Language Learning with a Bounded Number of Mind Changes

Steffen Lange
TH Leipzig
FB Mathematik und Informatik
PF 66
O–7030 Leipzig
steffen@informatik.th-leipzig.de

Thomas Zeugmann
TH Darmstadt
Institut für Theoretische Informatik
Alexanderstr. 10
W–6100 Darmstadt
zeugmann@iti.informatik.th-darmstadt.de

Abstract

We study the learnability of enumerable families $\mathcal{L}$ of uniformly recursive languages in dependence on the number of allowed mind changes, i.e., with respect to a well–studied measure of efficiency. We distinguish between *exact* learnability ($\mathcal{L}$ has to be inferred w.r.t. $\mathcal{L}$) and *class preserving* learning ($\mathcal{L}$ has to be inferred w.r.t. some suitable chosen enumeration of all the languages from $\mathcal{L}$) as well as between learning from *positive* and from both, *positive and negative* data.

The measure of efficiency is applied to prove the superiority of class preserving learning algorithms over exact learning. We considerably improve results obtained previously and establish two infinite hierarchies. Furthermore, we separate exact and class preserving learning from positive data that avoids *overgeneralization.* Finally, language learning with a bounded number of mind changes is completely characterized in terms of recursively generable finite sets. These characterizations offer a new method to handle overgeneralizations and resolve an open question of Mukouchi (1992).

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, 1987, Osherson, Stob and Weinstein, 1986).

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, 1982, Case, 1988, Fulk, 1990). Looking at potential applications it seemed reasonable to restrict ourselves to study language learning of families of uniformly recursive languages. Recently, this topic has attracted much attention (cf. e.g. Shinohara, 1990, Kapur and Bilardi, 1992, Lange and Zeugmann, 1992, Mukouchi, 1992). The general situation investigated in language learning can be described as follows: Given more and more information concerning the language to be learnt, the inference device has to produce, from time to time, a hypothesis about the phenomenon to be inferred. The set of all admissible hypotheses is called space of hypotheses. Furthermore, the information given may contain only *positive examples*, i.e., exactly all the strings contained in the language to be recognized, as well as both *positive and negative examples*, i.e., all strings over the underlying alphabet which are classified with respect to their containment to the unknown language. The sequence of hypotheses has to converge to a hypothesis correctly describing the object to be learnt. Consequently, the inference process is an ongoing one. If $d_1, d_2, \ldots$ denotes the sequence of data the inference machine M is successively fed with, then we use $h_1, h_2, \ldots$ to denote the corresponding hypotheses produced by M. We say that M changes its mind, or synonymously, M performs a mind change, iff $h_i \neq h_{i+1}$. The number of mind changes is a measure of efficiency and has been introduced by Barzdin and Freivalds (1972). Subsequently, this measure has been intensively studied. Barzdin and Freivalds (1972) proved the following remarkable result concerning inductive inference of enumerable classes of recursive functions. Gold's (1967) *identification by enumeration* technique yields successful inference *within the enumeration* but $n-1$ mind changes may be necessary to learn the nth function. On the other hand, there are a learning algorithm and a space of hypotheses such that the nth function in enumeration can be learnt with at most $O(\log n) + o(\log n)$ mind changes. This bound is optimal. Their result impressively shows that a careful choice of the space of hypotheses may considerably influence the efficiency of learning. Moreover, Case and Smith (1983) established a hierarchy in terms of mind changes and anomalies. Wiehagen, Freivalds and Kinber (1984) used the number of mind changes to prove advantages of probabilistic learning algorithms over deterministic ones. Gasarch and Velauthapillai (1992) studied *active learning* in dependence on the number of mind changes.

Hence, it is only natural to ask whether or not this measure of efficiency is of equal importance in language learning. Answering this question is by no means trivial, since, in general, at least inductive inference from positive data may behave totally different than inductive inference of recursive functions does (cf. e.g. Case, 1988, Fulk, 1990). This is already caused by the fact that Gold's (1967) identification by enumeration technique does not necessarily succeed. The main new problem consists in detecting or avoiding overgeneralizations, i.e., hypotheses describing proper supersets of the target language. Mukouchi (1992) studied the power of mind changes for learning algorithms that infer indexed families of recursive languages *within the given enumeration*. Moreover, he characterized language learning with a bounded number of mind changes in case that equality of languages within the given enumeration is decidable.

What we present in the sequel is an almost complete investigation of the power of mind changes. For the sake of presentation we introduce some notations. An indexed family $\mathcal{L}$ is said to be *exactly* learnable if there is a learning algorithm inferring $\mathcal{L}$ with respect to $\mathcal{L}$ itself. Furthermore, $\mathcal{L}$ is learnable by a *class preserving* learning algorithm M, if there is a space $\mathcal{G} = (G_j)_{j \in \mathbb{N}^+}$ of hypotheses such that any G_j describes a language from $\mathcal{L}$, and M infers $\mathcal{L}$ with respect to $\mathcal{G}$. In other words, any produced hypothesis is required to describe a language contained in $\mathcal{L}$ but we have the freedom to use a possibly *different enumeration* of $\mathcal{L}$ and possibly *different descriptions* of any $L \in \mathcal{L}$.

We compare exact and class preserving language learning in dependence on the allowed number of mind changes as well as in dependence on the choice of the space of hypotheses and on information presentation. The strongest possible separation is established, i.e., we prove that there are indexed families $\mathcal{L}$ which are exactly learnable from positive data with at most $k+1$ mind changes but that are not class preservingly learnable from *positive and negative* data with at most k mind changes. This result sheds considerably more light on the power of one additional mind change than Mukouchi's (1992) hierarchy of exact learning in terms of mind changes. Furthermore, we compare exact and class preserving language learning avoiding overgeneralization and separate them (cf. Corollary 10). Applying the proof technique developed we show that exact language learning from positive data with a bounded number of mind changes is always less powerful than class preserving inference restricted to the same number of mind changes (cf. Theorem 11). Finally, we completely characterize class preserving language learning in terms of recursively generable finite sets (cf. Theorem 14 and 15). In particular, we offer a different possibility to handle overgeneralization than Angluin (1980) did.

2. Preliminaries

By $\mathbb{N} = \{0, 1, 2, 3, ...\}$ we denote the set of all natural numbers. Moreover, we set $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$. In the sequel we assume familiarity with formal language theory. By Σ we denote any fixed finite alphabet of symbols. Let Σ^* be the free monoid over Σ. The length of a string $s \in \Sigma^*$ is denoted by $|s|$. Any subset $L \subseteq \Sigma^*$ is called a language. By $co - L$ we denote the complement of L. Let L be a language and $t = s_1, s_2, s_3, ...$ 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*. Furthermore, let $i = (s_1, b_1), (s_2, b_2), ...$ be a 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 $i^- = \{s_k \mid (s_k, b_k) = (s_k, -), k \in \mathbb{N}^+\} = co - L$. 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*. 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, respectively.

We restrict ourselves to deal exclusively with indexed families of recursive languages defined as follows (cf. Angluin, 1980):
A sequence $L_1, L_2, L_3, ...$ is said to be an *indexed family* $\mathcal{L}$ of 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

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

In the sequel we often denote an indexed family and its range by the same symbol $\mathcal{L}$. What is meant will be clear from the context.

As in Gold (1967) 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 (an informant i) and it either requires the next input string, or it first outputs a hypothesis, i.e., a number encoding a certain computer program, and then it requires the next input string (cf. e.g. Angluin, 1980).

At this point we have to clarify what space of hypotheses we should choose. Gold (1967) and Wiehagen (1977) pointed out that there is a difference in what can be inferred in dependence on whether we want to synthesize in the limit grammars or decision procedures. Case and Lynes (1982) 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 as long as uniform decidability of membership is required. Nevertheless, we decided to require the IIMs to output grammars, since this learning goal fits better with the intuitive idea of language learning. Furthermore, since we exclusively deal with indexed families $\mathcal{L} = (L_j)_{j \in \mathbb{N}^+}$ of recursive languages we almost always take as space of hypotheses an enumerable family of grammars $G_1, G_2, G_3, \ldots$ over the terminal alphabet Σ satisfying $\mathcal{L} = \{L(G_j) \mid j \in \mathbb{N}^+\}$. Moreover, we require that membership in $L(G_j)$ is uniformly decidable for all $j \in \mathbb{N}^+$ and all strings $s \in \Sigma^*$. The IIM outputs numbers j which we interpret as G_j.

A sequence $(j_x)_{x \in \mathbb{N}^+}$ of numbers is said to be convergent in the limit iff there is a number j such that $j_x = j$ for almost all numbers x.

Definition 1. (Gold, 1967) *Let $\mathcal{L}$ be an indexed family of languages, $L \in \mathcal{L}$, and let $\mathcal{G} = (G_j)_{j \in \mathbb{N}^+}$ be a space of hypotheses. An IIM M $LIM-TXT$ ($LIM-INF$)–identifies L on a text t (an informant i) with respect to $\mathcal{G}$ iff it almost always outputs a hypothesis and the sequence $(M(t_x))_{x \in \mathbb{N}^+}$ ($(M(i_x))_{x \in \mathbb{N}^+}$) converges in the limit to a number j such that $L = L(G_j)$.*
Moreover, M $LIM-TXT$ ($LIM-INF$)–identifies L, iff M $LIM-TXT$ ($LIM-INF$)–identifies L on every text (informant) for L. We set:
$LIM-TXT(M) = \{L \in \mathcal{L} \mid M\ LIM-TXT-identifies\ L\}$ and define $LIM-INF(M)$ analogously.
Finally, let $LIM-TXT$ ($LIM-INF$) denote the collection of all families $\mathcal{L}$ of indexed families of recursive languages for which there is an IIM M such that $\mathcal{L} \subseteq LIM-TXT(M)$ ($\mathcal{L} \subseteq LIM-INF(M)$).

Definition 1 could be easily generalized to arbitrary families of recursively enumerable languages (cf. Osherson et al., 1986). Nevertheless, we exclusively consider the restricted case defined above, since our motivating examples are all indexed families of recursive languages. Moreover, it may be well conceivable that the weakening of $\mathcal{L} = \{L(G_j) \mid j \in \mathbb{N}^+\}$ to $\mathcal{L} \subseteq \{L(G_j) \mid j \in \mathbb{N}^+\}$ may increase the collection of inferable indexed families. However, it does not, as the following proposition shows.

Proposition 1. *Let $\mathcal{L}$ be an indexed family and let $\mathcal{G} = (G_j)_{j\in\mathbb{N}^+}$ be any space of hypotheses such that $\mathcal{L} \subseteq \{L(G_j) \mid j \in \mathbb{N}^+\}$ and membership in $L(G_j)$ is uniformly decidable. Then we have: If there is an IIM M inferring $\mathcal{L}$ on text (informant) with respect to $\mathcal{G}$, then there is also an IIM $\hat{M}$ that learns $\mathcal{L}$ on text (informant) with respect to $\mathcal{L}$.*

Nevertheless, the proof of Proposition 1 does not preserve the number of mind changes. As we shall see later, the efficiency of learning may be well influenced by the choice of the space of hypotheses.

Within the next definition we consider the case that the number of allowed mind changes is bounded by an a priorily fixed number.

Definition 2. (Barzdin and Freivalds, 1972) *Let $\mathcal{L}$ be an indexed family of languages, $\mathcal{G} = (G_j)_{j\in\mathbb{N}^+}$ a space of hypotheses, $k \in \mathbb{N} \cup \{*\}$, and $L \in \mathcal{L}$. An IIM M $LIM_k - TXT$ ($LIM_k - INF$)–identifies L on text t (informant i) with respect to $\mathcal{G}$ iff for every text t (informant i) the following conditions are fulfilled:*

(1) $L \in LIM - TXT(M) \quad (L \in LIM - INF(M))$

(2) For any $L \in \mathcal{L}$ and any text t (informant i) of L the IIM M performs, when fed with t (i), at most k ($k = *$ means at most finitely many) mind changes.

$LIM_k - TXT(M)$, $LIM_k - INF(M)$ as well as $LIM_k - TXT$ and $LIM_k - INF$ are defined in the same way as above.

Obviously, $LIM_* - TXT = LIM - TXT$ as well as $LIM_* - INF = LIM - INF$.

Next to we sharpen Definition 1 in additionally requiring that any mind change has to be caused by a "provable misclassification" of the hypothesis to be rejected.

Definition 3. (Angluin, 1980) *Let $\mathcal{L}$ be an indexed family, $L \in \mathcal{L}$, and let $\mathcal{G} = (G_j)_{j\in\mathbb{N}^+}$ be a space of hypotheses. An IIM M CONSERVATIVE–TXT identifies L on text t with respect to $\mathcal{G}$ iff for every text t the following conditions are satisfied:*

(1) $L \in LIM - TXT(M)$

(2) If M on input t_x makes the guess j_x and then makes the guess $j_{x+k} \neq j_x$ at some subsequent step, then $L(G_{j_x})$ must fail to contain some string from t_{x+k}

CONSERVATIVE–TXT(M) as well as the collections of sets *CONSERVATIVE–TXT* are analogously defined as above.

For any mode of inference defined above we use the prefix E to denote exact learning, i.e., the fact that $\mathcal{L}$ has to be inferred with respect to $\mathcal{L}$ itself. For example, $ELIM_k - TXT$ denotes exact learnability with at most k mind changes from text. Despite the fact that $LIM - TXT = ELIM - TXT$, $LIM_0 - TXT = ELIM_0 - TXT$ as well as $LIM - INF = ELIM - INF$, $LIM_0 - INF = ELIM_0 - INF$ none of the analogous statements is true for the other modes of inference defined above, as we shall see.

3. Separations

The aim of the present chapter is to relate the different types of language learning defined above one to the other.

Theorem 1. (Mukouchi, 1992)

$ELIM_0 - TXT \subset ELIM_1 - TXT \subset ELIM_2 - TXT \subset ... \subset ELIM_* - TXT$

We want to strengthen the theorem above in two directions. Our first sharpening is a refinement of Theorem 1.

Theorem 2. $\bigcup_{k \in \mathbb{N}} LIM_k - TXT \subset LIM - TXT$

The proof of Theorem 2 allows the following corollary.

Corollary 3. $ECONSERVATIVE\text{-}TXT \setminus \bigcup_{k \in \mathbb{N}} LIM_k - TXT \neq \emptyset$

Our next theorem shows that, in general, one additional mind change can neither be traded versus information presentation nor versus an appropriate choice of the space of hypotheses.

Theorem 4. *For all* $k \geq 0$: $ELIM_{k+1} - TXT \setminus LIM_k - INF \neq \emptyset$

The following hierarchy is an immediate consequence of the latter theorems.

$LIM_0 - TXT \subset LIM_1 - TXT \subset ... \subset LIM_k - TXT \subset ... \subset \bigcup_{k \in \mathbb{N}} LIM_k - TXT \subset LIM - TXT$

We have shown that $LIM_0 - INF \subset CONSERVATIVE\text{-}TXT$ (cf. Lange, Zeugmann and Kapur, 1992). Surprisingly, it makes a real difference, if an IIM is allowed to change its mind at most one time.

Theorem 5. $ELIM_1 - INF \setminus LIM - TXT \neq \emptyset$

Moreover, $LIM_k - TXT \subseteq LIM_k - INF$. Since $LIM_0 - TXT \subset LIM_0 - INF$ Theorem 5 yields the following corollary.

Corollary 6.

(1) *For all* $k \geq 0$, $LIM_k - TXT \subset LIM_k - INF$

(2) $LIM_0 - INF \subset LIM_1 - INF \subset LIM_2 - INF \subset ... \subset LIM_* - INF$

As above, this result can be sharpened, too.

Lemma 7. $\bigcup_{k \in \mathbb{N}} LIM_k - INF \subset LIM - INF$

The latter result has the following consequence.

Corollary 8. $LIM - TXT \;\#\; \bigcup_{k \in \mathbb{N}} LIM_k - INF$.

Summarizing the results above we obtain the following hierarchy:

$LIM_0 - INF \subset LIM_1 - INF \subset ... \subset LIM_k - INF \subset ... \subset \bigcup_{k \in \mathbb{N}} LIM_k - INF \subset LIM - INF$

Finally, we want to compare exact and class preserving language learning with an a priori bounded number of mind changes. Moreover, we compare conservatively working IIMs with those ones performing a bounded number of mind changes. The next theorems and a corollary thereof relate these different modes of inference one to the other. In particular, we show class preserving learning with at most one mind change has to be performed by conservatively working IIMs. Moreover, one mind change is already sufficient to beat exact conservative learning.

Theorem 9.

(1) $LIM_1 - TXT \subset CONSERVATIVE\text{–}TXT$

(2) $LIM_1 - TXT \setminus ECONSERVATIVE\text{–}TXT \neq \emptyset$

The following corollary is an immediate consequence of the latter theorem.

Corollary 10.

(1) $ECONSERVATIVE\text{–}TXT \subset CONSERVATIVE\text{–}TXT$

(2) $ELIM_1 - TXT \subset LIM_1 - TXT$

Finally, a nontrivial modification of the proof technique above may be applied to obtain the desired separation of exact and class preserving language learning with a bounded number of mind changes.

Theorem 11. *For all $k \geq 1$:*

(1) $ELIM_k - TXT \subset LIM_k - TXT$

(2) $ELIM_k - INF \subset LIM_k - INF$

However, some problems remained open. The most intriguing question is whether $LIM_k - TXT \setminus CONSERVATIVE\text{–}TXT \neq \emptyset$. In Lange, Zeugmann and Kapur (1992) we have shown that there is an indexed family $\mathcal{L} \in LIM - TXT \setminus CONSERVATIVE\text{–}TXT$. Nevertheless, the proof given there does not yield any a priori bound for the number of allowed mind changes. On the other hand, a careful analysis of our proof showed that the IIM witnessing $\mathcal{L} \in LIM - TXT$ does not work *semantically finite*. An IIM is said to work *semantically finite* iff for all $L \in \mathcal{L}$, any text t of L the following condition is satisfied: Let j be the hypothesis the sequence $(M(t_x))_{x \in \mathbb{N}^+}$ converges to and let z be the least number such that $M(t_z) = j$. Then $L(G_{M(t_y)}) \neq L(G_j)$ for all $y < z$. That means, a semantically finite working IIM is never allowed to reject a guess that is correct for the language to be learnt. As it turns out, this phenomenon is a general one.

Theorem 12. *Let $\mathcal{L}$ be an indexed family, and let $\mathcal{G} = (G_j)_{j \in \mathbb{N}^+}$ be a space of hypotheses. If there is an IIM M working semantically finite such that $\mathcal{L} \in LIM - TXT(M)$, then $\mathcal{L} \in CONSERVATIVE\text{–}TXT$.*

Finally, we obtain the following characterization of conservatively working IIMs.

Theorem 13. *Let $\mathcal{L}$ be an indexed family. Then $\mathcal{L} \in CONSERVATIVE$–$TXT$ iff there is a space $\mathcal{G}$ of hypotheses and an IIM M inferring $\mathcal{L}$ semantically finite in the limit with respect to $\mathcal{G}$.*

4. Characterization Theorems

Characterizations play an important role in that they lead to a deeper insight into the problem how algorithms performing the inference process may work (cf. e.g. Blum and Blum, 1975, Wiehagen, 1977, Angluin, 1980, Zeugmann, 1983, Jain and Sharma, 1989). Moreover, characterizations may help gain a better understanding of the properties objects should have in order to be inferable in the desired sense. A very illustrative example is Angluin's (1980) characterization of those indexed families for which learning in the limit from positive data is possible. In particular, this theorem provides insight into the problem how to deal with overgeneralizations. Our next theorem offers an alternative way to resolve this question. We characterize $LIM_k - TXT$ in terms of recursively generable finite tell–tales. A family of finite sets $(T_j)_{j \in \mathbb{N}^+}$ is said to be recursively generable, iff there is a total effective procedure g which, on input j, generates all elements of T_j and stops. If the computation of $g(j)$ stops and there is no output, then T_j is considered to be empty. Finally, for notational convenience we use $L(\mathcal{G})$ to denote $\{L(G_j) \mid j \in \mathbb{N}^+\}$ for any space $\mathcal{G} = (G_j)_{j \in \mathbb{N}^+}$ of hypotheses.

Theorem 14. *Let $\mathcal{L}$ be an indexed family of recursive languages, and $k \in \mathbb{N}$. Then: $\mathcal{L} \in LIM_k - TXT$ if and only if there is a space of hypotheses $\hat{\mathcal{G}} = (\hat{G}_j)_{j \in \mathbb{N}^+}$, a computable relation $\prec$ over $\mathbb{N}^+$, and a recursively generable family $(\hat{T}_j)_{j \in \mathbb{N}^+}$ of finite and non–empty tell-tale sets such that*

(1) $range(\mathcal{L}) = L(\hat{\mathcal{G}})$.

(2) For all $z \in \mathbb{N}^+$, $\hat{T}_z \subseteq L(\hat{G}_z)$.

(3) For all $L \in \mathcal{L}$, any $z \in \mathbb{N}^+$, if $\hat{T}_z \subseteq L$, $L(\hat{G}_z) \neq L$, then there is a j such that $z \prec j$, $\hat{T}_z \subseteq \hat{T}_j$ and $L(\hat{G}_j) = L$.

(4) For all $L \in \mathcal{L}$, there is no sequence $(z_j)_{j=1,\dots,m}$ with $m > k+1$ such that $z_j \prec z_{j+1}$ as well as $\hat{T}_{z_j} \subseteq \hat{T}_{z_{j+1}} \subseteq L$, for all $j < m$.

Next we give a characterization of $LIM_k - INF$. For that purpose we define a relation $\prec$ over pairs of sets as follows. Let A, B, C, D be sets. Then $(A, B) \prec (C, D)$ iff $A \subseteq C$, $B \subseteq D$ and $A \subset C$ or $B \subset D$. Note that $\prec$ is computable if A, B, C, D are finitely generable. Now we are ready to state the announced characterization.

Theorem 15. *Let $\mathcal{L}$ be an indexed family of recursive languages and $k \in \mathbb{N}$. Then: $\mathcal{L} \in LIM_k - INF$ iff there are a space of hypotheses $\hat{\mathcal{G}} = (\hat{G}_j)_{j \in \mathbb{N}^+}$ and recursively generable families $(\hat{P}_j)_{j \in \mathbb{N}^+}$ and $(\hat{N}_j)_{j \in \mathbb{N}^+}$ of finite sets such that*

(1) $range(\mathcal{L}) = L(\hat{\mathcal{G}})$

(2) *For all* $j \in \mathbb{N}^+$, $\emptyset \neq \hat{P}_j \subseteq L(\hat{G}_j)$ *and* $\hat{N}_j \subseteq co - L(\hat{G}_j)$

(3) *For all* $L \in \mathcal{L}$ *and* $z \in \mathbb{N}^+$, *if* $\hat{P}_z \subseteq L \neq L(\hat{G}_z)$ *and* $\hat{N}_z \subseteq co - L$, *then there is a* $j \in \mathbb{N}^+$ *such that* $(\hat{P}_z, \hat{N}_z) \prec (\hat{P}_j, \hat{N}_j)$ *as well as* $L = L(\hat{G}_j)$.

(4) *For all* $L \in \mathcal{L}$ *there is no sequence* $(\hat{P}_{z_j}, \hat{N}_{z_j})_{j=1,\dots,m}$ *with* $m > k+1$ *such that* $(\hat{P}_{z_j}, \hat{N}_{z_j}) \prec (\hat{P}_{z_{j+1}}, \hat{N}_{z_{j+1}}) \prec (L, co - L)$, *for all* $j < m$.

5. Conclusions and Open Problems

We have dealt with the learnability of enumerable families $\mathcal{L}$ of uniformly recursive languages in dependence on the number of allowed mind changes. Applying this measure of efficiency we could prove that class preserving learning algorithms are superior to exact learnability. Moreover, in improving Mukouchi's (1992) results we established two new infinite hierarchies. On the other hand, we also proved that even a single additional mind change can neither be compensated by a suitable choice of the space of hypotheses nor by information presentation. Furthermore, we have separated exact and class preserving language learning that avoids overgeneralization. Finally, we presented a complete characterization of language learning in terms of recursively generable finite sets. These theorems resolved the problem that remained open in Mukouchi (1992). Additionally, they offer a new approach to handle overgeneralized hypotheses. However, some problems remained open. It would be very interesting to know how many mind changes are necessary to learn indexed families that cannot be inferred by class preserving conservatively working IIMs.

Acknowledgement

The first author has been supported by the German Ministry for Research and Technology (BMFT) under grant no. 01 IW 101.

We gratefully acknowledge many valuable comments on the preparation of the paper by Yasuhito Mukouchi.

6. References

Angluin, D. (1980), Inductive inference of formal languages from positive data, *Inf. and Control* **45**, 117 - 135.

Angluin, D., and Smith, C.H. (1987), Formal inductive inference, *in* "Encyclopedia of Artificial Intelligence" (St.C. Shapiro, Ed.), Vol. 1, pp. 409 - 418, Wiley-Interscience Publication, New York.

Barzdin, Ya.M., and Freivalds, R.V. (1972), On the prediction of general recursive functions, *Sov. Math. Dokl.* **13**, 1224 - 1228.

Blum, L., and Blum, M. (1975), Toward a mathematical theory of inductive inference, *Inf. and Control* **28**, 122 - 155

Case, J. (1988), The power of vacillation, *in* "Proc. 1st Workshop on Computational Learning Theory," (D. Haussler and L. Pitt, Eds.), pp. 196 -205, Morgan Kaufmann Publishers Inc.

CASE, J., AND LYNES, C. (1982), Machine inductive inference and language identification, *in* "Proc. Automata, Languages and Programming, 9th Colloquium," (M. Nielsen and E.M. Schmidt, Eds.), Lecture Notes in Computer Science Vol. 140, pp. 107 - 115, Springer-Verlag, Berlin.

CASE, J., AND SMITH, C. (1983), Comparison of identification criteria for machine inductive inference, *Theoretical Computer Science* **25**, 193 - 220.

FULK, M.(1990), Prudence and other restrictions in formal language learning, *Inf. and Computation* **85**, 1 - 11.

GASARCH, W.I., AND VELAUTHAPILLAI, M. (1992), Asking questions versus verifiability, *in* "Proc. 3rd International Workshop on Analogical and Inductive Inference," (K.P. Jantke, ed.) Lecture Notes in Artificial Intelligence Vol. 642, pp. 197 - 213, Springer-Verlag, Berlin.

GOLD, E.M. (1967), Language Identification in the Limit, *Inf. and Control* **10**, 447 - 474.

JAIN, S., AND SHARMA, A. (1989), Recursion theoretic characterizations of language learning, Univ. of Rochester, Dept. of Comp. Sci., TR 281.

KAPUR, S., AND BILARDI, G. (1992), Language learning without overgeneralization, *in* "Proc. 9th Annual Symposium on Theoretical Aspects of Computer Science," (A. Finkel and M. Jantzen, Eds.), Lecture Notes in Computer Science Vol. 577, pp. 245 - 256, Springer-Verlag, Berlin.

LANGE, S., AND ZEUGMANN, T. (1992), Types of monotonic language learning and their characterization, *in* "Proc. 5th Annual ACM Workshop on Computational Learning Theory," pp. 377 - 390, ACM Press.

LANGE, S.,ZEUGMANN, T., AND KAPUR, S. (1992), Class preserving monotonic language learning, GOSLER–Report 14/92, FB Mathematik und Informatik, TH Leipzig.

MUKOUCHI, Y. (1992), Inductive Inference with Bounded Mind Changes, *in* Proc. "Algorithmic Learning Theory," October 1992, Tokyo, Japan, JSAI.

OSHERSON, D., STOB, M., AND WEINSTEIN, S. (1986), "Systems that Learn, An Introduction to Learning Theory for Cognitive and Computer Scientists," MIT-Press, Cambridge, Massachusetts.

SHINOHARA, T. (1990), Inductive Inference from Positive Data is Powerful, *in* "Proc. 3rd Annual Workshop on Computational Learning Theory," (M. Fulk and J. Case, Eds.), pp. 97 - 110, Morgan Kaufmann Publishers Inc.

WIEHAGEN, R. (1977), Identification of formal languages, *in* "Proc. Mathematical Foundations of Computer Science," (J. Gruska, Ed.), Lecture Notes in Computer Science Vol. 53, pp. 571 - 579, Springer-Verlag, Berlin.

WIEHAGEN, R., FREIVALDS, R., AND KINBER, B. (1984), On the power of probabilistic strategies in inductive inference, *Theoretical Computer Science* **28**, 111 - 133.

ZEUGMANN, T. (1983), A–posteriori characterizations in inductive inference of recursive functions, *J. of Inf. Processing and Cybernetics (EIK)* **19**, 559 - 594.