数理解析研究所講究録
906 巻 1995 年 73-80

Two Variations of Inductive Inference of Languages from Positive Data

Takashi Tabe (田部 尚志) Thomas Zeugmann

Research Institute of Fundamental Information Science,
Kyushu University 33, Fukuoka 812, Japan
e-mail: {tabe, thomas} @rifis.kyushu-u.ac.jp

Abstract

The present paper deals with the learnability of indexed families of uniformly recursive languages by single inductive inference machines (abbr. IIM) and teams of IIMs from positive and both positive and negative data. We study the learning power of single IIMs in dependence on the hypothesis space and the number of anomalies the synthesized language may have. Our results are fourfold. First, we show that allowing anomalies does not increase the learning power as long as inference from positive and negative data is considered. Second, we establish an infinite hierarchy in the number of allowed anomalies for learning from positive data. Third, we prove that every learnable indexed family $\mathcal{L}$ may be even inferred with respect to the hypothesis space $\mathcal{L}$ itself. Fourth, we characterize learning with anomalies from positive data.

Finally, we investigate the error correcting power of team learners, and relate the inference capabilities of teams in dependence on their size to one another. Again, an infinite hierarchy is established and the learnability is characterized in terms of recursively generable families of finite and non-empty sets.

1 Introduction

Inductive inference is the process of hypothesizing a general rule from eventually incomplete data. Within the last three decades it received much attention from computer scientists.

The present paper deals with inductive inference of formal languages. Looking at potential applications, Angluin [1] started the systematic study of learning enumerable families of uniformly recursive languages, henceforth called *indexed families*. First we consider about inferability of indexed family from *text* with anomalies. A *text* of a language L is an infinite sequence of strings that eventually contains all strings of L. Since a text contains exclusively positive examples concerning the target language, we sometimes refer to text as to *positive data*.

An algorithmic learner, henceforth called *inductive inference machine* (abbr. IIM), takes as input initial segments of a text, and outputs, from time to time, a hypothesis about the target language. The set $\mathcal{G}$ of all admissible hypotheses is called *hypothesis space*. Furthermore, the sequence of hypotheses has to converge to a hypothesis *approximately describing* the target language. That is, the cardinality of the symmetric difference of the target language and the language generated by the hypothesis the IIM converges to is required to be bounded by some *a priori* fixed number or to be finite, respectively. Hence, the hypothesis synthesized in the limit is allowed to contain *anomalies* with respect to the target language. Therefore, we synonymously refer to approximate inference as to learning with anomalies. If the number of allowed anomalies is equal to zero, then we just arrive at Gold's [6] classical definition of learning in the limit (cf. Definition 1).

Approximate inference has been introduced by Blum and Blum [2] in the context of learning recursive functions. Subsequently, this topic has been studied by various authors (cf., e.g., Case and Smith [4], Kinber and Zeugmann [10]). The study of language learning with anomalies goes back to Case and Lynes [3] (cf. Osherson, Stob and Weinstein [14] for further information). However, the present paper is the first one dealing with the inferability of indexed families when anomalies are allowed.

Moreover, we study the learnability of indexed families by teams of IIMs. In this setting, originally introduced by Smith [17], the learning task has to be realized by a finite collection of IIMs called team. The number n of IIMs in the collection is referred to as team size. Every team member is receiving the same information, i.e., it is successively fed a text or informant of the target language, respectively. However, the learning task is successfully finished if at least m, $m \leq n$, of the team members learn the target language (cf. Definition 3).

We study the learnability of approximate and team inference in dependence on the set of admissible hypothesis spaces, the number of allowed anomalies, and the success ratio m/n of teams, respectively.

The results obtained are manifold. First, we show that allowing anomalies does not increase the learning power as long as inference from positive and negative data is considered. Second, we establish an infinite hierarchy in the number of allowed anomalies for learning from positive data. Moreover, we show that every approximately learnable indexed family $\mathcal{L}$ may be even properly inferred thereby maintaining the number of allowed anomalies. The latter result is obtained via a characterization of learning with anomalies. Then we investigate the error correcting power of team learners, and relate the inference capabilities of teams in dependence on their size to one another. Again, an infinite hierarchy is established.

2 Preliminaries

Let $\mathbb{N} = \{0, 1, 2, ...\}$ be the set of all natural numbers. We set $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$. By Σ we denote any fixed finite alphabet of symbols. Let Σ^+ be the set of all non-empty strings over Σ. Every subset $L \subseteq \Sigma^+$ is called a language. Let L be a language, then we use $|L|$ to denote the cardinality of L. Furthermore, let L and $\hat{L}$ be any two languages, and let $a \in \mathbb{N}$; then we write $L =_a \hat{L}$ iff $|L \triangle \hat{L}| \leq a$. Here $\triangle$ denotes the symmetric difference of L and $\hat{L}$, i.e., $L \triangle \hat{L} = (L \setminus \hat{L}) \cup (\hat{L} \setminus L)$. Finally, we write $L =_* \hat{L}$ iff $|L \triangle \hat{L}|$ is finite (abbr. $|L \triangle \hat{L}| \leq *$).

Let L be a language and let $t = s_0, s_1, s_2, ...$ be an infinite sequence of strings from Σ^+ such that $range(t) = \{s_k \mid k \in \mathbb{N}\} = L$. Then t is said to be a ***text*** for L or, synonymously, a ***positive presentation***. Let L be a language. By $text(L)$ we denote the set of all positive presentations of L. Let L be a language. Moreover, let t be a text, and let x be a number. Then t_x denote the initial segment of t of length $x+1$. Let t be a text and let $x \in \mathbb{N}$. Then we define $t_x^+ = \{s_k \mid k \leq x\}$. Following Angluin [1], we restrict ourselves to deal exclusively with indexed families of uniformly recursive languages defined as follows: A sequence $L_0, L_1, L_2, ...$ is said to be an ***indexed family*** $\mathcal{L}$ of uniformly recursive languages provided all L_j are non-empty and there is a recursive function f such that for all numbers j and all strings $s \in \Sigma^*$ we have

$$f(j, s) = \begin{cases} 1, & \text{if } s \in L_j, \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, we set $range(\mathcal{L}) = \{L_j \mid j \in \mathbb{N}\}$ for every indexed family $\mathcal{L} = (L_j)_{j \in \mathbb{N}}$.

As in Gold [6], we define an ***inductive inference machine*** (abbr. IIM) to be an algorithmic device which works as follows: The IIM takes as its input larger and larger initial segments of a text t and it either requests the next input, or it first outputs a hypothesis, i.e., a number encoding a certain computer program, and then it requests the next input (cf., e.g., Angluin [1]).

At this point we have to clarify what hypothesis space we should choose, thereby also specifying the goal of the learning process. Gold [6] and Wiehagen [18] pointed out that there is a difference in what can be inferred depending on whether we want to synthesize in the limit grammars (i.e., procedures generating languages) or decision procedures, i.e., programs of characteristic functions. Case and Lynes [3] investigated this phenomenon in detail. However, in the context of identification of indexed families, both concepts are of equal power. Since we exclusively deal with the learnability of indexed families $\mathcal{L} = (L_j)_{j \in \mathbb{N}}$ we always take as the hypothesis space an enumerable family of grammars $\mathcal{G} = G_0, G_1, G_2, ...$ over the terminal alphabet Σ such that membership in $L(G_j)$ is uniformly decidable for all $j \in \mathbb{N}$ and all strings $s \in \Sigma^+$. For notational convenience we use $\mathcal{L}(\mathcal{G})$ to denote $(L(G_j))_{j \in \mathbb{N}}$. Note that $\mathcal{L}(\mathcal{G})$ constitutes itself an indexed family for all hypothesis spaces $\mathcal{G}$. When an IIM outputs a number j, we interpret it to mean that the machine is hypothesizing the grammar G_j. Let $\mathcal{L}$ be an indexed family, and let $a \in \mathbb{N} \cup \{*\}$; a hypothesis space $\mathcal{G} = (G_j)_{j \in \mathbb{N}}$ is said to be ***class admissible*** (***class preserving***) for $\mathcal{L}$ with respect to a provided that for every $L \in range(\mathcal{L})$ there exists an index j such that $L =_a L(G_j)$ ($range(\mathcal{L}) = range(\mathcal{L}(\mathcal{G}))$).

Let σ be a text, and $x \in \mathbb{N}$. Then we use $M(\sigma_x)$ to denote the last hypothesis produced by M when successively fed σ_x. The sequence $(M(\sigma_x))_{x \in \mathbb{N}}$ is said to ***converge in the limit*** to the number j if and only if either $(M(\sigma_x))_{x \in \mathbb{N}}$ is infinite and all but finitely many terms of it are equal to j, or $(M(\sigma_x))_{x \in \mathbb{N}}$ is non-empty and finite, and its last term is j. Now we are ready to define learning in the limit.

Definition 1. (Gold [6]) *Let $\mathcal{L}$ be an indexed family, let L be a language, and let $\mathcal{G} = (G_j)_{j \in \mathbb{N}}$ be a hypothesis space.* ***An IIM M CLIM-TXT-infers L from text with respect to $\mathcal{G}$*** *iff for every text t for L, there exists a $j \in \mathbb{N}$ such that the sequence $(M(t_x))_{x \in \mathbb{N}}$ converges in the limit to j and $L = L(G_j)$.*

Furthermore, M CLIM-TXT-identifies $\mathcal{L}$ with respect to $\mathcal{G}$ iff, for each $L \in range(\mathcal{L})$, M CLIM-TXT-identifies L with respect to $\mathcal{G}$.

Finally, let CLIM-TXT denote the collection of all indexed families $\mathcal{L}$ for which there are an IIM M and *a hypothesis space $\mathcal{G}$ such that M CLIM-TXT-identifies $\mathcal{L}$ with respect to $\mathcal{G}$.*

Since, by the definition of convergence, only finitely many data of L were seen by the IIM up to the (unknown) point of convergence, whenever an IIM identifies the language L, some form of learning must have taken place. For this reason, hereinafter the terms *infer*, *learn*, and *identify* are used interchangeably.

In Definition 1, LIM stands for "limit." Furthermore, the prefix C is used to indicate ***class admissible*** learning, i.e., the fact that $\mathcal{L}$ may be learned with respect to some suitably chosen class admissible hypothesis space. The restriction of $CLIM$ to ***class preserving*** inference is denoted by LIM. That means LIM is the collection of all indexed families $\mathcal{L}$ that can be learned in the limit with respect to a suitably chosen class preserving hypothesis space. Moreover, if a target indexed family $\mathcal{L}$ has to be inferred with respect to the hypothesis space $\mathcal{L}$ itself, then we replace the prefix C by P, i.e., $PLIM$ is the collection of indexed families that can be ***properly*** learned in the limit. Note that proper learning is sometimes also referred to as exact learning (cf., e.g., Zeugmann and Lange [19]).

Proposition 1. (Zeugmann and Lange [11]) $PLIM\text{-}TXT = LIM\text{-}TXT = CLIM\text{-}TXT$,

Next, we generalize Definition 1 to learning in the limit with anomalies. That is, now the hypotheses an IIM converges to are only required to suitably approximating the target languages.

Definition 2. (Case and Lynes [3]) *Let $\mathcal{L}$ be an indexed family, let L be a language, let $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$ be a hypothesis space, and let $a \in \mathbb{N} \cup \{*\}$.* ***An IIM M $CLIM^a$-TXT-infers L from text with respect to $\mathcal{G}$*** *iff for every text t for L, there exists a $j \in \mathbb{N}$ such that the sequence $(M(t_x))_{x\in\mathbb{N}}$ converges in the limit to j and $L =_a L(G_j)$.*

Furthermore, M $CLIM^a$-TXT-identifies $\mathcal{L}$ with respect to $\mathcal{G}$ iff, for each $L \in range(\mathcal{L})$, M $CLIM^a$-TXT-identifies L with respect to $\mathcal{G}$.

Finally, let $CLIM^a$-TXT denote the collection of all indexed families $\mathcal{L}$ for which there are an IIM M and a hypothesis space $\mathcal{G}$ such that M $CLIM^a$-TXT-identifies $\mathcal{L}$ with respect to $\mathcal{G}$.

Obviously, $\lambda LIM^0\text{-}TXT = \lambda LIM\text{-}TXT$ for all $\lambda \in \{P, \varepsilon, C\}$.

Finally, we define learning in the limit by a team of IIMs. Team inference has been introduced by Smith [17] in the context of inferring recursive functions. Subsequently various authors have studied it intensively (cf., e.g., Pitt [15], Pitt and Smith [16], Jain and Sharma [8, 9]).

Definition 3. (Smith [17]) *Let $\mathcal{L}$ be an indexed family, let L be a language, let $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$ be a hypothesis space, and let $n, m \in \mathbb{N}^+$, $m \leq n$.* ***A team $(M_1,\ldots,M_n)$ of IIMs $(m,n)CLIM$-TXT-infers L from text with respect to $\mathcal{G}$*** *iff for every text t for L, there exist at least m team members $M_{k_1}, \cdots, M_{k_m}$ and indices $j_1, \cdots, j_m$ such that the corresponding sequences $(M_{k_1}(t_x))_{x\in\mathbb{N}}, \cdots, (M_{k_m}(t_x))_{x\in\mathbb{N}}$ converge in the limit to $j_1, \cdots, j_m$ and $L = L(G_{j_z})$ for all $1 \leq z \leq m$, respectively.*

Furthermore, $(M_1, \cdots, M_n)$ $(m,n)CLIM$-TXT -identifies $\mathcal{L}$ with respect to $\mathcal{G}$ iff, for each $L \in range(\mathcal{L})$, $(M_1, \cdots, M_n)$ $(m,n)CLIM$-TXT-identifies L with respect to $\mathcal{G}$.

Finally, let $(m,n)CLIM$-TXT denote the collection of all indexed families $\mathcal{L}$ for which there are a team $(M_1, \cdots, M_n)$ of IIMs and a hypothesis space $\mathcal{G}$ such that $(M_1, \cdots M_n)$ $(m,n)CLIM$-TXT-identifies $\mathcal{L}$ with respect to $\mathcal{G}$.

Recently, Meyer [12] extended Pitt's [15] unification results to the case of learning indexed families from positive data. In particular, she showed that, for every $p \in (0, 1]$ the power of probabilistic IIMs learning with probability p equals the power of (1,n)–team learning, where n is the unique integer such that $1/(n+1) < p \leq 1/n$. This result immediately allows the following conclusion.

Proposition 2. *For all $m, n \in \mathbb{N}^+$, $m \leq n$, we have:* $(m,n)CLIM\text{-}TXT = (1, \lfloor m/n \rfloor)CLIM\text{-}TXT$.

Hence, in the following it suffices to deal exclusively with $(1,n)CLIM\text{-}TXT$. Furthermore, the proof technique of Lange and Zeugmann [11] can be directly applied to relate the learning capabilities of proper, class preserving and class admissible team learning to one another.

Proposition 3. *For all $n \in \mathbb{N}^+$ we have:* $(1,n)PLIM\text{-}TXT = (1,n)LIM\text{-}TXT = (1,n)CLIM\text{-}TXT$.

3 Inferability with Anomalies from Text

We start our investigations by characterizing learning in the limit with anomalies in terms of finite tell-tales. The first such theorem goes back to Angluin [1] who characterized proper learning in the limit accordingly. In order to characterize learning in the limit with anomalies, we had to generalize Angluin's [1] definition of tell-tales as follows.

Definition 4. *Let $\mathcal{L} = (L_j)_{j\in\mathbb{N}}$ be an indexed family, let $a \in \mathbb{N} \cup \{*\}$, and let $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$ be a class preserving hypothesis space for $\mathcal{L}$. A set Q is said to be an a-tell-tale for $L \in range(\mathcal{L})$ with respect to $\mathcal{L}(\mathcal{G})$ provided Q satisfies the following conditions:*

(1) *Q is finite,*

(2) *$Q \subseteq L$, and*

(3) *for every $j \in \mathbb{N}$, if $Q \subseteq L(G_j) \subseteq L$, then $L(G_j) =_a L$.*

Note that the definition made above essentially coincides with Angluin's [1] definition of a tell-tale in case $a = 0$. Therefore, we refer to 0-tell-tales as to tell-tales for short.

Proposition 4. (Angluin [1]) *Let $\mathcal{L} = (L_j)_{j\in N}$ be an indexed family. Then following two conditions are equivalent.*

(1) *$\mathcal{L} \in PLIM\text{-}TXT$,*

(2) *there exists an effective procedure which, for every $j \in \mathbb{N}$, uniformly enumerates a tell-tale for L_j with respect to $\mathcal{L}$.*

Next, we want to extend this characterization theorem to identification in the limit with anomalies.

Theorem 1. *Let $\mathcal{L} = (L_j)_{j\in N}$ be an indexed family, let $a \in \mathbb{N} \cup \{*\}$, and let $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$ be a class preserving hypothesis space for $\mathcal{L}$. If there exists an effective procedure g which, for all $j \in \mathbb{N}$, uniformly enumerates a-tell-tales for $L(G_j)$ with respect to $\mathcal{L}(\mathcal{G})$, then $\mathcal{L}$ is LIM^a-TXT-inferable with respect to $\mathcal{G}$.*

Proof. Using just the same procedure as in Angluin [1], we can enumerate a-tell-tales. □

However, in generalizing Theorem 1 to class admissible learning with anomalies, and in proving the converse of Theorem 1, and its desired generalization we have to overcome some difficulties. The next lemma actually states that any IIM which $CLIM^a$-TXT–infers an indexed family $\mathcal{L}$ can be replaced by another one that converges on every text for every language $L \in range(\mathcal{L})$ to a superset of it, and that also witnesses $\mathcal{L} \in CLIM^a$-TXT.

Lemma 2. *Let $a \in \mathbb{N} \cup \{*\}$. Furthermore, let $\mathcal{L}$ be an indexed family, let $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$ be a hypothesis space and let M be an IIM witnessing $\mathcal{L} \in CLIM^a$-TXT with respect to $\mathcal{G}$. Then one can effectively construct a hypothesis space $\hat{\mathcal{G}} = (\hat{G}_j)_{j\in\mathbb{N}}$ and an IIM $\hat{M}$ such that*

(1) *$\hat{M}$ $CLIM^a$-TXT-infers $\mathcal{L}$ with respect to $\hat{\mathcal{G}}$, and*

(2) *for all $L \in range(\mathcal{L})$ and for all $t \in text(L)$, if $(\hat{M}(t_x))_{x\in\mathbb{N}}$ converges to k, then $L \subseteq L(\hat{G}_k)$.*

Proof. The desired hypothesis space $\hat{\mathcal{G}}$ is defined as follows. For all $j \in \mathbb{N}$, we set

$$\hat{G}_j = \begin{cases} G_{\frac{j}{2}}, & \text{if } j \text{ is even,} \\ \text{grammar of} L_{\frac{j-1}{2}}, & \text{if } j \text{ is odd.} \end{cases}$$

Furthermore, let $w_0, w_1, \ldots$ be any fixed effective enumeration of Σ^+. For every $A \subseteq \Sigma^+$ and $x \in \mathbb{N}$, we use $A^{(x)}$ to denote $A \cap \{w_0, w_1, \ldots, w_x\}$. Now we are ready to define the desired IIM $\hat{M}$.

IIM $\hat{M}$: "On input t_x, execute Stage x.

Stage x : Let $j_x = M(t_x)$. Search for the least $j \le x$ satisfying $(t_x^+)^{(x)} \subseteq L(\hat{G}_j)^{(x)} \subseteq L(G_{j_x}) \cup t_x^+$. If such a j is found, then output it, and request the next input.
Else output x, and request the next input."

It remains to show that $\hat{M}$ satisfies Properties (1) and (2). Let $L \in range(\mathcal{L})$, let $t \in text(L)$, and $x \in \mathbb{N}$. By assumption, M $CLIM^a$-TXT-infers L with respect to $\mathcal{G}$. Hence, there exist $\tilde{x}$ and m such that $M(t_x) = j_x = m$ for all $x \ge \tilde{x}$ and $L(G_m) =_a L$. Let $\tilde{\jmath}$ be the least j which satisfies $L \subseteq L(\hat{G}_j) \subseteq L(G_m) \cup L$. There exists such a $\tilde{\jmath}$ because $\mathcal{L}(\hat{\mathcal{G}})$ contains L. Inevitably $L =_a L(\hat{G}_{\tilde{\jmath}})$. It is straightforward to see $\hat{M}$ converges to $\tilde{\jmath}$. □

Now we are ready to establish both the desired generalization of Theorem 1 as well as its converse.

Theorem 3. *Let $a \in \mathbb{N} \cup \{*\}$, and let $\mathcal{L} = (L_j)_{j\in\mathbb{N}}$ be an indexed family. If $\mathcal{L}$ is $CLIM^a$-TXT-inferable, then there exists an effective procedure g which, for each $j \in \mathbb{N}$, uniformly enumerates an a-tell-tale for L_j with respect to $\mathcal{L}$.*

Proof. Let M be any IIM which $CLIM^a$-TXT-infers $\mathcal{L}$ with respect to $\mathcal{G}$. Let $\hat{M}$ and $\hat{\mathcal{G}}$ be chosen in accordance with Lemma 2. We show that the following procedure g uniformly enumerates a-tell-tales with respect to $\mathcal{L}$. Note that this procedure is just the same one Angluin [1] has used. For every $j \in \mathbb{N}$ let $w_0, w_1, w_2, \ldots$ be any fixed effective enumeration of L_j, and let $\tau_0, \tau_1, \tau_2, \ldots$ be any fixed effective enumeration of all finite sequences of elements of L_j.

Procedure g:

"On input $j \in \mathbb{N}$, do the following: Initialize $\sigma_0 = w_0$, output w_0. Execute Stage 0 (If the computation of $\hat{M}(w_0)$ halts without any output, consider that $\hat{M}(w_0)$ is not equal to any integers.)

Stage x, $x \in \mathbb{N}$ **:** For $y = 0, 1, 2, \cdots$, compute $\hat{M}(\sigma_0 \cdot \ldots \cdot \sigma_x \cdot \tau_y)$ until the first y is found such that $\hat{M}(\sigma_0 \cdot \ldots \cdot \sigma_x) \neq \hat{M}(\sigma_0 \cdot \ldots \cdot \sigma_x \cdot \tau_y)$. Then, let $\sigma_{x+1} = \tau_y \cdot w_{x+1}$, output all elements that occur in σ_{x+1}, and go to Stage $(x+1)$."

It is clear that this procedure is effective since $\hat{M}$ $CLIM^a$-TXT-infers $\mathcal{L}$. If this procedure executes infinitely many stages, then $t = \sigma_0 \cdot \sigma_1 \cdot \ldots$ becomes a text for L_j, because it is containing all and only the elements of L_j. However for this text t, the IIM $\hat{M}$ does not converge. Hence, the assumption is contradicted. Thus there exists an m such that for all τ_k, $\hat{M}(\sigma_0 \cdot \ldots \cdot \sigma_m \cdot \tau_k) = \hat{M}(\sigma_0 \cdot \ldots \cdot \sigma_m)$. Let $\ell = \hat{M}(\sigma_0 \cdot \ldots \cdot \sigma_m)$. Consequently, we may conclude that $L_j =_a L(\hat{G}_\ell)$ as well as $L_j \subseteq L(\hat{G}_\ell)$, since $\hat{M}$ is assumed to $CLIM^a$-TXT-infer $\mathcal{L}$ with respect to $\hat{\mathcal{G}}$ and in accordance with Lemma 2 it converges to supersets. Moreover, in accordance with the definition of Procedure g we see that the set Q_j enumerated on input j satisfies $Q_j = (\sigma_0 \cdot \ldots \cdot \sigma_m)^+$.

Next we prove that Q_j fulfills the Properties (1) through (3) of Definition 4. Obviously Q_j is finite and $Q_j \subseteq L_j$, thus Properties (1) and (2) are satisfied. In order to prove Property (3) assume any $L' \in range(\mathcal{L})$ that satisfies $Q_j \subseteq L' \subseteq L_j$. We have to show that $L' =_a L_j$. Let $\tilde{t}$ be any text for L'. Since $Q_j \subseteq L'$, we may directly conclude that $\sigma_0 \cdot \ldots \cdot \sigma_m \cdot \tilde{t}$ is a text for L', too. Furthermore, every initial segment $\tilde{t}_x$ of $\tilde{t}$ constitutes a finite sequence of elements from L_j, since $L' \subseteq L_j$. Therefore, the construction of the finite sequence $\sigma_0 \cdot \ldots \cdot \sigma_m$ ensures that the sequence $(\hat{M}((\sigma_0 \cdot \ldots \cdot \sigma_m \cdot \tilde{t})_x))_{x\in\mathbb{N}}$ converges to ℓ. Finally, since $\hat{M}$ $CLIM^a$-TXT-infers $\mathcal{L}$ with respect to $\hat{\mathcal{G}}$, and since $L' \in range(\mathcal{L})$, we get $L' =_a L(\hat{G}_\ell)$. Taking into account that $L' \subseteq L_j \subseteq L(\hat{G}_\ell)$ we can conclude that $L' \setminus L_j = \emptyset$ as well as $a \geq |L(\hat{G}_\ell) \setminus L'| \geq |L_j \setminus L'|$. Thus, $L' =_a L_j$ and Q_j is an a-tell-tale for L_j. This proves the theorem. □

Now, we can completely characterize inference with anomalies. This is done with the following proposition.

Proposition 5. *Let $a \in \mathbb{N} \cup \{*\}$ and $\mathcal{L} = (L_j)_{j\in N}$ be an index family. Then the following conditions are equivalent.*

(i) There exists an effective procedure which, for every $j \in \mathbb{N}$, uniformly enumerates an a-tell-tale for L_j with respect to $\mathcal{L}$.

(ii) $\mathcal{L} \in PLIM^a$-TXT.

(iii) $\mathcal{L} \in CLIM^a$-TXT.

Proof. By Theorem 1, (i) implies (ii). The implication $(ii) \to (iii)$ is clear by definition. By Theorem 3, we have $(iii) \to (i)$. □

The following corollaries follow immediately.

Corollary 4. *For all $a \in \mathbb{N}^+$ there exists an indexed family $\mathcal{L}_a$ such that $\mathcal{L}_a \in PLIM^a$-$TXT \backslash PLIM^{a-1}$-$TXT$.*

Proof. Let $a \in \mathbb{N}^+$, we define the desired indexed family $\mathcal{L}_a = (L_j)_{j\in\mathbb{N}}$ as follows. Set $L_0 = \Sigma^+$, and let $L_1, L_2, \ldots$ be the canonical enumeration of all languages obtained by removing just a strings from Σ^+. It is easy to see $\mathcal{L}_a \in PLIM^a$-$TXT \setminus PLIM^{a-1}$-TXT. □

Corollary 5.

$$\bigcup_{a\in\mathbb{N}} PLIM^a\text{-}TXT \subset PLIM^*\text{-}TXT$$

Proof. Let $\mathcal{L}_1, \mathcal{L}_2, \ldots$ be the indexed families defined above, and let $\mathcal{L}$ be the canonical enumeration of all the languages enumerated in the indexed families $\mathcal{L}_1, \mathcal{L}_2, \ldots$. It is easily shown that $\mathcal{L} \in PLIM^*\text{-}TXT \setminus \bigcup_{a\in\mathbb{N}} PLIM^a\text{-}TXT$. □

Corollary 6. *Let $\mathcal{L}$ be any super-finite indexed family, i.e., $\mathcal{L}$ involves all finite sets and at least one infinite language L. Then $\mathcal{L}$ is not $PLIM^*\text{-}TXT$-inferable.*

We finish this section with the following figure that summarizes the results obtained.

$$PLIM^0\text{-}TXT \subset PLIM^1\text{-}TXT \subset \cdots \subset \bigcup_{a\in\mathbb{N}} PLIM^a\text{-}TXT \subset PLIM^*\text{-}TXT$$

4 Inferability by a Team of IIMs from Text

The next theorem shows that a team of two IIMs has sometimes more learning power than every IIM learning with anomalies.

Theorem 7. *For all $a \in \mathbb{N}$, $(1,2)PLIM\text{-}TXT \setminus PLIM^a\text{-}TXT \neq \emptyset$.*

Moreover, the next theorem shows that teams of size $a+1$ can be used to correct at most a anomalies a single machine may make. Note that a similar result has been obtained by Daley [5] in case of learning recursive functions.

Theorem 8. *Let $a \in \mathbb{N}^+$. Then, $PLIM^a\text{-}TXT \subseteq (1, a+1)PLIM\text{-}TXT$.*

Furthermore, the number $a+1$ of team members used in the above theorem to correct a anomalies cannot be decreased, as we shall show (cf. Theorem 10). In order to prove this, we need a further generalization of the tell-tale concept which is provided by the next definition.

Definition 5. *Let $\mathcal{L}$ be an indexed family and $L \in range(\mathcal{L})$.*

A set Q is said to be a 0-depth tell-tale for L with respect to $\mathcal{L}$ if it satisfies the following conditions:

(i) Q is finite,

(ii) $Q \subseteq L$, and

(iii) no $L' \in range(\mathcal{L})$ exists such that $Q \subseteq L' \subset L$.

(that is, Q is an ordinary tell-tale set for L).

We proceed inductively. Let $n \geq 1$. Then a set Q is said to be a n-depth tell-tale for L with respect to $\mathcal{L}$ if it satisfies following conditions:

(i) Q is finite,

(ii) $Q \subseteq L$ and

(iii) for all $\hat{L} \in range(\mathcal{L})$, $Q \subseteq \hat{L} \subset L$ implies the existence of an $(n-1)$-depth tell-tale for $\hat{L}$ with respect to $\mathcal{L}$.

Note that an n-depth tell-tale is a variation of *n-bounded finite tell-tales* introduced by Mukouchi [13]. The next lemma describes a necessary condition for team inference from positive data in terms of n-depth tell-tales.

Lemma 9. *Let $n \in \mathbb{N}$ and let $\mathcal{L}$ be any indexed family that is $(1, n+1)PLIM\text{-}TXT$–inferable. Then, for all $L \in range(\mathcal{L})$, there exists an n-depth tell-tale for L with respect to $\mathcal{L}$.*

Proof. Since $\mathcal{L} \in (1, n+1)PLIM\text{-}TXT$, there exists a team $(M_0, M_1, \ldots, M_n)$ of IIMs that $(1, n+1)PLIM\text{-}TXT$–infers $\mathcal{L}$. We continue with the proof of a technical claim that is very helpful in showing the existence of n-depth tell-tale. This claim describes that there exists a "locking sequence" (see, [14]) in case of team inference also.

Claim. Let $L' \in range(\mathcal{L})$ and let σ be an arbitrary finite sequence which satisfies $\sigma^+ \subseteq L'$. Then, there exist a finite sequence τ and $k \in \{0, 1, \ldots, n\}$ such that

(i) $\tau^+ \subseteq L'$, and

(ii) there exists j such that $M_k(\sigma \cdot \tau \cdot \psi) = j$ for every finite (maybe empty) sequence ψ with $\psi^+ \subseteq L'$ and $L' = L_j$.

Proof of Claim. Assume there does not exist a pair (τ, k) which satisfies Conditions (i) and (ii). Let $w_0, w_1, \ldots$ be a fixed effective enumeration of L'. We define finite sequences $\sigma_0, \sigma_1, \ldots$ inductively as follows:

(a) $\sigma_0 = w_0$

(b) $(\sigma \cdot \sigma_0 \cdot \ldots \cdot \sigma_{x-1})^+ \subseteq L'$ by inductive definition of $\sigma_0, \ldots, \sigma_{x-1}$. Since the team infers $\mathcal{L}$, there exists τ such that $\tau^+ \subseteq L'$ and $P(\tau) = \{(k,j) |\ M_k(\sigma \cdot \sigma_0 \cdot \sigma_1 \cdot \ldots \cdot \sigma_{x-1} \cdot \tau) = j \text{ and } L_j = L'\} \neq \emptyset$. Note that $|P(\tau)| \leq n+1$. Let $(k_1, j_1), \cdots, (k_r, j_r)$ be any enumeration of $P(\tau)$. By the assumption, a pair $(\sigma_0 \cdot \ldots \cdot \sigma_{x-1} \cdot \tau, k_1)$ does not satisfy (ii). Hence we may conclude that there is a finite sequence ψ_1 such that $\psi_1^+ \subseteq L'$ and $j_1 \neq M_{k_1}(\sigma \cdot \sigma_0 \cdot \sigma_1 \cdot \ldots \cdot \sigma_{x-1} \cdot \tau \cdot \psi_1)$. Since a pair $(\sigma_0 \cdot \ldots \cdot \sigma_{x-1} \cdot \tau \cdot \psi_1, k_1)$ does not satisfy (ii), there exists ψ_2 such that $\psi_2^+ \subseteq L'$ and $j_2 \neq M_{k_1}(\sigma \cdot \sigma_0 \cdot \sigma_1 \cdot \ldots \cdot \sigma_{x-1} \cdot \tau \cdot \psi_1 \cdot \psi_2)$. By iterating this construction, we effectively find a finite sequence $\psi = \psi_1 \cdot \psi_2 \cdot \ldots \cdot \psi_r$ which satisfies (i) $\psi^+ \subseteq L'$ and (ii) M_k outputs another number than j between $\sigma \cdot \sigma_0 \cdot \ldots \cdot \sigma_{x-1} \cdot \tau$ and $\sigma \cdot \sigma_0 \cdot \ldots \cdot \sigma_{x-1} \cdot \tau \cdot \psi$ for all $(k,j) \in P(\tau)$. Let $\sigma_x = \tau \cdot \psi \cdot w_x$.

$t = \sigma_0 \cdot \sigma_1 \cdot \ldots$ constitutes a text for L'. However the team can not infer L' from t by construction (infinitely many mind changes occur). This is a contradiction. Here, the claim is proved.

In order to finish the proof of the lemma, we have to show that every $L \in \mathit{range}(\mathcal{L})$ possesses an n-depth tell-tale. Assume L does not have any n-depth tell-tale. Consider the following $(n+1)$ stages.

Stage 0: By the claim, there exist a finite sequence τ_0 and $k_0 \leq n$ such that $\tau_0^+ \subseteq L$ and M_{k_0} is locked to j_0 with $L_{j_0} = L$ having τ_0. Since L does not have any n-depth tell-tale, there exists $L^1 \in \mathit{range}(\mathcal{L})$ which satisfies $\tau_0^+ \subseteq L^1 \subset L$ and L^1 does not have any $(n-1)$-depth tell-tale.

Stage x $(1 \leq x \leq n-1)$**:** Since $(\tau_0 \cdot \tau_1 \cdot \ldots \cdot \tau_{x-1})^+ \subseteq L^x$, by the claim there exist a finite sequence τ_x and $k_x \leq n$ such that $\tau_x^+ \subseteq L^x$ and M_{k_x} is locked to j_x with $L_{j_x} = L^x$ having $\tau_0 \cdot \ldots \cdot \tau_x$. Since L^x does not have any $(n-x)$-depth tell-tale, there exists $L^{x+1} \in \mathit{range}(\mathcal{L})$ which satisfies $(\tau_0 \cdot \ldots \cdot \tau_x)^+ \subseteq L^{x+1} \subset L^x$ and L^{x+1} does not have any $(n-x-1)$-depth tell-tale.

Stage n**:** Since $(\tau_0 \cdot \tau_1 \cdot \ldots \cdot \tau_{n-1})^+ \subseteq L^n$, by the claim there exist a finite sequence τ_n and $k_n \leq n$ such that $\tau_n^+ \subseteq L^n$ and M_{k_n} is locked to j_n with $L_{j_n} = L^n$ having $\tau_0 \cdot \ldots \cdot \tau_n$. Since L^n does not have any 0-depth tell-tale, there exists $L^{n+1} \in \mathit{range}(\mathcal{L})$ which satisfies $(\tau_0 \cdot \ldots \cdot \tau_n)^+ \subseteq L^{n+1} \subset L^n$.

In Stage n, $L^{n+1} \in \mathit{range}(\mathcal{L})$ is defined. Let t be a text for L^{n+1}. Then, $\tau_0 \cdot \ldots \cdot \tau_n \cdot t$ is also a text for L^{n+1}. However $M_{k_0}, M_{k_1}, \cdots, M_{k_n}$ are locked to $j_0, j_1, \cdots, j_n$ respectively, having $\tau_0 \cdot \ldots \cdot \tau_n$. And by the construction, $L^{n+1} \subset L_{j_n} \subset L_{j_{n-1}} \subset \cdots \subset L_{j_0}$. That is, all IIMs are locked to indices of languages which are not equal to L^{n+1} having $\tau_0 \cdot \ldots \cdot \tau_n$. Hence the team cannot infer L^{n+1} from its text $\tau_0 \cdot \ldots \cdot \tau_n \cdot t$, a contradiction. □

Theorem 10. *Let $a \in \mathbb{N}^+$. Then, $PLIM^a\text{-}TXT \setminus (1,a)PLIM\text{-}TXT \neq \emptyset$.*

Proof. For all $k \in \{0, 1, \ldots, a\}$, let $\mathcal{L}^{(k)}$ be the canonical enumeration of all sets obtained by removing just $(a-k)$ strings from Σ^+. And let $\mathcal{L}$ be the canonical enumeration of all languages in the indexed families $\mathcal{L}^{(0)}, \mathcal{L}^{(1)}, \ldots, \mathcal{L}^{(a)}$. Clearly, $\mathcal{L} \in PLIM^a\text{-}TXT$.

On the other hand, we can show $\Sigma^+ \in \mathit{range}(\mathcal{L})$ does not have an $(a-1)$-depth tell-tale by mathematical induction. Thus, by lemma 9, $\mathcal{L}$ is not $(1,a)PLIM\text{-}TXT$–inferable. □

$$
\begin{array}{ccc}
PLIM^0\text{-}TXT & = & (1,1)PLIM\text{-}TXT \\
\cap & & \cap \\
PLIM^1\text{-}TXT & \subset & (1,2)PLIM\text{-}TXT \\
\cap & & \cap \\
PLIM^2\text{-}TXT & \subset & (1,3)PLIM\text{-}TXT \\
\vdots & & \vdots
\end{array}
$$

5 Conclusion

We mainly studied the learnability of indexed families from positive data by IIMs that are allowed to converge to approximations as well as by teams. In particular, two new infinite hierarchies have been established.

The uniform recursive enumerability of n-depth tell-tales could only be proved to be necessary for learning by a team of n machines with success ratio $1/n$. On the other hand, it remained open whether or not this condition is sufficient, too.

6 References

[1] D. Angluin, Inductive inference of formal languages from positive data, *Information and Control* **45** (1980) 117–135.

[2] L. Blum and M. Blum, Toward a mathematical theory of inductive inference, *Information and Control* **28** (1975), 122–155.

[3] J. Case and C. Lynes, Machine inductive inference and language identification, in: *Proc. of the Ninth International Conference on Automata, Languages and Programming* Lecture Notes in Computer Science **140** (Springer, Berlin, 1982) 107–115.

[4] J. Case and C. H. Smith, Comparison of identification criteria for machine inductive inference, *Theoretical Computer Science* **25**, 193–220.

[5] R. P. Daley, On the error correcting power of pluralism in BC–type inductive inference, *Theoretical Computer Science* **24** (1983), 95 – 104.

[6] M. E. Gold, Language identification in the limit, *Information and Control* **10** (1967) 447–474.

[7] J. E. Hopcroft and J. D. Ullman, *Formal Languages and their Relation to Automata* (Addison-Wesley, Reading, Massachusetts, 1979).

[8] S. Jain and A. Sharma, Language Learning by a "Team", Lecture Notes in Computer Science **443**, (Springer, Berlin, 1989) 153–166.

[9] S. Jain and A. Sharma, Computational Limits on Team Identification of Languages.

[10] E. B. Kinber and T. Zeugmann, Inductive inference of almost everywhere correct programs by reliably working strategies, *Journal of Information Processing and Cybernetics* **21**, 91–100.

[11] S. Lange and T. Zeugmann, Learning recursive languages with bounded mind changes, *International Journal of Foundations of Computer Science* **4** (1993), 157–178.

[12] L. Meyer, personal communication

[13] Y. Mukouchi, Inductive inference of recursive concepts, PhD thesis, Kyushu University, Research Institute of Fundamental Information Science, Technical Report RIFIS-TR-CS 82, 1994.

[14] D. Osherson, M. Stob and S. Weinstein, *Systems that Learn, An Introduction to Learning Theory for Cognitive and Computer Scientists* (MIT-Press, Cambridge, Massachusetts, 1986).

[15] L. Pitt, A characterization of probabilistic inference, in: *Proc. of the 25th Annual Symposium on Foundations of Computer Science, Palm Beach, Florida*, 485–494.

[16] L. Pitt, C. H. Smith, Probability and Plurality for Aggregations of Learning Machines, *Information and Computation* **77** (1988), 77–92.

[17] C. H. Smith, The power of pluralism for automatic program synthesis, *Journal Association for Computing Machinery* **29** (1982), 1144 – 1165.

[18] R. Wiehagen, Identification of formal languages, in: *Proc. Mathematical Foundations of Computer Science* Lecture Notes in Computer Science **53** (Springer-Verlag, Berlin, 1977) 571–579.

[19] T. Zeugmann and S. Lange, A guided tour across the boundaries of learning recursive languages, TH Leipzig, Fachbereich Mathematik und Informatik, GOSLER–Report 26/94, 1994.