Two Variations of Inductive Inference of Languages from Positive Data

Takashi Tabe 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 allowed 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. Nowadays inductive inference can be considered as a form of machine learning with potential applications to artificial intelligence (cf., e.g., Angluin and Smith [3, 4], Osherson, Stob and Weinstein [38]). For more information concerning recent developments in inductive inference, the reader is referred to the annual Workshops on Computational Learning Theory, COLT (cf., e.g., Rivest et al. [42], Fulk and Case [15], and Haussler [18], the International Workshops on Algorithmic Learning Theory, ALT (cf., e.g., Arikawa et al. [5, 6, 7]), the International Workshops on Analogical and Inductive Inference, AII (cf., e.g., Jantke [22, 23], and Arikawa and Jantke [7]), and the European conference on Computational Learning Theory, EuroCOLT (cf. Shawe-Taylor and Anthony [44]).

The present paper deals with inductive inference of formal languages, a field in which many interesting and sometimes surprising results have been obtained (cf., e.g., Case and Lynes [11], Case [10], Fulk [14]). Looking at potential applications, Angluin [1, 2] started the systematic study of learning enumerable families of uniformly recursive languages, henceforth called *indexed families.* A sequence $L_0, L_1, L_2, \cdots$ of languages is said to be indexed family provided all L_j are non-empty and membership in L_j is uniformly decidable for all indices j. As a matter of fact, the definition of an indexed family contains both, a description for every enumerated language L_j and a particular enumeration of all the languages from its range. Recently, this topic has attracted much attention (cf., e.g., [24, 25, 27, 28, 29, 30, 31, 32, 33, 35, 36, 37, 43, 45, 50, 51]).

Next we specify the information from which the target languages have to be learned. 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.* Alternatively, one can consider learning from *informant.* An informant of a language L is an infinite sequence of all strings over the underlying alphabet that are classified with respect to their containment in L. Consequently, an informant contains both *positive and negative examples* concerning the language to be learned. Therefore, we sometimes refer to informants as to *positive and negative data.*

An algorithmic learner, henceforth called *inductive inference machine* (abbr. IIM), takes as input initial segments of a text (an informant), 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 there is an IIM that learns the target language L from all its texts (informants), then L is said to be approximately *learnable from text* (*learnable from informant*) in the limit with respect to the hypothesis space $\mathcal{G}$. An indexed family $\mathcal{L}$ is said to be learnable from text (learnable from informant) provided there is an IIM that learns every language contained in the range of $\mathcal{L}$ (cf. Definition 2). If the number of allowed anomalies is equal to zero, then we just arrive at Gold's [17] classical definition of learning in the limit (cf. Definition 1).

Approximate inference has been introduced by Blum and Blum [9] in the context of learning recursive functions. Subsequently, this topic has been studied by various authors (cf., e.g., Case and Smith [12], Kinber and Zeugmann [26]). The study of language learning with anomalies goes back to Case and Lynes [11] (cf. Osherson, Stob and Weinstein [38] 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 [47], 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.

Concerning the choice of the hypothesis space we distinguish between *proper* learnability, *class preserving* inference, *class admissible* learnability, and *absolute* learning. Next, we explain these notions in some more detail. Obviously, the hypothesis space must contain a suitable description for every language enumerated in the target indexed family. Therefore, one might be tempted to choose the indexed family itself as hypothesis space. If an indexed family $\mathcal{L}$ has to be inferred with respect to the hypothesis space $\mathcal{L}$, then we refer to this learning model as to *proper* inference. Note that proper inference has been studied by various authors (cf., e.g., Angluin [1, 2], Shinohara [45], Mukouchi [36]). Nevertheless, one may also allow any recursive enumeration of the range of $\mathcal{L}$ as well as any description of the enumerated languages provided membership remains uniformly decidable. The resulting learning model is referred to as *class preserving* inference. Moreover, when dealing with learning with anomalies, the following further generalization concerning hypothesis spaces is appropriate. A hypothesis space $\mathcal{G}$ is called *class admissible* for $\mathcal{L}$ provided that $\mathcal{G}$ contains for every L from the range of the target indexed family $\mathcal{L}$ at least one description that constitutes an approximation for L within the required precision. The resulting learning model is referred to as *class admissible* learning. Finally, we call an indexed family $\mathcal{L}$ *absolute* learnable if it can be inferred with respect to every class admissible hypothesis space for it.

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. Nevertheless, absolute learnability has its peculiarities as we shall show. 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.

The paper is structured as follows. Section 2 presents preliminaries, i.e., notations and definitions as well as further motivation for the research undertaken. Subsequently, we present our results concerning the learnability of indexed families with anomalies (cf. Section 3). The 4th Section deals with team inference from positive data in the setting of learning indexed families. Finally, we outline conclusions and present open problems (cf. Section 5). All references are given in Section 6.

2 Preliminaries

Let $\mathbb{N} = \{0, 1, 2, ...\}$ be the set of all natural numbers. We set $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$. In the sequel we assume familiarity with formal language theory (cf., e.g., Hopcroft and Ullman [19]). By Σ we denote any fixed finite alphabet of symbols. Let Σ^* be the free monoid over Σ, and let $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$, where ε denotes the empty string. Every subset $L \subseteq \Sigma^*$ is called a language. By *co-L* we denote the complement of L. 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 $|L \triangle \hat{L}| \le a$. 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. $|L \triangle \hat{L}| \le *$).

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. Furthermore, let $i = (s_0, b_0), (s_1, b_1), ...$ be an infinite sequence of elements of $\Sigma^* \times \{+, -\}$ such that $range(i) = \{s_k \mid k \in \mathbb{N}\} = \Sigma^*$, $i^+ = \{s_k \mid (s_k, b_k) = (s_k, +), k \in \mathbb{N}\} = L$ and $i^- = \{s_k \mid (s_k, b_k) = (s_k, -), k \in \mathbb{N}\} = co\text{-}L$. Then we refer to i as an ***informant***. If L is classified via an informant then we also say that L is represented by ***positive and negative***

data. Let L be a language. By $info(L)$ we denote the set of all informants for L. Moreover, let t, i be a text and an informant, respectively, and let x be a number. Then t_x, i_x denote the initial segment of t and i of length $x+1$, respectively, e.g., $i_2 = (s_0, b_0), (s_1, b_1), (s_2, b_2)$. Let t be a text and let $x \in \mathbb{N}$. Then we define $t_x^+ = \{s_k \mid k \leq x\}$. Furthermore, by i_x^+ and i_x^- we denote the sets $\{s_k \mid (s_k, +) \in i, k \leq x\}$ and $\{s_k \mid (s_k, -) \in i, k \leq x\}$, respectively. Finally, we write $t_x \sqsubseteq t_y$ ($t_x \sqsubset t_y$), iff t_x is a (proper) prefix of t_y.

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

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

In the following we refer to indexed families of uniformly recursive languages as indexed families for short. Examples for indexed families are the canonical enumeration of all regular languages, of all context-free languages, and of all context-sensitive languages over Σ, respectively (cf. Hopcroft and Ullman [19]). Moreover, we set $range(\mathcal{L}) = \{L_j \mid j \in \mathbb{N}\}$ for every indexed family $\mathcal{L} = (L_j)_{j \in \mathbb{N}}$. Note that the definition of an indexed family includes both, a description for every language L_j, and a particular enumeration of all the languages in its range. In particular, we may consider the indices of the enumerated languages as compiled grammars and acceptors, respectively (cf. Hopcroft and Ullman [19]).

As in Gold [17], 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 (or an informant i) 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 [17] and Wiehagen [49] 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 [11] investigated this phenomenon in detail. As it turns out, IIMs synthesizing grammars can be more powerful than those ones which are requested to output decision procedures. 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 or informant, respectively, 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 [17])** *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 [CLIM-INF]–infers L from text [informant] with respect to $\mathcal{G}$*** *iff for every text t [informant i] for L, there exists a $j \in \mathbb{N}$ such that the sequence $(M(t_x))_{x\in\mathbb{N}}$ $[(M(i_x))_{x\in\mathbb{N}}]$ converges in the limit to j and $L = L(G_j)$.*

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

Finally, let CLIM-TXT [CLIM-INF] 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 [CLIM-INF]–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. Note that class admissible learning is sometimes also referred to as class comprising learning (cf., e.g., Zeugmann and Lange [51] and the references therein). 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 sometime also referred to as exact learning (cf., e.g., Zeugmann and Lange [51]). Finally, we replace the prefix C by A to denote the collection of all those indexed families that can be learned in the limit with respect to *every* class admissible hypothesis space. The latter learning type is referred to as ***absolute learning***. We adopt this convention in the definitions of the learning types below.

Definition 1 could be easily generalized to arbitrary families of recursively enumerable languages (cf., e.g., Osherson et al. [38]). Nevertheless, we exclusively consider the restricted case defined above, since our motivating examples are all families of uniformly recursive languages. Furthermore, the following question naturally arises. Does the collection of inferable indexed families depend on the set of allowed hypothesis spaces introduced above? The answer to this question has been provided by Lange and Zeugmann [30]. We state their result in the following proposition that completely clarifies the relations between absolute, proper, class preserving and class admissible learning in the limit.

Proposition 1.

(1) *ALIM-TXT = PLIM-TXT = LIM-TXT = CLIM-TXT*,

(2) *ALIM-INF = PLIM-INF = LIM-INF = CLIM-INF.*

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 [11])** *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 [CLIM^a-INF]–infers L from text [informant] with respect to $\mathcal{G}$*** *iff for every text t [informant i] for L, there exists a $j \in \mathbb{N}$ such that the sequence $(M(t_x))_{x\in\mathbb{N}}$ $[(M(i_x))_{x\in\mathbb{N}}]$*

converges in the limit to j and $L =_a L(G_j)$.

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

Finally, let $CLIM^a$-TXT [$CLIM^a$-INF] 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 [$CLIM^a$-INF]–identifies $\mathcal{L}$ with respect to $\mathcal{G}$.

Obviously, λLIM^0-$TXT = \lambda LIM$-TXT for all $\lambda \in \{A, P, \varepsilon, C\}$. However, for $a \in \mathbb{N} \cup \{*\}$ there is a peculiarity we would like to mention here. Let $\mathcal{L}$ be an indexed family, and let $\mathcal{G} = (G_j)_{j\in\mathbb{N}}$ be any class admissible hypothesis space such that $\mathcal{L}$ is $CLIM^a$-TXT–($CLIM^a$-INF)–inferable with respect to $\mathcal{G}$. If $a = 0$, then class admissible learning implies $range(\mathcal{L}) \subseteq range(\mathcal{L}(\mathcal{G}))$. On the other hand, if $a \geq 1$, then class admissible learning only requires the existence of hypotheses in $range(\mathcal{G})$ that provide sufficiently precise approximations for every $L \in range(\mathcal{L})$. As a consequence, the proof technique of Lange and Zeugmann [30] cannot be directly applied to extend Proposition 1 to identification in the limit with anomalies.

Finally, we define learning in the limit by a team of IIMs. Team inference has been introduced by Smith [47] in the context of inferring recursive functions. Subsequently various authors have studied it intensively (cf., e.g., Pitt [40], Pitt and Smith [41], Jain and Sharma [20, 21]). For a survey concerning the results obtained and the problems studied the interested reader is referred to Smith [48]. Furthermore, in Chapter 3 we provide a summary of results that are most relevant to the problems studied in this paper.

Definition 3. (Smith [47]) *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,...,M_n) of IIMs (m,n)CLIM-TXT*** *[**(m,n)CLIM-INF**]–**infers L from text** [**informant**] **with respect to $\mathcal{G}$** iff for every text t [informant i] 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 $(M_{k_1}(t_x))_{x\in\mathbb{N}}, \cdots, (M_{k_m}(t_x))_{x\in\mathbb{N}}$ $[(M_{k_1}(i_x))_{x\in\mathbb{N}}, \cdots, (M_{k_m}(i_x))_{x\in\mathbb{N}}]$ 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 [$(m,n)CLIM$-INF]–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 [$(m,n)CLIM$- INF]–identifies L with respect to $\mathcal{G}$.

Finally, let $(m,n)CLIM$-TXT [$(m,n)CLIM$-INF] 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 [$(m,n)CLIM$-INF]–identifies $\mathcal{L}$ with respect to $\mathcal{G}$.

In the following we derive some easy consequences from the latter definition. First, it is easy to see that every indexed family can be inferred in the limit from informant by a single IIM. In particular, Gold's [17] identification by enumeration principle serves as a universal learning method (cf. Zeugmann and Lange [51] for a detailed discussion). Taking this observation as well as Proposition 1 into account, we directly obtain the following proposition.

Proposition 2. *For all $m, n \in \mathbb{N}^+$, $m \leq n$, we have:*

$(m,n)CLIM$-INF $=$ $CLIM$-INF $=$ $ALIM$-INF.

Next, we consider team inference from text. Recently, Meyer [34] extended Pitt's [39] 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 3. *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 [30] can be directly applied to relate the learning capabilities of absolute, proper, class preserving and class admissible team learning to one another. We display the resulting equality in the next proposition.

Proposition 4. *For all $n \in \mathbb{N}^+$ we have:*

$(1,n)ALIM\text{-}TXT \ = \ (1,n)PLIM\text{-}TXT \ = \ (1,n)LIM\text{-}TXT \ = \ (1,n)CLIM\text{-}TXT.$

In the following we aim to clarify the remaining relations between absolute, proper, class preserving and class admissible learning in the limit with anomalies. Moreover, we want to study the learning power of $\lambda LIM^a\text{-}TXT$, $\lambda \in \{A, P, \varepsilon, C\}$ and $(1,n)PLIM\text{-}TXT$ in dependence on the number of allowed anomalies the inferred grammars may have and the team size n, respectively. This is done in Sections 3 and 4.

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. Recently, several models of learning indexed families have been successfully characterized in terms of finite tell-tales, too (cf., e.g., Zeugmann, Lange and Kapur [52]). The characterizations obtained helped to gain a better understanding of what different learning models have in common and where the differences are. Hence, it is only natural to ask whether or not identification in the limit with anomalies may be characterized as well. This is indeed the case as we shall show. In order to do this, 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 5. (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 $\mathcal{L}$.*

Originally, this proposition characterized all those indexed families that are proper inferable in the limit. However, a straightforward application of Proposition 1 directly results in a complete characterization of indexed families that are learnable in the limit from positive data.

Next, we want to extend this characterization theorem to identification in the limit with anomalies. First, we show that, for every $a \in \mathbb{N} \cup \{*\}$, uniformly recursively enumerable a-tell-tales are sufficient to guarantee class preserving learning 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. Let $a \in \mathbb{N} \cup \{*\}$ be arbitrarily fixed, and let Q_j denote the a-tell-tale for $L(G_j)$. Furthermore, for every $x \in \mathbb{N}$, let Q_j^x be the set of all elements enumerated within xth steps of computation performed by the procedure g. Then, analogously as in Angluin [1] we define an IIM as follows:

IIM *M*: "On input t_x, execute Stage x.

Stage x **:** Search for the least index $j \le x$ which satisfies $Q_j^x \subseteq t_x^+ \subseteq L(G_j)$.
In case such an index j has been found, output it, and request the next input.
Otherwise, output x, and request the next input."

Let $L \in range(\mathcal{L})$ and $t \in text(L)$. We have to show that M LIM^a-TXT–infers L. Let j_0 be the least index j which satisfies $Q_j \subseteq L \subseteq L(G_j)$. Since $L \in range(\mathcal{L})$ and since $\mathcal{G}$ is a class preserving hypothesis space, such an index has to exist. Note that $L =_a L(G_{j_0})$ by Property (3) of a-tell-tales. Now it suffices to show that the IIM M converges to j_0.

For every $j < j_0$, take x_j which satisfies the following conditions.

(i) In case that $Q_j \not\subseteq L$, let x_j be sufficiently large such that $Q_j^{x_j} = Q_j \not\subseteq L$.

(ii) In case that $Q_j \subseteq L$ and $L \not\subseteq L(G_j)$, let x_j be sufficiently large such that $t_{x_j}^+ \not\subseteq L(G_j)$.

Furthermore, let x_{j_0} be the least x such that the a-tell-tale Q_{j_0} is completely enumerated, i.e., $Q_{j_0}^x = Q_{j_0}$. Let $X = \max\{x_1, x_2, \ldots, x_{j_0}, j_0\}$. Then for all $x \ge X$, (i) and (ii) imply that $Q_j^x \subseteq t_x^+ \subseteq L(G_j)$ does not hold for each $j < j_0$. On the other hand, $Q_{j_0}^x \subseteq t_x^+ \subseteq L(G_{j_0})$. Hence, M converges to j_0. □

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. Therefore, we continue with a lemma providing a useful normalization. The 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. Consequently, this lemma nicely contrasts a corresponding theorem for function learning, where one can always achieve convergence to *subfunctions* (cf. Case and Smith [12]).

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. A hypothesis space $\mathcal{G}$ may not have a superset of some language in $\mathcal{L}$. This problem is solved by mixing $\mathcal{G}$ with $\mathcal{L}$. 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 $j \leq x$ 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 $x \geq \tilde{x}$ 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 $L =_a L(\hat{G}_{\tilde{\jmath}})$. We show that $\hat{M}$ converges to $\tilde{\jmath}$.

For every $j < \tilde{\jmath}$, let $x(j)$ be the least x which satisfies the following conditions.

(i) if $L \not\subseteq L(\hat{G}_j)$, then $(t_x^+)^{(x)} \not\subseteq L(\hat{G}_j)$,

(ii) if $L \subseteq L(\hat{G}_j) \not\subseteq L(G_m) \cup L$, then $L(\hat{G}_j)^{(x)} \not\subseteq L(G_m) \cup L$.

Moreover let $x(\tilde{\jmath})$ be the least x such that $L(G_m) \cup t_x^+ = L(G_m) \cup L$. Note that such an $x(\tilde{\jmath})$ has to exist, since $|L \setminus L(G_m)| \leq a$. Let $X = \max\{\tilde{x}, x(0), x(1), x(2), \ldots, x(\tilde{\jmath}), \tilde{\jmath}\}$.

Claim. For every $x \geq X$, on input t_x, the IIM $\hat{M}$ outputs $\tilde{\jmath}$.

Obviously, the IIM $\hat{M}$ outputs $\tilde{\jmath}$, if $\tilde{\jmath}$ is the least $k \leq x$ such that $(t_x^+)^{(x)} \subseteq L(\hat{G}_k)^{(x)} \subseteq L(G_{j_x}) \cup t_x^+$. Hence, it suffices to show that no $j < \tilde{\jmath}$ satisfies this condition and $\tilde{\jmath}$ satisfies it, provided $x \geq X$. Consider any $j < \tilde{\jmath}$. We distinguish the following cases.

Case 1. $L \not\subseteq L(\hat{G}_j)$

In accordance with the choice of $x(j)$, we directly get $(t_x^+)^{(x)} \not\subseteq L(\hat{G}_j)$. Thus j does not satisfy the condition.

Case 2. $L \subseteq L(\hat{G}_j) \not\subseteq L(G_m) \cup L$

In accordance with the choice of $\tilde{x}$ and $x(j)$, $L(\hat{G}_j)^{(x)} \not\subseteq L(G_m) \cup L = L(G_{j_x}) \cup t_x^+$. Thus j does not satisfy the condition.

On the other hand, since $L \subseteq L(\hat{G}_{\tilde{\jmath}})$, we have $(t_x^+)^{(x)} \subseteq L(\hat{G}_{\tilde{\jmath}})$. Furthermore, $L(\hat{G}_{\tilde{\jmath}})^{(x)} \subseteq L(G_m) \cup L = L(G_{j_x}) \cup t_x^+$. Thus $\tilde{\jmath}$ satisfies the condition.

Since the search reaches $\tilde{\jmath}$ in Stage x, the IIM $\hat{M}$ outputs $\tilde{\jmath}$. This proves the claim.

By this claim, Lemma 2 immediately follows. □

Now we are ready to establish both the desired generalization of Proposition 1 as well as its converse. The next theorem actually states that a-tell-tales may be always uniformly enumerated with respect to the target indexed family $\mathcal{L}$ itself.

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. □

Note that the above proof exploits a special property of the sequence $\sigma_0 \cdot \ldots \cdot \sigma_m$. This property is usually referred to as locking sequence, and we shall extensively use it hereafter. Therefore, we continue with a formal definition of it.

Definition 5. (Osherson, Stob and Weinstein [38]) *Let M be an IIM and let L be a language. Furthermore, let $a \in \mathbb{N} \cup \{*\}$, and let $\mathcal{G} = (G_j)_{j \in \mathbb{N}}$ be a class admissible hypothesis space for $\{L\}$ with respect to a. Then, a sequence σ is called a* ***locking sequence*** *for L iff*

(1) *σ is a finite sequence, and $\sigma^+ \subseteq L$,*

(2) *for all finite sequences τ with $\tau^+ \subseteq L$, $M(\sigma \cdot \tau) = M(\sigma)$ and $L(G_{M(\sigma)}) =_a L$.*

Note that within the demonstration of the latter theorem, we have implicitly reproved the following fact. If $\mathcal{L}$ is an indexed family that can be $CLIM^a$-TXT-inferred, then for all $j \in \mathbb{N}$, a locking sequence for L_j is *limiting recursive*.

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

Proposition 6. *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 latter theorem immediately allows the following corollary.

Corollary 4. *For all $a \in \mathbb{N} \cup \{*\}$, $PLIM^a$-$TXT = CLIM^a$-TXT.*

Hence, in the following it suffices to deal exclusively with $PLIM^a$-TXT. We continue our investigations by separating $PLIM^a$-TXT and $PLIM^b$-TXT for all $a, b \in \mathbb{N}$, $a \neq b$.

Theorem 5. *For all $a \in \mathbb{N}^+$ there exists an indexed family $\mathcal{L}_a$ such that*

$$\mathcal{L}_a \in PLIM^a\text{-}TXT \setminus PLIM^{a-1}\text{-}TXT.$$

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^+. Obviously, $\mathcal{L}_a \in PLIM^a$-TXT; one may simply take the IIM M that always outputs 0. Now, suppose that $\mathcal{L}_a \in PLIM^{a-1}$-$TXT$. Then, by Proposition 6 every language $L \in range(\mathcal{L}_a)$ must possess an $(a-1)$-tell-tale with respect to $\mathcal{L}_a$. Consider any finite set Q which satisfies $Q \subseteq L_0 = \Sigma^+$. Clearly, there exists a language $L \in range(\mathcal{L}_a)$ such that $Q \subseteq L \subseteq L_0$ and $|L \bigtriangleup L_0| = a$, i.e., $L =_a L_0$ but $L \neq_{a-1} L_0$. This violates Property (3) of Definition 4. Therefore L_0 does not possess an $(a-1)$-tell-tale with respect to $\mathcal{L}_a$, a contradiction. □

Next we extend Theorem 5 to the $a = *$ case.

Corollary 6.

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

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 $\mathcal{L}_1, \mathcal{L}_2, \ldots$. Obviously $\mathcal{L} \in PLIM^*$-TXT. Suppose $\mathcal{L} \in \bigcup_{a \in \mathbb{N}} PLIM^a$-$TXT$. Then there exists an $a \in \mathbb{N}$ such that $\mathcal{L} \in PLIM^a$-TXT. However there exists no a-tell-tale for Σ^+ with respect to $\mathcal{L}$. In fact, for every finite set $Q \subseteq \Sigma^+$, there exists a language $L \in range(\mathcal{L}_{a+1}) \subseteq range(\mathcal{L})$ such that $Q \subseteq L \subset \Sigma^+$. This is a contradiction. Hence $\mathcal{L} \notin \bigcup_{a \in \mathbb{N}} PLIM^a$-$TXT$. □

With the next corollary we generalize a classical result of Gold [17]. Note that Case and Lynes [11] obtained a similar generalization but provided a completely different proof for it.

Corollary 7. *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 $PLIM^*$-TXT-inferable.*

Proof. Assume $Q \subseteq L$ is an arbitrary finite set. Since $\mathcal{L}$ is super-finite, $Q \in \mathit{range}(\mathcal{L})$ and of course $Q \subseteq Q \subseteq L$ hold. However $|L \triangle Q| = \infty$. Therefore, no $*$-tell-tale for L exists. By Theorem 6, $\mathcal{L}$ is not $PLIM^*$-TXT-inferable. □

Finally in this section we study the learning capabilities of absolute learning with anomalies. Our next theorem shows that the requirement to learn absolutely considerably restricts the learning power.

Proposition 7. *Let $a \in \mathbb{N}$. Then, $ALIM^a$-$TXT \subset PLIM^a$-TXT.*

Proof. $ALIM^a$-$TXT \subseteq PLIM^a$-TXT is clear by definition. Let $\mathcal{L}_a = (L_j)_{j\in\mathbb{N}}$ be the indexed family defined in Theorem 5. Hence, we know that $\mathcal{L}_a \in PLIM^a$-TXT. Next we show that $\mathcal{L}_a \notin ALIM^a$-$TXT$.

Let $\mathcal{G}$ be a hypothesis space such that $\mathcal{L}(\mathcal{G}) = (L_{j+1})_{j\in\mathbb{N}}$. Note that $\Sigma^+ \notin \mathit{range}(\mathcal{L}(\mathcal{G}))$. Obviously, $\mathcal{G}$ is an admissible hypothesis space for $\mathcal{L}_a$. Assume any IIM M that $ALIM^a$-TXT-infers $\mathcal{L}_a$ with respect to $\mathcal{G}$. Then there exists a locking sequence σ for $L_0 = \Sigma^+$. In particular, there has to be a $j \in \mathbb{N}$ such that

(A) for every finite sequence τ with $\tau^+ \subseteq \Sigma^+$, $j = M(\sigma \cdot \tau) = M(\sigma)$, and $L(G_j) =_a \Sigma^+$.

Note that by construction $L(G_j)$ has to be a language obtained by removing just a strings from Σ^+. However there exists a language $L \in \mathit{range}(\mathcal{L}_a)$ such that $\sigma^+ \subseteq L \subset \Sigma^+$ and $|L \triangle L(G_j)| = 2a$. Hence M cannot $CLIM^a$–infer L from every text for it that contains σ as prefix, a contradiction. □

The proof of the latter theorem points to the following interesting problem. As we have seen, for every $a \in \mathbb{N}^+$ there are indexed families $\mathcal{L}$, class admissible hypothesis spaces $\mathcal{G}$ for $\mathcal{L}$ with respect to a but no IIM can $CLIM^a$–infer $\mathcal{L}$ with respect to $\mathcal{G}$. Furthermore, the example provided by Proposition 7 forces any IIM M to converge sometimes to hypotheses that contain $2a$ anomalies. Hence, it is only natural to ask whether or not there is a universal upper bound for the number of anomalies that are inevitable provided the given hypothesis space is class admissible with respect to a. The answer is provided by our next theorem.

Theorem 8. *For all $a \in \mathbb{N}$ we have:*
Let $\mathcal{L}$ be an indexed family such that $\mathcal{L} \in CLIM^a$-TXT. Then, for every class admissible hypothesis space $\mathcal{G}$ for $\mathcal{L}$ with respect to a there exists an IIM M such that M $CLIM^{2a}$-TXT–learns $\mathcal{L}$ with respect to $\mathcal{G}$.

Proof. (*Sketch*) Suppose an IIM $\hat{M}$ witnesses $\mathcal{L} \in CLIM^a$-TXT with respect to a hypothesis space $\hat{\mathcal{G}}$. Without loss of generality, we may assume $\hat{M}$ and $\hat{\mathcal{G}}$ satisfies conditions in lemma 2. Let $\mathcal{G}$ be arbitrary class admissible hypothesis for $\mathcal{L}$ with respect to a. An IIM M works as following: On input t_x, simulate $\hat{M}$ and let $\hat{G}$ be a hypothesis an index of which is the last guess of $\hat{M}$. Search for the least index of a hypothesis G from $\mathcal{G}$ which satisfies $|L(G) \setminus L(\hat{G})| + |t_x^+ \setminus L(G)| \le a$ (condition #) and, if any, output the index.

Next, we show that M $CLIM^{2a}$-TXT-learns $\mathcal{L}$. Suppose a text t for a language $L \in \mathit{range}(\mathcal{L})$ is given. After sufficiently long time, $\hat{G}$ satisfies $L =_a L(\hat{G})$ and $L \subseteq L(\hat{G})$ by the assumption. Since $\mathcal{G}$ is class admissible for $\mathcal{L}$ with respect to a, there exists $G \in \mathit{range}(\mathcal{G})$ such that $L(G) =_a L$. Then, $|L(G) \setminus L(\hat{G})| + |L \setminus L(G)| \leq |L(G) \setminus L| + |L \setminus L(G)| = |L(G) \triangle L| \leq a$. Hence, there exists a hypothesis which satisfies the condition # in the limit.

On the other hand, assume M converged to the index of a hypothesis G'. Then, it has to hold that $|L(G') \setminus L(\hat{G})| + |L \setminus L(G')| \leq a$. Therefore $|L \triangle L(G')| = |L(G') \setminus L| + |L \setminus L(G')| \leq |L(G') \setminus L(\hat{G})| + |L(\hat{G}) \setminus L| + |L \setminus L(G')| \leq 2a$. Hence, $L =_{2a} L(G')$. M works correctly. This completes the proof. □

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

Learning in dependence on the number of anomalies allowed
and on the hypothesis spaces admissible

$$\begin{array}{ccccccccc}
PLIM^0\text{-}TXT & \subset & PLIM^1\text{-}TXT & \subset & \cdots & \subset & \bigcup_{a\in\mathbb{N}} PLIM^a\text{-}TXT & \subset & PLIM^*\text{-}TXT \\
\| & & \| & & & & \| & & \| \\
LIM^0\text{-}TXT & \subset & LIM^1\text{-}TXT & \subset & \cdots & \subset & \bigcup_{a\in\mathbb{N}} LIM^a\text{-}TXT & \subset & LIM^*\text{-}TXT \\
\| & & \| & & & & \| & & \| \\
CLIM^0\text{-}TXT & \subset & CLIM^1\text{-}TXT & \subset & \cdots & \subset & \bigcup_{a\in\mathbb{N}} CLIM^a\text{-}TXT & \subset & CLIM^*\text{-}TXT
\end{array}$$

Figure 1

4 Inferability by a Team of IIMs from Text

In this section, we investigate the relation between inference allowing anomalies and team inference. As we shall see, our results mainly resemble similar results obtained in the setting of function learning instead those ones established in learning recursively enumerable languages.

We start our investigations by comparing the learning capabilities of teams of two IIMs and learning with an arbitrarily but *a priori* fixed number of allowed anomalies. The next theorem shows that a team of two IIMs has sometimes more learning power than every IIM learning with anomalies.

Theorem 9. *For all $a \in \mathbb{N}$, $(1,2)PLIM\text{-}TXT \setminus PLIM^a\text{-}TXT \neq \emptyset$.*

Proof. Let $\mathcal{L}_{a+1} = (L_j)_{j\in\mathbb{N}}$ be the indexed family defined in theorem 5. Define a team of two IIMs M_1 and M_2 as follows: M_1 always outputs 0. The IIM M_2 can be straightforwardly defined to identify $L_1, L_2, \cdots$ from positive data. It is easy to see this team $(1,2)PLIM\text{-}TXT$-infers $\mathcal{L}_{a+1}$. On the other hand, we already know $\mathcal{L}_{a+1} \notin PLIM^a\text{-}TXT$ by theorem 5. □

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 [13] in case of learning recursive functions.

Theorem 10. *Let $a \in \mathbb{N}^+$. Then,*

$$PLIM^a\text{-}TXT \subseteq (1, a+1)PLIM\text{-}TXT.$$

Proof. (*Sketch*) Let M be any IIM witnessing $\mathcal{L} \in PLIM^a\text{-}TXT$. Without loss of generality, we may assume that M converges to supersets (cf. Lemmasuperset). Construct a team $(M_0, M_1, \ldots, M_a)$ of $a+1$ IIMs as follows:

(i) $M_0 = M$.

(ii) For every $k = 1, 2, \ldots, a$, we define an IIM M_k. On input t_x, the values of $M_k(t_x)$ are uniformly defined via the following Stage x.

Stage x**:** Let $j_x = M(t_x)$ and let $v_1, v_2, \ldots, v_k$ be first k strings in $L_{j_x} \setminus t_x^+$ occurring in the fixed enumeration of Σ^+. Search for the least index $j \leq x$ which satisfies $L_j^{(x)} = (L_{j_x} \setminus \{v_1, v_2, \ldots, v_k\})^{(x)}$ by generating the lexicographically ordered informants of lengths $x+1$ of L_j and L_{j_x}, respectively. If such a j is found, then output it, else output x.

Suppose a text t of language $L \in \mathit{range}(\mathcal{L})$ is input and that M converges to m. Then, $L_m =_a L$ and $L \subseteq L_m$ holds. Let $v_1, v_2, \ldots, v_k$ be all strings of $L_m \setminus L$ in the order occurring in the enumeration of Σ^+ (inevitably $k \leq a$). After sufficiently large stages, the first k strings in $L_{j_x} \setminus t_x^+$ become equal to $v_1, v_2, \ldots, v_k$. Moreover, if x is sufficiently large, then the kth IIM M_k always outputs the first index of language $L_m \setminus \{v_1, v_2, \ldots, v_k\} = L$. □

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 12). In order to prove this, we need a further generalization of the tell-tale concept which is provided by the next definition.

Definition 6. *Let $\mathcal{L}$ be an indexed family and $L \in \mathit{range}(\mathcal{L})$.*

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 $L' \in \mathit{range}(\mathcal{L})$ 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 $\hat{L} \in \mathit{range}(\mathcal{L})$, $Q \subseteq \hat{L} \subset L$ 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 [37]. The next lemma describes a necessary condition for team inference from positive data in terms of n-depth tell-tales.

Lemma 11. Let $n \in \mathbb{N}$ and let $\mathcal{L}$ be any indexed family that is $(1, n+1)\mathit{PLIM\text{-}TXT}$–inferable. Then, for all $L \in \mathit{range}(\mathcal{L})$, there exists an n-depth tell-tale for L with respect to $\mathcal{L}$.

Proof. Since $\mathcal{L} \in (1, n+1)\mathit{PLIM\text{-}TXT}$, there exists a team $(M_0, M_1, \ldots, M_n)$ of IIMs that $(1, n+1)\mathit{PLIM\text{-}TXT}$–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" 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 $P(\tau) = \{(k, j) \mid M_k(\sigma \cdot \sigma_0 \cdot \sigma_1 \cdot \ldots \cdot \sigma_{x-1} \cdot \tau) = j \text{ and } L_j = L'\} \neq \emptyset$. 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 $L \in range(\mathcal{L})$ 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 $L^1 \in range(\mathcal{L})$ 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 $L^{x+1} \in range(\mathcal{L})$ 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 $L^{n+1} \in range(\mathcal{L})$ 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 12. *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, $\mathcal{L} \in PLIM^a\text{-}TXT$.

Claim. Let $k = 1, \ldots, a$ and $L \in \mathit{range}(\mathcal{L}^{(k)})$. Then, L does not have any $(k-1)$-depth tell-tale with respect to $\mathcal{L}$.

Proof of Claim. We prove this claim by induction.

Base step: Let $L \in \mathit{range}(\mathcal{L}^{(1)})$. For every finite set $Q \subseteq L$, there exists $L' \in \mathit{range}(\mathcal{L}^{(0)})$ such that $Q \subseteq L' \subset L$. Hence L does not have any 0-depth tell-tale.

Induction step: Let $L \in \mathit{range}(\mathcal{L}^{(k+1)})$. For every finite set $Q \subseteq L$, there exists $L' \in \mathit{range}(\mathcal{L}^{(k)})$ such that $Q \subseteq L' \subset L$. From the induction hypothesis, L' does not have $(k-1)$-depth tell-tale. Hence L does not have any k-depth tell-tale.

By mathematical induction, the claim was proved.

As we have seen by the claim made above $\Sigma^+ \in \mathit{range}(\mathcal{L}^{(a)})$ does not have an $(a-1)$-depth tell-tale. Thus, by lemma 11, $\mathcal{L}$ is not $(1,a)PLIM\text{-}TXT$–inferable. □

Figure 1 illustrates the relations between inference with anomalies and team learning. Note that every relation not explicitly mentioned indicates incomparability of the relevant identification criteria.

Figure 1: a relation between two criteria

$$\begin{array}{ccc}
PLIM^0\text{-}TXT & = & (1,1)PLIM\text{-}TXT \\
\cap & & \cap \\
PLIM^1\text{-}TXT & \subset & (1,2)PLIM\text{-}TXT \\
\cap & & \cap \\
PLIM^2\text{-}TXT & \subset & (1,3)PLIM\text{-}TXT \\
\cap & & \cap \\
PLIM^3\text{-}TXT & \subset & (1,4)PLIM\text{-}TXT \\
\vdots & & \vdots
\end{array}$$

5 Conclusion

In the present paper, 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.

Moreover, within Definition 4 we introduced the notion of a-tell-tales, and succeed in characterizing inference allowing anomalies. These results extend Angluin's [1] Theorem to the case of approximate inference. Furthermore, the characterization mentioned enabled us to completely clarify the relation between proper learning, class preserving and class admissible learning with anomalies. The proved equality generalized Lange and Zeugmann's [30] relevant result to approximate inference. Additionally, in the process of proving the characterization theorem we found Lemma 2. This lemma describes an important property of anomaly allowing inference, and it also plays an important role in clarifying the relationship between team inference and approximative learning (cf. Theorem 10).

In order to investigate team inference, we introduced a notion of n-depth tell-tales, a further generalization of Angluin's [1] fundamental concept. However, 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.

Finally, concerning absolute learning several problems also remain open. Finding necessary and sufficient conditions for absolute learning with anomalies is a challenge for further research. Furthermore, we did not succeed in completely clarifying whether or not $ALIM^a \subset ALIM^{a+1}$ for all $a \in \mathbb{N}$.

6 References

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

[2] D. Angluin, Finding patterns common to a set of strings, *Journal of Computer and System Sciences* **21** (1980) 46–62.

[3] D. Angluin and C. H. Smith, Inductive inference: theory and methods, *Computing Surveys* **15** (1983) 237–269.

[4] D. Angluin and C. H. Smith, Formal inductive inference, in: *Encyclopedia of Artificial Intelligence, Vol. 1* (Wiley-Interscience Publication, New York, 1987) 409–418.

[5] S. Arikawa, S. Goto, S. Ohsuga and T. Yokomori (eds.), *Proc. of the 1st International Workshop on Algorithmic Learning Theory*, (Japanese Society for Artificial Intelligence, Tokyo, 1990).

[6] S. Arikawa, A. Maruoka and T. Sato (eds.), *Proc. of the 2nd International Workshop on Algorithmic Learning Theory*, (Japanese Society for Artificial Intelligence, Tokyo, 1991).

[7] S. Arikawa and K. P. Jantke (Eds.), *Proc. of the 5th International Workshop on Algorithmic Learning Theory* and *Proc. of the 4th International Workshop on Analogical and Inductive Inference.* Lecture Notes in Artificial Intelligence Vol. **872** (1994), Springer Verlag, Berlin.

[8] S. Arikawa, T. Shinohara and T. Miyahara, Knowledge Acquisition and Learning, Chapter 6 Theory of Inductive Inference (in Japanese), (OHMSHA, LTD.) 147–197.

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

[10] J. Case, The power of vacillation, in: *Proc. of the 1st Workshop on Computational Learning Theory* (Morgan Kaufmann Publishers Inc., San Mateo, 1988) 196–205.

[11] 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.

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

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

[14] M. Fulk, Prudence and other restrictions in formal language learning, *Information and Computation* **85** (1990) 1–11.

[15] M. Fulk and J. Case (eds.), *Proc. of the 3rd Annual Workshop on Computational Learning Theory*, (Morgan Kaufmann Publishers Inc., San Mateo, 1990).

[16] M. E. Gold, Limiting Recursion, *The Journal of Symbolic Logic* **30-1** (1965) 28–48.

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

[18] D. Haussler (ed.), *Proc. of the 5th Annual Workshop on Computational Learning Theory* (ACM Press, New York, 1992).

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

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

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

[22] K. P. Jantke (ed.), *Proc. of the 2nd International Workshop on Analogical and Inductive Inference*, Lecture Notes in Artificial Intelligence Vol. 397, (Springer-Verlag, Berlin, 1989).

[23] K. P. Jantke (ed.), *Proc. of the 3rd International Workshop on Analogical and Inductive Inference*, Lecture Notes in Artificial Intelligence Vol. 642, (Springer-Verlag, Berlin, 1992).

[24] M. Kanazawa, Learnable classes of categorical grammars, PhD thesis, Stanford University, Department of Linguistics, 1994.

[25] S. Kapur, Computational learning of languages, Ph.D. thesis, Cornell University, Computer Science Department Technical Report 91-1234, 1991.

[26] 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.

[27] S. Lange, The representation of recursive languages and its impact on the efficiency of learning, in: *Proc. of the 7th Annual ACM Conference on Computational Learning Theory*, (ACM Press, New York, 1994) 256–267.

[28] S. Lange and T. Zeugmann, On the power of monotonic language learning, TH Leipzig, Fachbereich Mathematik und Informatik, GOSLER–Report 05/92, 1992.

[29] S. Lange and T. Zeugmann, Monotonic versus non-monotonic language learning, in: *Proc. of the 2nd International Workshop on Nonmonotonic and Inductive Logic* Lecture Notes in Artificial Intelligence **659** (Springer-Verlag, Berlin, 1993) 254–269.

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

[31] S. Lange and T. Zeugmann, Language learning in dependence on the space of hypotheses, in: *Proc. of the 6th Annual ACM Conference on Computational Learning Theory* (ACM Press, New York, 1993) 127–136.

[32] S. Lange and T. Zeugmann, On the impact of order independence to the learnability of recursive languages, FUJITSU Laboratories Ltd., Institute for Social Information Science, Numazu, Research Report ISIS-RR-93-17E, 1993.

[33] S. Lange and T. Zeugmann, Characterization of language learning on informant under various monotonicity constraints, *Journal of Experimental and Theoretical Artificial Intelligence* **6** (1994) 73–94.

[34] L. Meyer, personal communication

[35] T. Moriyama and M. Sato, Properties of language classes with finite elasticity, in: *Proc. of the 4th International Workshop on Algorithmic Learning Theory* Lecture Notes in Artificial Intelligence **744** (Springer-Verlag, Berlin, 1993) 187–196.

[36] Y. Mukouchi, Inductive inference of an approximate concept from positive data, Kyushu University, Research Institute of Fundamental Information Science, Technical Report RIFIS-TR-CS 74, 1993.

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

[38] 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).

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

[40] L. Pitt, Probabilistic Inductive Inference, *Journal Association for Computing Machinery* **36** (1989), 383 – 433.

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

[42] R. Rivest, D. Haussler and M.K. Warmuth (eds.), *Proc. of the 2nd Annual Workshop on Computational Learning Theory*, (Morgan Kaufmann Publishers Inc., San Mateo, 1989).

[43] M. Sato and K. Umayahara, Inductive inferability for formal languages from positive data, *IEICE Transactions on Information and Systems* **E–75D**, 415–419.

[44] J. Shawe-Taylor and M. Anthony (eds.), *Proc. of the 1st European conference on Computational Learning Theory: EuroCOLT'93*, The Institute of Mathematics and its Applications Conference Series, New Series Number 53, Oxford University Press, Oxford 1994.

[45] T. Shinohara, Polynomial time inference of extended regular pattern languages, in: *Proc. of the RIMS Symposia on Software Science and Engineering* Lecture Notes in Computer Science **147** (Springer-Verlag, Berlin, 1982) 115–127.

[46] T. Shinohara, Rich classes inferable from positive data: length bounded elementary formal systems, *Information and Computation* **108** (1994), 175 – 186.

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

[48] C. H. Smith, Three decades of team learning, in: *Proc. of the 5th International Workshop on Algorithmic Learning Theory* Lecture Notes in Artificial Intelligence **872** (Springer-Verlag, Berlin, 1994) 211–228.

[49] 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.

[50] R. Wiehagen and T. Zeugmann, Ignoring data may be the only way to learn efficiently, *Journal of Experimental and Theoretical Artificial Intelligence* **6** (1994) 131–144.

[51] 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.

[52] T. Zeugmann, S. Lange and S. Kapur, Characterizations of monotonic and dual monotonic language learning, *Information and Computation***120**, No. 2 (1995) 155 – 173.