On the non-existence of maximal inference degrees for language identification

Sanjay Jain
Institute of Systems Science
National University of Singapore
Singapore 0511
Republic of Singapore
sanjay@iss.nus.sg

Arun Sharma
School of Computer Science and Engineering
University of New South Wales
Sydney, NSW 2033
Australia
arun@cs.unsw.oz.au

Abstract

Identification of grammars (r. e. indices) for recursively enumerable languages from positive data by algorithmic devices is a well studied problem in learning theory. The present paper considers identification of r. e. languages by machines that have access to membership oracles for noncomputable sets. It is shown that for any set A there exists another set B such that the collections of r. e. languages that can be identified by machines with access to a membership oracle for B is strictly larger than the collections of r. e. languages that can be identified by machines with access to a membership oracle for A. In other words, there is no maximal inference degree for language identification.

Keywords: Machine learning, inductive inference, methodology of science, recursion theory

1 Introduction

A model for a subject learning a concept could be described thus. At any given time, the subject receives some finite data about the concept. Based on this data, the subject conjectures a hypothesis about the concept. Availability of more data may cause the subject to revise its hypotheses. The subject is said to learn the concept just in case the sequence of hypotheses conjectured by the subject converges to a fixed hypothesis and this final hypothesis is a correct representation of the concept. This is essentially the theme of Gold's [9] identification in the limit when the subject is an algorithmic device.

Identification of programs for computable functions from their graphs and identification of grammars (r. e. indices) for recursively enumerable languages from positive data are two extensively studied problems in the above framework. Lately, with a view to gain a deeper insight into the process of identification, there has been considerable interest in investigating identification by devices that have access to membership oracles for noncomputable sets [1, 8, 7]. This study has mainly concentrated on identification of functions; a representative result being that there exists a single machine with access to an oracle for the halting set that identifies every computable function [1].

The present paper investigates identification of recursively enumerable languages from positive data by machines with access to membership oracles for noncomputable sets. It is shown that for this problem, there is no maximal inference degree in the sense that for every set A there exists another set B such that the collections of r. e. languages that can be identified by machines with access to a membership oracle for B is strictly larger than the collections of r. e. languages that can be identified by machines with access to a membership oracle for A. Similar results are shown about more general identification criteria, namely vacillatory language identification and behaviorally correct language identification.

We now proceed formally. In Section 2 and Section 3, we introduce the notation and definitions, respectively. Section 4 contains the results.

2 Notation

Recursion-theoretic concepts not explained below are treated in [16]. N is the set of natural numbers, $\{0, 1, 2, \ldots\}$; N^+ is the set of positive integers, $\{1, 2, 3, \ldots\}$. The symbols a, b, i, j, m, n, s, t, x, and y, with or without decorations (decorations are subscripts, superscripts and the like), range over natural numbers unless otherwise specified. $\subseteq, \subset$, denote subset, proper subset, respectively. $\in$ denotes 'element of.' $\emptyset$ denotes the empty set. A, B, C, D, S range over sets of natural numbers; we usually reserve D to denote finite sets. We denote the cardinality of the set S by $\mathrm{card}(S)$. $A \boldsymbol{\Delta} B$ denotes the symmetric difference of A and B, i.e. $(A - B) \cup (B - A)$. The symbol $*$ denotes 'finite but unbounded' such that $(\forall n)[n < * < \infty]$. For $a \in N \cup \{*\}$, we say that $A =^a B$ iff $\mathrm{card}(A \boldsymbol{\Delta} B) \leq a$. Thus, $A =^* B$ means that $\mathrm{card}(A \boldsymbol{\Delta} B)$ is finite. $\max(\ ), \min(\ )$ denote the maximum and minimum of a set, respectively. By convention $\max(\emptyset) = 0$ and $\min(\emptyset) = \infty$.

For a partial recursive function η, $\mathrm{domain}(\eta)$ denotes the domain of η. $\downarrow$ denotes defined. $\uparrow$ denotes undefined.

L, with or without decorations, ranges over recursively enumerable (r.e.) subsets of N. $\mathcal{E}$ denotes the class of all r.e. languages. $\mathcal{L}$, with or without decorations, ranges over subsets of $\mathcal{E}$. φ denotes a standard acceptable programming system (also referred to as standard acceptable numbering) [15, 16]. φ_i denotes the partial recursive function computed by the i^{th} program in the standard acceptable programming system φ. Φ denotes an arbitrary fixed Blum complexity measure [3, 10] for the φ-system. W_i denotes the domain of φ_i. W_i is, then, the r.e. set/language ($\subseteq N$) accepted by φ-program i. We also think of the r.e. index i for W_i as a grammar for accepting W_i. $W_{i,s}$ denotes the set $\{x \leq s \mid \Phi_i(x) \leq s\}$.

FIN denotes the set of all *non-empty* finite sets. We do not include the empty set in our definition of **FIN** to facilitate the presentation of our proof of Theorem 1. Note that an effective canonical indexing for **FIN** can be easily obtained. For $n \in N$, Σ_n^C and Π_n^C denote the n^{th} level of the Kleene hierarchy relativized with respect to C [16]. $\langle i, j \rangle$ stands for an arbitrary computable one to one encoding of all pairs of natural numbers onto N [16].

The quantifiers '$\stackrel{\infty}{\forall}$' and '$\stackrel{\infty}{\exists}$' mean 'for all but finitely many' and 'there exist infinitely many', respectively.

3 Preliminaries

In this section, we briefly adopt notions and results from the recursion-theoretic machine learning literature to learning with oracles. We first introduce a notion that facilitates discussion about elements of a language being fed to a learning machine.

A *finite sequence* is a mapping from $\{x \mid x < a\}$, for some $a \in N$, into $(N \cup \{\#\})$. We let σ and τ, with or without decorations, range over finite sequences. The *content* of a finite sequence σ, denoted $\text{content}(\sigma)$, is the set of natural numbers in the range of σ. Intuitively, #'s represent pauses in the presentation of data. The *length* of σ, denoted $|\sigma|$, is the number of elements in the domain of σ. $\sigma \subseteq \tau$ means that σ is an initial sequence of τ. SEQ denotes the set of all finite sequences.

An *oracular learning machine* is an algorithmic device which has access to a membership oracle and which computes a mapping from SEQ into N. Without loss of generality, we assume our oracle learning machines to be total with respect to any oracle. We let $\mathbf{M}$, with or without decorations, range over oracular learning machines. $\mathbf{M}^A(\sigma)$ denotes the hypothesis conjectured by machine $\mathbf{M}$ with access to a membership oracle for A on input σ.

A *text* T for a language L is a mapping from N into $(N \cup \{\#\})$ such that L is the set of natural numbers in the range of T. The *content* of a text T, denoted $\text{content}(T)$, is the set of natural numbers in the range of T. We let T, with or without decorations, range over texts. $T(n)$ denotes the n^{th} element of text T. $T[n]$ denotes the finite initial sequence of T with length n. Hence, $\text{domain}(T[n]) = \{x \mid x < n\}$. The reader should note that $T[n]$ does not include $T(n)$. $\sigma \subseteq T$ means that σ is an initial sequence of T.

We now describe three distinct criteria for successful identification of languages from texts.

3.1 Explanatory Language Identification

Suppose $\mathbf{M}$ is an oracular learning machine and T is a text. Let $A \subseteq N$. $\mathbf{M}^A(T)\downarrow$ (read: $\mathbf{M}^A(T)$ *converges*) $\iff$ $(\exists i)(\overset{\infty}{\forall} n)[\mathbf{M}^A(T[n]) = i]$. If $\mathbf{M}^A(T)\downarrow$, then $\mathbf{M}^A(T)$ is defined to be the unique i such that $(\overset{\infty}{\forall} n)[\mathbf{M}^A(T[n]) = i]$; otherwise we say that $\mathbf{M}^A(T)$ diverges (written: $\mathbf{M}^A(T)\uparrow$).

Definition 1 Let $A \subseteq N$. Let $a \in N \cup \{*\}$.

1. $\mathbf{M}$ $\mathbf{O}^A\mathbf{TxtEx}^a$-*identifies* L (written: $L \in \mathbf{O}^A\mathbf{TxtEx}^a(\mathbf{M})$) $\iff$

   $(\forall \text{ texts } T \text{ for } L)[\mathbf{M}^A(T)\downarrow \ \wedge \ W_{\mathbf{M}^A(T)} =^a L]$.

2. $\mathbf{O}^A\mathbf{TxtEx}^a = \{\mathcal{L} \subseteq \mathcal{E} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{O}^A\mathbf{TxtEx}^a(\mathbf{M})]\}$.

If A is recursive then $\mathbf{O}^A\mathbf{TxtEx}^a$ is the same class as the class $\mathbf{TxtEx}^a$ defined in the literature [6]. $\mathbf{O}^A\mathbf{TxtEx}^0$ is usually written $\mathbf{O}^A\mathbf{TxtEx}$.

3.2 Vacillatory Language Identification

We now extend the notion of oracular identification to vacillatory learning [4, 13]. Intuitively, according to this criteria of success, instead of converging to a single index for the language, the machine converges to a finite set of indices, all of which are correct.

We first introduce the notion of an oracular learning machine finitely converging on a text. Let $\mathbf{M}$ be an oracular learning machine and T be a text. Let $A \subseteq N$. $\mathbf{M}^A(T)$ *finitely-converges* (written: $\mathbf{M}^A(T)\Downarrow$) $\iff$ $\{\mathbf{M}^A(\sigma) \mid \sigma \subset T\}$ is finite, otherwise we say that $\mathbf{M}^A(T)$ *finitely-diverges* (written: $\mathbf{M}^A(T)\Uparrow$). If $\mathbf{M}^A(T)\Downarrow$, then $\mathbf{M}^A(T)$ is defined as P, where $P = \{i \mid (\stackrel{\infty}{\exists}\ \sigma \subset T)[\mathbf{M}^A(\sigma) = i]\}$.

Definition 2 Let $A \subseteq N$. Let $b \in N^+ \cup \{*\}$. Let $a \in N \cup \{*\}$.

1. $\mathbf{M}$ $\mathbf{O}^A\mathbf{TxtFex}^a_b$-*identifies* L (written: $L \in \mathbf{O}^A\mathbf{TxtFex}^a_b(\mathbf{M})$) $\iff$

   $(\forall$ texts T for $L)(\exists P \mid \mathrm{card}(P) \leq b\ \wedge\ (\forall i \in P)[W_i =^a L])[\mathbf{M}^A(T)\Downarrow \wedge \mathbf{M}^A(T) = P]$.

2. $\mathbf{O}^A\mathbf{TxtFex}^a_b = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{O}^A\mathbf{TxtFex}^a_b(\mathbf{M})]\}$.

If A is recursive then $\mathbf{O}^A\mathbf{TxtFex}^a_b$ is the same class as the class $\mathbf{TxtFex}^a_b$ defined in the literature [13, 4].

3.3 Behaviorally Correct Language Identification

We extend the notion of oracular identification to $\mathbf{TxtBc}^a$-identification [6, 14, 13].

Definition 3 Let $A \subseteq N$. Let $a \in N \cup \{*\}$.

1. $\mathbf{M}$ $\mathbf{O}^A\mathbf{TxtBc}^a$-*identifies* L (written: $L \in \mathbf{O}^A\mathbf{TxtBc}^a(\mathbf{M})$) $\iff$

   $(\forall$ texts T for $L)(\stackrel{\infty}{\forall}\ n)[W_{\mathbf{M}^A(T[n])} =^a L]$.

2. $\mathbf{O}^A\mathbf{TxtBc}^a = \{\mathcal{L} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{O}^A\mathbf{TxtBc}^a(\mathbf{M})]\}$.

4 Results

4.1 $\mathbf{O}^A\mathbf{TxtEx}^a$-identification

The following lemma is a relativization of a central result about $\mathbf{TxtEx}^a$-identification [2, 12].

Lemma 1 *Let $a \in N \cup \{*\}$. Let $A \subseteq N$. If $\mathbf{M}$ $\mathbf{O}^A\mathbf{TxtEx}^a$-identifies L, then there exists a σ such that*

1. $\mathrm{content}(\sigma) \subseteq L$,

2. $W_{\mathbf{M}^A(\sigma)} =^a L$, *and*

3. $(\forall \tau \mid \sigma \subseteq \tau\ \wedge\ \mathrm{content}(\tau) \subseteq L)[\mathbf{M}^A(\sigma) = \mathbf{M}^A(\tau)]$.

σ is referred to as an $\mathbf{O}^A\mathbf{TxtEx}^a$-locking sequence for $\mathbf{M}$ on L.

Theorem 1 *Let $a \in N$. Consider a sequence of sets, $A_0, A_1, A_2, \ldots$. Consider the set, $B = \{\langle i, x\rangle \mid x \in A_i\ \wedge\ i \in N\}$. Let C be such that $\{i \mid A_i$ is finite $\} \notin \Sigma_2^C$. Then there exists a class of languages $\mathcal{L}$ such that $\mathcal{L} \in \mathbf{O}^B\mathbf{TxtEx} - \mathbf{O}^C\mathbf{TxtEx}^a$.*

PROOF. Consider the following languages and classes of languages:
For $i \in N$, $S \subseteq N$, let

$$L_{i,S} \text{ denote } \{\langle i, x\rangle \mid x \in S\}$$

For $i \in N$, let

$$\mathcal{L}_i = \begin{cases} \{L_{i,N}\}, & \text{if } A_i \text{ is finite;} \\ \{L_{i,D} \mid D \in \mathbf{FIN}\}, & \text{otherwise.} \end{cases}$$

Then, let

$$\mathcal{L} = \bigcup_{i \in N} \mathcal{L}_i.$$

Note that according to our definition **FIN** consists of all the finite sets except the empty set. We leave out empty set for ease of writing the proof of Claim 1 below. For $D \in \mathbf{FIN}$, let $G_{i,D}$ denote a grammar for $L_{i,D}$ such that $G_{i,D}$ can be found effectively from i and D. Let $G_{i,N}$ denote a grammar for $L_{i,N}$.

Now the theorem follows from Claims 1 and 2 below.

Claim 1 $\mathcal{L} \in \mathbf{O}^B\mathbf{TxtEx}$.

PROOF. We construct a machine $\mathbf{M}$ that $\mathbf{O}^B\mathbf{TxtEx}$-identifies $\mathcal{L}$. Suppose T is a text for $L \in \mathcal{L}$. Then, $\mathbf{M}$, on seeing the initial sequence $T[x]$, will make use of a number of functions to issue its hypothesis $\mathbf{M}^B(T[x])$. Each of these functions will either be recursive or recursive in B.

It is clear that elements of L are of the form $\langle i, n\rangle$ for some fixed i; if A_i is finite then n ranges over N, else n ranges over some finite set. We refer to i as the *key* of L. Formally,

$$\text{key}(T[x]) = \min(\{x\} \cup \{j \mid (\exists y)[\langle j, y\rangle \in \text{content}(T[x])]\}).$$

We now define three functions g_1, g_2, and g_3.

$$g_1(T[x]) = \text{card}(\{y \leq x \mid \langle \text{key}(T[x]), y\rangle \in B\})$$

$$g_2(T[x]) = \text{card}(\text{content}(T[x]))$$

The information provided by g_1 and g_2 is collected in the following function g_3. If L is $L_{i,N}$, then $g_3(T[x])$ is 0 for all but finintely many x. If L is finite, then $g_3(T[x])$ is 1 for all but finitely many x. Clearly, g_3 is recursive in B.

$$g_3(T[x]) = \begin{cases} 1, & \text{if } g_1(T[x]) \geq g_2(T[x]); \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to verify that the following two statements summarize the properties of g_1, g_2, and g_3.

1. A_i is finite $\iff$ $L = L_{i,N}$ $\iff$ $(\overset{\infty}{\forall} x)[g_2(T[x]) > g_1(T[x])]$ $\iff$ $(\overset{\infty}{\forall} x)[g_3(T[x]) = 0]$.

2. A_i is infinite $\iff$ $L = L_{i,D}$ for some $D \in \mathbf{FIN}$ $\iff$ $(\stackrel{\infty}{\forall} x)[g_1(T[x]) \geq g_2(T[x])]$ $\iff$ $(\stackrel{\infty}{\forall} x)[g_3(T[x]) = 1]$.

Now, let $\mathbf{M}$ be defined as follows:

$$\mathbf{M}^B(T[x]) = \begin{cases} G_{\mathrm{key}(T[x]),N}, & \text{if } g_3(T[x]) = 0 \ ; \\ G_{\mathrm{key}(T[x]),D}, & \text{if } g_3(T[x]) = 1 \ \wedge \ D = \{x \mid \langle \mathrm{key}(T[x]), x\rangle \in \mathrm{content}(T[x])\}. \end{cases}$$

It is easy to verify that $\mathbf{M}$ indeed $\mathbf{O}^B\mathbf{TxtEx}$-identifies $\mathcal{L}$. □

Claim 2 $\mathcal{L} \notin \mathbf{O}^C\mathbf{TxtEx}^a$.

Proof. Suppose by way of contradiction $\mathcal{L} \in \mathbf{O}^C\mathbf{TxtEx}^a$. Then, we show that the set $\{i \mid A_i \text{ is finite }\} \in \Sigma_2^C$. Suppose $\mathbf{M}$ $\mathbf{O}^C\mathbf{TxtEx}^a$-identifies $\mathcal{L}$.

Let $i \in N$. Recall that $\mathcal{L}_i = \{L_{i,N}\}$ if A_i is finite. Thus, by Lemma 1, there is an $\mathbf{O}^C\mathbf{TxtEx}^a$-locking sequence for $\mathbf{M}$ on $L_{i,N}$ if A_i is finite. On the other hand, if A_i is infinite, then for every σ such that $\mathrm{content}(\sigma) \subseteq L_{i,N}$, there exist infinitely many $L \in \mathcal{L}_i$ such that $L \neq^a W_{\mathbf{M}(\sigma)}$. Thus we have,

$$A_i \text{ is finite } \iff (\exists \sigma \mid \mathrm{content}(\sigma) \subseteq L_{i,N})(\forall \tau \mid \sigma \subseteq \tau \ \wedge \ \mathrm{content}(\tau) \subseteq L_{i,N})[\mathbf{M}^C(\sigma) = \mathbf{M}^C(\tau)].$$

Thus, $\{i \mid A_i \text{ is finite }\} \in \Sigma_2^C$. A contradiction. □
▮ (Theorem 1)

We would like to note that the above theorem can be extended to show that there exists a class of languages $\mathcal{L}'$ such that $\mathcal{L}' \in \mathbf{O}^B\mathbf{TxtEx} - \mathbf{O}^C\mathbf{TxtEx}^*$. A proof can be worked out on the lines of the above proof by taking $\mathcal{L}'$ to be a cylindrification of $\mathcal{L}$. We leave details to the reader.

The following Corollary is essentially a restatement of Osherson, Stob and Weinstein's result that $\mathcal{E}$ cannot be $\mathbf{TxtEx}^*$-identified by even noncomputable learning machines [12]. Osherson, Stob, and Weinstein's proof follows from their observation that Blum and Blum's [2] locking sequence construction also holds for noncomputable devices.

Corollary 1 $(\forall A)[\mathcal{E} \notin \mathbf{O}^A\mathbf{TxtEx}^*]$.

The following Corollary says that there doesn't exist a maximal inference degree for $\mathbf{TxtEx}^a$-identification.

Corollary 2 *Let* $a \in N \cup \{*\}$. $(\forall A)(\exists B)[\mathbf{O}^A\mathbf{TxtEx}^a \subset \mathbf{O}^B\mathbf{TxtEx}^a]$.

Kummer and Stephan [11] have recently refined our result.

4.2 $\mathbf{O}^A\mathbf{TxtFex}_b^a$-identification

We now show that an analog of Theorem 1 holds for $\mathbf{O}^A\mathbf{TxtFex}_b^a$-identification. As in the case of Theorem 1, we only give a proof of the cases where $a \in N$; the $a = *$ case can similarly be obtained by considering a cylindrification of class used in the proof of other cases. But first, we present the following lemma, which is a relativization of a result about $\mathbf{TxtFex}_b^a$-identification [13, 4].

Lemma 2 *Let $A \subseteq N$. Let $a \in N \cup \{*\}$. Let $b \in N^+ \cup \{*\}$. If* $\mathbf{M}$ $\mathbf{O}^A\mathbf{TxtFex}^a_b$*-identifies L, then there exists a $\sigma \in SEQ$ and $D \in$ **FIN** such that*

1. content$(\sigma) \subseteq L$,
2. card$(D) \leq b \ \wedge \ (\forall i \in D)[W_i =^a L]$, *and*
3. $(\forall \tau \mid \sigma \subseteq \tau \ \wedge \ \text{content}(\tau) \subseteq L)[\mathbf{M}^A(\tau) \in D]$.

Theorem 2 *Let $a \in N$. Consider a sequence of sets, $A_0, A_1, A_2, \ldots$. Consider the set, $B = \{\langle i, x\rangle \mid x \in A_i \ \wedge \ i \in N\}$. Let C be such that $\{i \mid A_i$ is finite $\} \notin \Sigma_2^C$. Then there exists a class of languages $\mathcal{L}$ such that $\mathcal{L} \in \mathbf{O}^B\mathbf{TxtEx} - \mathbf{O}^C\mathbf{TxtFex}^a_*$.*

Proof. Consider the collection of languages $\mathcal{L}$ from the proof of Theorem 1. We only need to show that $\mathcal{L} \notin \mathbf{O}^C\mathbf{TxtFex}^a_*$.

Suppose by way of contradiction, $\mathcal{L} \in \mathbf{O}^C\mathbf{TxtFex}^a_*$. We show that $\{i \mid A_i$ is finite $\} \in \Sigma_2^C$. Suppose $\mathbf{M}$ $\mathbf{O}^C\mathbf{TxtFex}^a_*$-identifies $\mathcal{L}$. Now, using Lemma 2 and proceeding as in the proof of Claim 2, we have, A_i is finite $\iff$ $(\exists \sigma \mid \text{content}(\sigma) \subseteq L_{i,N})(\exists$ a finite set $D)(\forall \tau \mid \sigma \subseteq \tau \ \wedge \ \text{content}(\tau) \subseteq L_{i,N})[\mathbf{M}^C(\tau) \in D]$. Thus, $\{i \mid A_i$ is finite $\} \in \Sigma_2^C$. A contradiction. ∎

As already stated, we leave it to the reader to extend the above result for the $a = *$ case. The following corollary is essentially a restatement of Osherson, Stob and Weinstein's result that $\mathcal{E}$ cannot be $\mathbf{TxtFex}^*_*$-identified by even noncomputable learning machines [12].

Corollary 3 $(\forall A)[\mathcal{E} \notin \mathbf{O}^A\mathbf{TxtFex}^*_*]$.

The following corollary says that there doesn't exist a maximal inference degree for $\mathbf{TxtFex}^a_b$-identification.

Corollary 4 *Let $a \in N \cup \{*\}$. Let $b \in N^+ \cup \{*\}$.* $(\forall A)(\exists B)[\mathbf{O}^A\mathbf{TxtFex}^a_b \subset \mathbf{O}^B\mathbf{TxtFex}^a_b]$.

4.3 $\mathbf{O}^A\mathbf{TxtBc}^a$-identification

We now show that a similar result also holds for behaviorally correct language identification. The following lemma is a relativized version of a result about $\mathbf{TxtBc}^a$-identification [13, 6].

Lemma 3 *If* $\mathbf{M}$ $\mathbf{O}^A\mathbf{TxtBc}^a$*-identifies L, then there exists a σ such that*

1. content$(\sigma) \subseteq L$,
2. $W_{\mathbf{M}^A(\sigma)} =^a L$, *and*
3. $(\forall \tau \mid \sigma \subseteq \tau \ \wedge \ \text{content}(\tau) \subseteq L)[W_{\mathbf{M}^A(\tau)} =^a L]$.

Theorem 3 *Consider a sequence of sets, $A_0, A_1, A_2, \ldots$. Consider the set, $B = \{\langle i, x\rangle \mid x \in A_i \ \wedge \ i \in N\}$. Let C be such that $\{i \mid A_i$ is finite $\} \notin \Sigma_3^C$. Then there exists a class of languages $\mathcal{L}$ such that $\mathcal{L} \in \mathbf{O}^B\mathbf{TxtEx} - \mathbf{O}^C\mathbf{TxtBc}^*$.*

PROOF. Again, consider the collection of languages $\mathcal{L}$ from the proof of Theorem 1. We only need to show that $\mathcal{L} \notin \mathbf{O}^C\mathbf{TxtBc}^*$.

Suppose by way of contradiction, $\mathcal{L} \in \mathbf{O}^C\mathbf{TxtBc}^*$. We then show that $\{i \mid A_i \text{ is finite }\} \in \Sigma_3^C$. Suppose $\mathbf{M}$ $\mathbf{O}^C\mathbf{TxtBc}^*$-identifies $\mathcal{L}$. Now, A_i is finite implies $\mathbf{M}$ $\mathbf{O}^C\mathbf{TxtBc}^*$-identifies $L_{i,N}$, which implies (by Lemma 3) that

$(\exists\sigma \mid \text{content}(\sigma) \subseteq L_{i,N})(\forall\tau \mid \sigma \subseteq \tau \ \wedge\ \text{content}(\tau) \subseteq L_{i,N})[W_{\mathbf{M}^C(\tau)} =^* L_{i,N}]$,

which in turn implies that

$(\exists\sigma \mid \text{content}(\sigma) \subseteq L_{i,N})(\forall\tau \mid \sigma \subseteq \tau \ \wedge\ \text{content}(\tau) \subseteq L_{i,N})[W_{\mathbf{M}^C(\tau)} \text{ is infinite}]$.

Also,

$(\exists\sigma \mid \text{content}(\sigma) \subseteq L_{i,N})(\forall\tau \mid \sigma \subseteq \tau \ \wedge\ \text{content}(\tau) \subseteq L_{i,N})[W_{\mathbf{M}^C(\tau)} \text{ is infinite}]$

implies that $\mathbf{M}$ does not $\mathbf{O}^C\mathbf{TxtBc}^*$-identify any $L_{i,D}$ such that $\text{content}(\sigma) \subseteq L_{i,D}$, thereby implying that A_i is finite.

Thus, we have,

$$A_i \text{ is finite} \iff (\exists\sigma \mid \text{content}(\sigma) \subseteq L_{i,N})(\forall\tau \mid \sigma \subseteq \tau \ \wedge\ \text{content}(\tau) \subseteq L_{i,N})[W_{\mathbf{M}^C(\tau)} \text{ is infinite}].$$

Now, it is easy to see that $[W_{\mathbf{M}^C(\tau)}$ is infinite $]$ is Π_2^C (since, $[W_{\mathbf{M}^C(\tau)}$ is infinite $] \iff (\forall m)(\exists n > m)(\exists t)[n \in W_{\mathbf{M}^C(\tau),t}]$). Hence, $\{i \mid A_i \text{ is finite }\} \in \Sigma_3^C$. A contradiction. ∎

The following corollary is essentially a restatement of Osherson, Stob and Weinstein's result that $\mathcal{E}$ cannot be $\mathbf{TxtBc}^*$-identified by even noncomputable learning machines [12].

Corollary 5 $(\forall A)[\mathcal{E} \notin \mathbf{O}^A\mathbf{TxtBc}^*]$.

The following corollary says that there doesn't exist a maximal inference degree for $\mathbf{TxtBc}^a$-identification.

Corollary 6 *Let* $a \in N \cup \{*\}$. $(\forall A)(\exists B)[\mathbf{O}^A\mathbf{TxtBc}^a \subset \mathbf{O}^B\mathbf{TxtBc}^a]$.

5 Acknowledgements

We wish to thank the referees for many helpful comments, for pointing out an oversight in an earlier version of the paper, and for suggesting that we also consider **TxtBc**-identification by oracular machines.

References

[1] L. Adleman and M. Blum. Inductive inference and unsolvability. *Journal of Symbolic Logic*, 56(3):891–900, 1991.

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

[3] M. Blum. A machine independent theory of the complexity of recursive functions. *Journal of the ACM*, 14:322–336, 1967.

[4] J. Case. The power of vacillation. In D. Haussler and L. Pitt, editors, *Proceedings of the Workshop on Computational Learning Theory*, pages 133–142. Morgan Kaufmann Publishers, Inc., 1988. Expanded in [5].

[5] J. Case. The power of vacillation in language learning. Technical Report 93-08, University of Delaware, 1992. Expands on [4]; journal article under review.

[6] J. Case and C. Lynes. Machine inductive inference and language identification. In M. Nielsen and E. M. Schmidt, editors, *Proceedings of the 9th International Colloquium on Automata, Languages and Programming*, volume 140, pages 107–115. Springer-Verlag, Berlin, 1982.

[7] P Cholak, R. Downey, L. Fortnow, W. Gasarch, Knber E., M. Kummer, S. Kurtz, and T. Slaman. Degrees of inferability. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, Pennsylvania*, pages 180–192. ACM Press, July 1992.

[8] W. I. Gasarch and M. B. Pleszkoch. Learning via queries to an oracle. In R. Rivest, D. Haussler, and M. K. Warmuth, editors, *Proceedings of the Second Annual Workshop on Computational Learning Theory, Santa Cruz, California*, pages 214–229. Morgan Kaufmann Publishers, Inc., 1989.

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

[10] J. Hopcroft and J. Ullman. *Introduction to Automata Theory Languages and Computation*. Addison-Wesley Publishing Company, 1979.

[11] M. Kummer and F. Stephan. On the structure of degrees of inferability. In *Proceedings of the Sixth Annual Workshop on Computational Learning Theory, Santa Cruz, California*. ACM Press, July 1993.

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

[13] D. Osherson and S. Weinstein. Criteria of language learning. *Information and Control*, 52:123–138, 1982.

[14] D. Osherson and S. Weinstein. A note on formal learning theory. *Cognition*, 11:77–88, 1982.

[15] H. Rogers. Gödel numberings of partial recursive functions. *Journal of Symbolic Logic*, 23:331–341, 1958.

[16] H. Rogers. *Theory of Recursive Functions and Effective Computability*. McGraw Hill, New York, 1967. Reprinted by MIT Press, Cambridge, Massachusetts in 1987.