Inf \subset \cdots \subset It^*Inf = ConsvInf$.

Finally, it is not hard to verify that the results obtained so far prove the existence of an infinite hierarchy of more and more powerful finite learners parameterized in the number of allowed anomalies.

References

1. D. Angluin. Finding patterns common to a set of strings. *Journal of Computer and System Sciences*, 21:46–62, 1980.
2. D. Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.
3. G.R. Baliga, J. Case, and S. Jain. The synthesis of language learners. *Information and Computation*, 152:16–43, 1999.
4. J. Bārzdiņš. Two theorems on the limiting synthesis of functions. In *Theory of Algorithms and Programs Vol. 1*, pages 82–88, Latvian State University, 1974, (Russian).
5. L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:122–155, 1975.
6. J. Case and C. Lynes. Machine inductive inference and language identification. In *Proc. 9th International Colloquium on Automata, Languages and Programming*, Lecture Notes in Computer Science 140, pages 107–115. Springer-Verlag, Berlin, 1982.
7. J. Case and C.H. Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science* 25:193–220, 1983.
8. J. Case, S. Jain, S. Lange, and T. Zeugmann, Incremental concept learning for bounded data mining. *Information and Computation* 152:74–110, 1999.
9. M. Fulk. Prudence and other restrictions in formal language learning. *Information and Computation*, 85:1–11, 1990.
10. E.M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.
11. S. Jain, D. Osherson, J. Royer, and A. Sharma. *Systems that Learn - 2nd Edition, An Introduction to Learning Theory.* MIT Press, Cambridge, Mass., 1999.
12. S. Lange and G. Grieser. On the strength of incremental learning. In *Proc. 10th International Conference on Algorithmic Learning Theory*, Lecture Notes in Artificial Intelligence 1720, pages 118–131. Springer-Verlag, Berlin, 1999.
13. S. Lange and T. Zeugmann. Types of monotonic language learning and their characterization. In *Proc. 5th Annual ACM Workshop on Computational Learning Theory*, pages 377–390. ACM Press, New York, 1992.
14. S. Lange and T. Zeugmann. Language learning in dependence on the space of hypotheses. In *Proc. 6th Annual ACM Conference on Computational Learning Theory*, pages 127–136. ACM Press, New York, 1993.
15. S. Lange and T. Zeugmann. Incremental learning from positive data. *Journal of Computer and System Sciences*, 53:88–103, 1996.
16. S. Lange and T. Zeugmann. Set-driven and rearrangement-independent learning of recursive languages. *Mathematical Systems Theory*, 29:599–634, 1996.
17. T. Tabe and T. Zeugmann. Two variations of inductive inference of languages from positive data. Technical Report RIFIS-TR-CS-105, Kyushu University, 1995.
18. K. Wexler and P. Culicover. *Formal Principles of Language Acquisition.* MIT Press, Cambridge, Mass., 1980.
19. R. Wiehagen. Limes-Erkennung rekursiver Funktionen durch spezielle Strategien. *Journal of Information Processing and Cybernetics (EIK)*, 12:93–99, 1976.
20. T. Zeugmann and S. Lange. A guided tour across the boundaries of learning recursive languages. In K.P. Jantke and S. Lange, editors, *Algorithmic Learning for Knowledge-Based Systems*, Lecture Notes in Artificial Intelligence 961, pages 190–258. Springer-Verlag, Berlin, 1995.