The Intrinsic Complexity of Language Identification

Sanjay Jain
Department of Information Systems and Computer Science
National University of Singapore
Singapore 119260, Republic of Singapore
Email: sanjay@iscs.nus.sg

Arun Sharma
School of Computer Science and Engineering
The University of New South Wales
Sydney, NSW 2052, Australia
Email: arun@cse.unsw.edu.au

August 24, 2010

Abstract

A new investigation of the complexity of language identification is undertaken using the notion of reduction from recursion theory and complexity theory. The approach, referred to as the intrinsic complexity of language identification, employs notions of 'weak' and 'strong' reduction between learnable classes of languages. The intrinsic complexity of several classes is considered and the results agree with the intuitive difficulty of learning these classes. Several complete classes are shown for both the reductions and it is also established that the weak and strong reductions are distinct.

An interesting result is that the self referential class of Wiehagen in which the minimal element of every language is a grammar for the language and the class of pattern languages introduced by Angluin are equivalent in the strong sense.

This study has been influenced by a similar treatment of function identification by Freivalds, Kinber, and Smith.

1   Introduction

The present paper introduces a novel way to look at the difficulty of learning collections of languages from positive data. Most studies on feasibility issues in learning have concentrated on the complexity of the learning algorithm. The present paper describes a model which provides an insight into why certain classes are more easily learned than others. Our model adopts a similar study in the context of learning functions by Freivalds [9], and by Freivalds, Kinber, and Smith [10]. The main idea of the approach is to introduce reductions between learnable classes of languages. If a collection of languages, $\mathcal{L}_1$, can be reduced to another collection of languages, $\mathcal{L}_2$, then the learnability of $\mathcal{L}_1$ is no more difficult than that of $\mathcal{L}_2$. We illustrate this idea with the help of simple examples.

Consider the following collections of languages over N, the set of natural numbers.

$SINGLE = \{L \mid L \text{ is singleton }\}$.

$COINIT = \{L \mid (\exists n)[L = \{x \mid x \geq n\}]\}$.

$FIN = \{L \mid \text{cardinality of } L \text{ is finite }\}$.

So, $SINGLE$ is the collection of all singleton languages, $COINIT$ is the collection of languages that contain all natural numbers except a finite initial segment, and FIN is the collection of all finite languages. Clearly, each of these three classes is identifiable in the limit from only positive data. For example, a machine $\mathbf{M}_1$ that upon encountering the first data element, say n, keeps on emitting a grammar for the singleton language $\{n\}$ identifies $SINGLE$. A machine $\mathbf{M}_2$ that, at any given time, finds the minimum element among the data seen so far, say n, and emits a grammar for the language $\{x \mid x \geq n\}$ can easily be seen to identify $COINIT$. Similarly, a machine $\mathbf{M}_3$ that continually outputs a grammar for the finite set of data seen so far identifies FIN.

Now, although all three of these classes are identifiable, it can be argued that they present learning problems of varying difficulty. One way to look at the difficulty is to ask the question, "At what stage in the processing of the data can a learning machine confirm its success?" In the case of $SINGLE$, the machine can be confident of success as soon as it encounters the first data element. In the case of $COINIT$, the machine cannot always be sure that it has identified the language. However, at any stage after it has seen the first data element, the machine can provide an upper bound on the number of mind changes that the machine will make before converging to a correct grammar. For example, if at some stage the minimum element seen is m, then $\mathbf{M}_2$ will make no more than m mind changes because it changes its mind only if a smaller element appears. In the case of FIN, the learning machine can neither be confident about its success nor can it, at any stage, provide an upper bound on the number of further mind changes that it may have to undergo before it is rewarded with success. Clearly, these three collections of languages pose learning problems of varying difficulty where $SINGLE$ appears to be the least difficult to learn and FIN is seen to be the most difficult to learn with $COINIT$ appearing to be of intermediate difficulty. The model described in the present paper captures this gradation in difficulty of various identifiable collections of languages. Following Freivalds, Kinber, and Smith [10], we refer to such a notion of difficulty as "intrinsic complexity."

We next present an informal description of the reductions that are central to our

analysis of the intrinsic complexity of language learning. To facilitate our discussion, we first present some technical notions about language learning.

Informally, a *text* for a language L is just an infinite sequence of elements, with possible repetitions, of all and only the elements of L. A text for L is thus an abstraction of the presentation of positive data about L. A learning machine is essentially an algorithmic device. Elements of a text are sequentially fed to a learning machine one element at a time. The learning machine, as it receives elements of the text, outputs an infinite sequence of grammars. Several criteria for the learning machine to be successful on a text have been proposed. In the present paper we will concern ourselves with Gold's [11] criterion of identification in the limit (referred to as **TxtEx**-identification). A sequence of grammars, $G = g_0, g_1, \ldots$, is said to *converge to g* just in case, for all but finitely many n, $g_n = g$. We say that the sequence of grammars, $G = g_0, g_1, \ldots$, converges just in case there exists a g such that G converges to g; if no such g exists, then we say that the sequence G diverges. We say that **M** converges on T (to g), if the sequence of grammars emitted by **M** on T converges (to g). If the sequence of grammars emitted by the learning machine converges to a correct grammar for the language whose text is fed to the machine, then the machine is said to **TxtEx**-*identify* the text. A machine is said to **TxtEx**-*identify* a language just in case it **TxtEx**-identifies each text for the language.

It is also useful to call an infinite sequence of grammars, $g_0, g_1, g_2, \ldots$, **TxtEx**-admissible for a text T just in case the sequence of grammars converges to a single correct grammar for the language whose text is T.

Our reductions are based on the idea that for a collection of languages $\mathcal{L}$ to be reducible to $\mathcal{L}'$, we should be able to transform texts T for languages in $\mathcal{L}$ to texts T' for languages in $\mathcal{L}'$ and further transform **TxtEx**-admissible sequences for T' into **TxtEx**-admissible sequences for T. This is achieved with the help of two enumeration operators. Informally, enumeration operators are algorithmic devices that map infinite sequences of objects (for example, texts and infinite sequences of grammars) into infinite sequences of objects. The first operator, Θ, transforms texts for languages in $\mathcal{L}$ into texts for languages in $\mathcal{L}'$. The second operator, Ψ, behaves as follows: if Θ transforms a text T for some language in $\mathcal{L}$ into text T' (for some language in $\mathcal{L}'$), then Ψ transforms **TxtEx**-admissible sequences for T' into **TxtEx**-admissible sequences for T.

To see that the above satisfies the intuitive notion of reduction consider collections $\mathcal{L}$ and $\mathcal{L}'$ such that $\mathcal{L}$ is reducible to $\mathcal{L}'$. We now argue that if $\mathcal{L}'$ is identifiable then so is $\mathcal{L}$.

Let **M'** **TxtEx**-identify $\mathcal{L}'$. Let enumeration operators Θ and Ψ witness the reduction of $\mathcal{L}$ to $\mathcal{L}'$. Then we describe a machine **M** that **TxtEx**-identifies $\mathcal{L}$. **M**, upon being fed a text T for some language $L \in \mathcal{L}$, uses Θ to construct a text T' for a language in $\mathcal{L}'$. It then simulates machine **M'** on text T' and feeds conjectures of **M'** to the operator Ψ to produce its conjectures. It is easy to verify that the properties of Θ, Ψ, and **M'** guarantee the success of **M** on each text for each language in $\mathcal{L}$.

We show that under the above reduction, *SINGLE* is reducible to *COINIT* but *COINIT* is not reducible to *SINGLE*. We also show that *COINIT* is reducible to *FIN* while *FIN* is not reducible to *COINIT*, thereby justifying our intuition about the intrinsic complexity of these classes. We also show that *FIN* is in fact complete with respect

to the above reduction. Additionally, we study the status of numerous language classes with respect to this reduction and show several of them to be complete.

We also consider a stronger notion of reduction than the one discussed above. The reader should note that in the above reduction, different texts for the same language may be transformed into texts for different languages by Θ. If we further require that Θ is such that it transforms every text for a language into texts for some unique language then we have a stronger notion of reduction. In the context of function learning [10], these two notions of reduction are the same. However, surprisingly, in the context of language identification this stronger notion of reduction turns out to be different from its weaker counterpart as we are able to show that FIN is not complete with respect to the stronger reduction. We give an example of complete class with respect to the strong reduction.

We now discuss two interesting collections of languages that are shown not to be complete with respect to either reduction.

The first one is a class of languages introduced by Wiehagen [19] which contains all those languages L such that the minimum element in L is a grammar for L; we refer to this collection of languages as $WIEHAGEN$. This self-referential class, which can be **TxtEx**-identified, is a very interesting class as it contains a finite variant of every recursively enumerable language. We show that this class is not complete and is in fact equivalent to $COINIT$ under the strong reduction.

The second class is the collection of pattern languages introduced by Angluin [1]. Pattern languages have been studied extensively in the computational learning theory literature since their introduction as a nontrivial class of languages that could be learned in the limit from only positive data. We show that pattern languages are also equivalent to $COINIT$ in the strong sense, thereby implying that they pose a learning problem of similar difficulty to that of Wiehagen's class.

Finally, we also study intrinsic complexity of identification from both positive and negative data. As in the case of functions, the weak and strong reductions result in the same notion. We show that FIN is complete for identification from both positive and negative data, too.

We now proceed formally. In Section 2, we present notation and preliminaries from language learning theory. In Section 3, we introduce our reducibilities. Results are presented in Section 4.

2   Notation and Preliminaries

Any unexplained recursion theoretic notation is from [18]. The symbol N denotes the set of natural numbers, $\{0, 1, 2, 3, \ldots\}$. Unless otherwise specified, e, g, i, j, k, l, m, n, q, r, s, t, w, x, y, with or without decorations[1], range over N. Symbols $\emptyset$, $\subseteq$, $\subset$, $\supseteq$, and $\supset$ denote empty set, subset, proper subset, superset, and proper superset, respectively. Symbols A and S, with or without decorations, range over sets of numbers. S, with or without decorations, ranges over finite sets of numbers. $D_0, D_1, \ldots,$

---

[1]Decorations are subscripts, superscripts and the like.

denotes a canonical recursive indexing of all the finite sets [18, Page 70]. We assume that if $D_i \subseteq D_j$ then $i \leq j$ (the canonical indexing defined in [18] satisfies this property).

Cardinality of a set S is denoted by $\mathrm{card}(S)$. The maximum and minimum of a set are denoted by $\max(\cdot), \min(\cdot)$, respectively, where $\max(\emptyset) = 0$ and $\min(\emptyset) = \infty$.

Unless otherwise specified, letters f, F and h, with or without decorations, range over *total* functions with arguments and values from N. Symbol $\mathcal{R}$ denotes the set of all total computable functions. We let $\langle \cdot, \cdot \rangle$ stand for an arbitrary, computable, bijective mapping from $N \times N$ onto N [18]. We define $\pi_1(\langle x, y \rangle) = x$ and $\pi_2(\langle x, y \rangle) = y$. $\langle \cdot, \cdot \rangle$ can be extended to n-tuples in a natural way.

By φ we denote a fixed *acceptable* programming system for the partial computable functions: $N \to N$ [18, 15]. By φ_i we denote the partial computable function computed by the program with number i in the φ-system. The letter, p, in some contexts, with or without decorations, ranges over programs; in other contexts p ranges over total functions with its range being construed as programs. By Φ we denote an arbitrary fixed Blum complexity measure [3, 12] for the φ-system. By W_i we denote $\mathrm{domain}(\varphi_i)$. W_i is, then, the r.e. set/language ($\subseteq N$) accepted (or equivalently, generated) by the φ-program i. We also say that i is a grammar for W_i. Symbol $\mathcal{E}$ will denote the set of all r.e. languages. Symbol L, with or without decorations, ranges over $\mathcal{E}$. Symbol $\mathcal{L}$, with or without decorations, ranges over subsets of $\mathcal{E}$. We denote by $W_{i,s}$ the set $\{x \leq s \mid \Phi_i(x) < s\}$.

We now present concepts from language learning theory. The definition below introduces the concept of a *sequence* of data.

Definition 1

(a) A *sequence* σ is a mapping from an initial segment of N into $(N \cup \{\#\})$. The empty sequence is denoted by Λ.

(b) The *content* of a sequence σ, denoted $\mathrm{content}(\sigma)$, is the set of natural numbers in the range of σ.

(c) The *length* of σ, denoted by $|\sigma|$, is the number of elements in σ. So, $|\Lambda| = 0$.

(d) For $n \leq |\sigma|$, the initial sequence of σ of length n is denoted by $\sigma[n]$. So, $\sigma[0]$ is Λ.

(e) The last element of a nonempty sequence σ is denoted $\mathrm{last}(\sigma)$; the last element of Λ is defined to be 0. Formally, $\mathrm{last}(\sigma) = \sigma(|\sigma| - 1)$ if $\sigma \neq \Lambda$, otherwise $\mathrm{last}(\sigma)$ is defined to be 0.

(f) The result of stripping the last element from the sequence σ is denoted $\mathrm{prev}(\sigma)$. Formally, if $\sigma \neq \Lambda$, then $\mathrm{prev}(\sigma) = \sigma[|\sigma| - 1]$, else $\mathrm{prev}(\sigma) = \Lambda$.

Intuitively, #'s represent pauses in the presentation of data. We let σ, τ, and γ, with or without decorations, range over finite sequences. We denote the sequence formed by the concatenation of τ at the end of σ by $\sigma \diamond \tau$. Sometimes we abuse the notation and use $\sigma \diamond x$ to denote the concatenation of sequence σ and the sequence of length 1 which contains the element x. SEQ denotes the set of all finite sequences.

Definition 2 A *language learning machine* is an algorithmic device which computes a mapping from SEQ into N.

We let **M**, with or without decorations, range over learning machines.

Definition 3

(a) 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.

(b) The *content* of a text T, denoted content(T), is the set of natural numbers in the range of T.

(c) $T[n]$ denotes the finite initial sequence of T with length n.

Thus, $\mathbf{M}(T[n])$ is interpreted as the grammar (index for an accepting program) conjectured by learning machine **M** on initial sequence $T[n]$. We say that **M** converges on T to i, (written $\mathbf{M}(T){\downarrow} = i$) if $(\overset{\infty}{\forall} n)[\mathbf{M}(T[n]) = i]$.

There are several criteria for a learning machine to be successful on a language. Below we define identification in the limit introduced by Gold [11].

Definition 4 [11]

(a) **M TxtEx**-identifies a text T just in case $(\exists i \mid W_i = \text{content}(T))\ (\overset{\infty}{\forall} n)[\mathbf{M}(T[n]) = i]$.

(b) **M TxtEx**-identifies an r.e. language L (written: $L \in \mathbf{TxtEx}(\mathbf{M})$) just in case **M TxtEx**-identifies each text for L.

(c) $\mathbf{TxtEx} = \{\mathcal{L} \subseteq \mathcal{E} \mid (\exists \mathbf{M})[\mathcal{L} \subseteq \mathbf{TxtEx}(\mathbf{M})]\}$.

Other criteria of success are finite identification [11], behaviorally correct identification [8, 17, 7], and vacillatory identification [17, 5]. In the present paper, we only discuss results about **TxtEx**-identification.

3 Weak and Strong Reductions

We first present some technical machinery.

We write "$\sigma \subseteq \tau$" if σ is an initial segment of τ, and "$\sigma \subset \tau$" if σ is a proper initial segment of τ. Likewise, we write $\sigma \subset T$ if σ is an initial finite sequence of text T. Let finite sequences $\sigma^0, \sigma^1, \sigma^2, \dots$ be given such that $\sigma^0 \subseteq \sigma^1 \subseteq \sigma^2 \subseteq \cdots$ and $\lim_{i \to \infty} |\sigma^i| = \infty$. Then there is a unique text T such that for all $n \in N$, $\sigma^n = T[|\sigma^n|]$. This text is denoted $\bigcup_n \sigma^n$. Let $\mathcal{T}$ denote the set of all texts, that is, the set of all infinite sequences over $N \cup \{\#\}$.

We define an *enumeration operator*, Θ, to be an algorithmic mapping from SEQ into SEQ such that for all $\sigma, \tau \in$ SEQ, if $\sigma \subseteq \tau$, then $\Theta(\sigma) \subseteq \Theta(\tau)$. We further assume that

for all texts T, $\lim_{n \to \infty} |\Theta(T[n])| = \infty$. By extension, we think of Θ as also defining a mapping from $\mathcal{T}$ into $\mathcal{T}$ such that $\Theta(T) = \bigcup_n \Theta(T[n])$.

A final notation about the operator Θ. If for a language L, there exists an L' such that for each text T for L, $\Theta(T)$ is a text for L', then we write $\Theta(L) = L'$, else we say that $\Theta(L)$ is undefined. The reader should note the overloading of this notation because the type of the argument to Θ could be a sequence, a text, or a language; it will be clear from the context which usage is intended.

We also need the notion of an infinite sequence of grammars. We let G, with or without decorations, range over infinite sequences of grammars. From the discussion in the previous section it is clear that infinite sequences of grammars are essentially infinite sequences over N. Hence, we adopt the machinery defined for sequences and texts over to finite sequences of grammars and infinite sequences of grammars. So, if $G = g_0, g_1, g_2, g_3, \ldots$, then $G[3]$ denotes the sequence g_0, g_1, g_2, $G(3)$ is g_3, $\text{last}(G[3])$ is g_2, and $\text{prev}(G[3])$ is the sequence g_0, g_1.

We now formally introduce our reductions. Although we develop the theory of these reductions for only **TxtEx**-identification, we present the general case of the definition.

Let $\mathbf{I}$ be an identification criterion. We say that an infinite sequence of grammars G is $\mathbf{I}$-*admissible* for text T just in case G is an infinite sequence of grammars witnessing $\mathbf{I}$-identification of text T. So, if $G = g_0, g_1, g_2, \ldots$ is a **TxtEx**-admissible sequence for T, then there exists n such that for all $n' \geq n$, $g_{n'} = g_n$ and $W_{g_n} = \text{content}(T)$.

We now introduce our first reduction.

Definition 5 Let $\mathcal{L}_1 \subseteq \mathcal{E}$ and $\mathcal{L}_2 \subseteq \mathcal{E}$ be given. Let identification criteria $\mathbf{I}_1$ and $\mathbf{I}_2$ be given. Let $\mathcal{T}_1 = \{T \mid T \text{ is a text for } L \in \mathcal{L}_1\}$. Let $\mathcal{T}_2 = \{T \mid T \text{ is a text for } L \in \mathcal{L}_2\}$. We say that $\mathcal{L}_1 \leq^{\mathbf{I}_1, \mathbf{I}_2}_{\text{weak}} \mathcal{L}_2$ just in case there exist operators Θ and Ψ such that for all $T \in \mathcal{T}_1$ and for all infinite sequences of grammars $G = g_0, g_1, \ldots$ the following hold:

(a) $\Theta(T) \in \mathcal{T}_2$ and

(b) if G is an $\mathbf{I}_2$-admissible sequence for $\Theta(T)$, then $\Psi(G)$ is an $\mathbf{I}_1$-admissible sequence for T.

We say that $\mathcal{L}_1 \leq^{\mathbf{I}}_{\text{weak}} \mathcal{L}_2$ iff $\mathcal{L}_1 \leq^{\mathbf{I}, \mathbf{I}}_{\text{weak}} \mathcal{L}_2$. We say that $\mathcal{L}_1 \equiv^{\mathbf{I}}_{\text{weak}} \mathcal{L}_2$ iff $\mathcal{L}_1 \leq^{\mathbf{I}}_{\text{weak}} \mathcal{L}_2$ and $\mathcal{L}_2 \leq^{\mathbf{I}}_{\text{weak}} \mathcal{L}_1$.

As noted before, we have deliberately made the above definition general. In this paper, most of our results are about $\leq^{\mathbf{TxtEx}}_{\text{weak}}$ reduction. We now define the corresponding notions of hardness and completeness for the above reduction.

Definition 6 Let $\mathbf{I}$ be an identification criterion. Let $\mathcal{L} \subseteq \mathcal{E}$ be given.

(a) If for all $\mathcal{L}' \in \mathbf{I}$, $\mathcal{L}' \leq^{\mathbf{I}}_{\text{weak}} \mathcal{L}$, then $\mathcal{L}$ is $\leq^{\mathbf{I}}_{\text{weak}}$-*hard*.

(b) If $\mathcal{L}$ is $\leq^{\mathbf{I}}_{\text{weak}}$-hard and $\mathcal{L} \in \mathbf{I}$, then $\mathcal{L}$ is $\leq^{\mathbf{I}}_{\text{weak}}$-*complete*.

Intuitively, $\mathcal{L}_1 \leq^{\mathbf{I}}_{\text{weak}} \mathcal{L}_2$ just in case there exists an operator Θ that transforms texts for languages in $\mathcal{L}_1$ into texts for languages in $\mathcal{L}_2$ and there exists another operator Ψ that behaves as follows: if Θ transform text T to text T', then Ψ transforms $\mathbf{I}$-admissible sequences for T' into $\mathbf{I}$-admissible sequences for T. It should be noted that there is no requirement that Θ map every text for a language in $\mathcal{L}_1$ into texts for a unique language in $\mathcal{L}_2$. If we further place such a constraint on Θ, we get the following stronger notion.

Definition 7 Let $\mathcal{L}_1 \subseteq \mathcal{E}$ and $\mathcal{L}_2 \subseteq \mathcal{E}$ be given. We say that $\mathcal{L}_1 \leq^{\mathbf{I}_1, \mathbf{I}_2}_{\text{strong}} \mathcal{L}_2$ just in case there exist operators Θ, Ψ witnessing that $\mathcal{L}_1 \leq^{\mathbf{I}_1, \mathbf{I}_2}_{\text{weak}} \mathcal{L}_2$, and for all $L_1 \in \mathcal{L}_1$, there exists an $L_2 \in \mathcal{L}_2$, such that $(\forall \text{ texts } T \text{ for } L_1)[\Theta(T) \text{ is a text for } L_2]$.

We say that $\mathcal{L}_1 \leq^{\mathbf{I}}_{\text{strong}} \mathcal{L}_2$ iff $\mathcal{L}_1 \leq^{\mathbf{I}, \mathbf{I}}_{\text{strong}} \mathcal{L}_2$. We say that $\mathcal{L}_1 \equiv^{\mathbf{I}}_{\text{strong}} \mathcal{L}_2$ iff $\mathcal{L}_1 \leq^{\mathbf{I}}_{\text{strong}} \mathcal{L}_2$ and $\mathcal{L}_2 \leq^{\mathbf{I}}_{\text{strong}} \mathcal{L}_1$.

We can similarly define $\leq^{\mathbf{I}}_{\text{strong}}$-hardness and $\leq^{\mathbf{I}}_{\text{strong}}$-completeness.

It is easy to see the following.

Proposition 1 $\leq^{\mathbf{TxtEx}}_{\text{weak}}, \leq^{\mathbf{TxtEx}}_{\text{strong}}$ *are reflexive and transitive.*

The above proposition holds for most natural learning criteria. It is also easy to verify the next proposition stating that strong reducibility implies weak reducibility.

Proposition 2 *Let $\mathcal{L} \subseteq \mathcal{E}$ and $\mathcal{L}' \subseteq \mathcal{E}$ be given. Let $\mathbf{I}$ be an identification criterion. Then $\mathcal{L} \leq^{\mathbf{I}}_{\text{strong}} \mathcal{L}' \Rightarrow \mathcal{L} \leq^{\mathbf{I}}_{\text{weak}} \mathcal{L}'$.*

4   Results

In Section 4.1, we present results about reductions between the classes discussed in the introduction. Section 4.2 contains results about the status of two interesting collections of languages with respect to these reductions. Sections 4.3 and 4.4 contain results about complete classes with respect to weak and strong reductions, respectively. Finally, in Section 4.5, we consider identification from both positive and negative data.

4.1   Examples of Reductions

Recall the three language classes, *SINGLE*, *COINIT*, and *FIN*, discussed in the introduction. Our first result uses the notion of reducibility to show that in the context of **TxtEx**-identification *SINGLE* presents a strictly weaker learning problem than *COINIT*, as *SINGLE* is strong-reducible to *COINIT* whereas *COINIT* is not even weak-reducible to *SINGLE*. This is in keeping with our earlier intuitive discussion of these classes.

Theorem 1 *SINGLE $\leq^{\mathbf{TxtEx}}_{\text{strong}}$ COINIT $\wedge$ COINIT $\not\leq^{\mathbf{TxtEx}}_{\text{weak}}$ SINGLE.*

PROOF. We first construct a Θ such that $\Theta(\{n\}) = \{x \mid x \geq n\}$. Let $\tau_{m,n}$ be the lexicographically least sequence such that $\text{content}(\tau_{m,n}) = \{x \mid m \leq x \leq n\}$. Note that $\text{content}(\tau_{n+1,n}) = \emptyset$. Consider operator Θ such that if $\text{content}(\sigma) = \emptyset$, then $\Theta(\sigma) = \sigma$, else $\Theta(\sigma) = \Theta(\text{prev}(\sigma)) \diamond \tau_{\min(\text{content}(\sigma)), |\sigma|}$. For $i \in N$, let $f(i)$ denote the index of a

grammar (derived effectively from i) for the singleton language $\{i\}$. Let Ψ be defined as follows. Suppose G is a sequence of grammars, $g_0, g_1, \ldots$. Then $\Psi(G)$ denotes the sequence of grammars $g'_0, g'_1, \ldots$, where, for $n \in N$, $g'_n = f(\min(\{n\} \cup W_{g_n, n}))$.

We now show that Θ and Ψ witness $SINGLE \leq_{\text{strong}}^{\textbf{TxtEx}} COINIT$.

Let $L \in SINGLE$. We first show that Θ maps each text for L into texts for some unique language in $COINIT$. Let $L = \{e\}$. Let T be any text for L. It is easy to verify that $\Theta(T) = \cup_{n \in N}\Theta(T[n])$ is a text for the language $\{x \mid x \geq e\} \in COINIT$. Moreover, if T' is another text for L, distinct from T, then it is also easy to verify that $\text{content}(\Theta(T)) = \text{content}(\Theta(T')) = \{x \mid x \geq e\}$.

We next show that Ψ works. Suppose T is a text for $\{e\} \in SINGLE$. Let $T' = \Theta(T)$. Clearly, $\text{content}(T') = \{x \mid x \geq e\}$. Suppose $G = g_0, g_1, g_2, \ldots$ is a $\textbf{TxtEx}$-admissible sequence for T'. We claim that $\Psi(G)$ is a $\textbf{TxtEx}$-admissible sequence for T. To see the claim, let n_0 be so large that

(a) $(\forall n > n_0)[g_n = g_{n_0}]$;

(b) $n_0 > \min(W_{g_{n_0}})$; and

(c) $\min(W_{g_{n_0}}) \in W_{g_{n_0}, n_0}$.

There exists such an n_0, since G is a $\textbf{TxtEx}$-admissible sequence for T'. Let $\Psi(G) = g'_0, g'_1, g'_2, \ldots$. It is easy to verify from the definition of Ψ that, for all $n > n_0$, $g'_n = g'_{n_0}$ and g'_{n_0} is a grammar for the language $\{\min(W_{g_{n_0}})\} = \min(\text{content}(T')) = \{e\} = \text{content}(T)$. Thus Θ and Ψ witness that $SINGLE \leq_{\text{strong}}^{\textbf{TxtEx}} COINIT$.

Now suppose by way of contradiction that $COINIT \leq_{\text{weak}}^{\textbf{TxtEx}} SINGLE$ as witnessed by Θ and Ψ. Consider languages L_0 and L_1, where $L_0 = \{0, 1, 2, 3, \ldots\}$ and $L_1 = \{1, 2, 3, \ldots\}$. Clearly, both $L_0, L_1 \in COINIT$. Let σ be such that $\text{content}(\sigma) \subseteq L_1$ and $\text{content}(\Theta(\sigma)) \neq \emptyset$ (if no such σ exists then clearly Θ does not map any text for L_1 to a text for a language in $SINGLE$). Let T_0 be a text for L_0 and T_1 be a text for L_1 such that $\sigma \subset T_0$ and $\sigma \subset T_1$. Now either $\text{content}(\Theta(T_0)) = \text{content}(\Theta(T_1))$ or $\text{content}(\Theta(T_0)) \notin SINGLE$ or $\text{content}(\Theta(T_1)) \notin SINGLE$. It immediately follows that Θ and Ψ do not witness $COINIT \leq_{\text{weak}}^{\textbf{TxtEx}} SINGLE$. ∎

Our next result justifies the earlier discussion that $COINIT$ is a strictly weaker learning problem than FIN.

Theorem 2 $COINIT \leq_{\text{weak}}^{\textbf{TxtEx}} FIN \ \wedge \ FIN \not\leq_{\text{weak}}^{\textbf{TxtEx}} COINIT$.

PROOF. $COINIT \leq_{\text{weak}}^{\textbf{TxtEx}} FIN$ follows from Corollary 3 presented later. $FIN \not\leq_{\text{weak}}^{\textbf{TxtEx}} COINIT$ follows from Theorem 3 presented next. (The reader should contrast this result with theorem 11 later which implies that $COINIT \not\leq_{\text{strong}}^{\textbf{TxtEx}} FIN$.) ∎

We now present a theorem that turns out to be very useful in showing that certain classes are not complete with respect to $\leq_{\text{weak}}^{\textbf{TxtEx}}$ reduction. The theorem states that if a collection of languages $\mathcal{L}$ is such that each natural number x appears in only finitely many languages in $\mathcal{L}$, then FIN is not $\leq_{\text{weak}}^{\textbf{TxtEx}}$ reducible to $\mathcal{L}$. Since $FIN \in \textbf{TxtEx}$, this theorem immediately implies that $COINIT$ is not $\leq_{\text{weak}}^{\textbf{TxtEx}}$-complete.

Theorem 3 *Suppose $\mathcal{L}$ is such that $(\forall x)[\text{card}(\{L \in \mathcal{L} \mid x \in L\}) < \infty]$. Then FIN $\not\leq_{\text{weak}}^{\mathbf{TxtEx}}$ $\mathcal{L}$.*

PROOF. Suppose by way of contradiction that Θ and Ψ witness that $FIN \leq_{\text{weak}}^{\mathbf{TxtEx}} \mathcal{L}$. Let σ be such that content$(\Theta(\sigma)) \neq \emptyset$ (there exists such a σ, since otherwise clearly, Θ and Ψ do not witness the reduction from *FIN* to $\mathcal{L}$). Let $w = \min(\text{content}(\Theta(\sigma)))$. Let T_i be a text for content$(\sigma) \cup \{i\}$ such that $\sigma \subset T_i$. Thus for all i, we have $w \in \text{content}(\Theta(T_i))$. But since $\{\text{content}(T_i) \mid i \in N\}$ contains infinitely many languages and $\{L \in \mathcal{L} \mid w \in L\}$ is finite, there exist i, j such that content$(T_i) \neq$ content(T_j) but content$(\Theta(T_i)) =$ content$(\Theta(T_j))$. But then Θ and Ψ do not witness that $FIN \leq_{\text{weak}}^{\mathbf{TxtEx}} \mathcal{L}$. ∎

4.2   *WIEHAGEN* and Pattern Languages

Earlier results about identification in the limit from positive data turned out to be pessimistic because Gold [11] established that any collection of languages that contains an infinite language and all its finite subsets cannot be **TxtEx**-identified. As a consequence of this result no class in the Chomsky hierarchy can be identified in the limit from texts. However, later, two interesting classes were proposed that could be identified in the limit from texts. In this section, we describe these classes and locate their status with respect to the reductions introduced in this paper.

The first of these classes was introduced by Wiehagen [19]. We define, *WIEHAGEN* = $\{L \mid L \in \mathcal{E} \ \wedge \ L = W_{\min(L)}\}$.

WIEHAGEN is an interesting class because it can be shown that it contains a finite variant of every recursively enumerable language. It is easy to verify that *WIEHAGEN* $\in$ **TxtEx**. It is also easy to see that there exists a machine which **TxtEx**-identifies *WIEHAGEN* and that this machine, while processing a text for any language in *WIEHAGEN*, can provide an upper bound on the number of additional mind changes required before convergence. In this connection this class appears to pose a learning problem similar in nature to *COINIT* above. This intuition is indeed justified by the following two theorems as these two classes turn out to be equivalent in the strong sense.

Theorem 4 *WIEHAGEN* $\leq_{\text{strong}}^{\mathbf{TxtEx}}$ *COINIT*.

PROOF. Suppose Θ is such that $\Theta(L) = \{x \mid (\exists y)[y \in L \ \wedge \ x \geq y]\}$. Note that such a Θ can be easily constructed. Let Ψ be defined as follows. Suppose G is a sequence of grammars, $g_0, g_1, \ldots$. Then $\Psi(G)$ denotes the sequence of grammars $g'_0, g'_1, \ldots$, where, for $n \in N$, $g'_n = \min(\{n\} \cup W_{g_n, n})$. It is easy to see that Θ and Ψ witness *WIEHAGEN* $\leq_{\text{strong}}^{\mathbf{TxtEx}}$ *COINIT*; we omit the details. ∎

Theorem 5 *COINIT* $\leq_{\text{strong}}^{\mathbf{TxtEx}}$ *WIEHAGEN*.

PROOF. By operator recursion theorem [4] there exists a recursive 1–1 increasing function p such that for all i, $W_{p(i)} = \{x \mid x \geq p(i)\}$. Let Θ be such that $\Theta(L) = \{x \mid (\exists i)[i \in$

$L \; \wedge \; x \geq p(i)]\}$. Note that such a Θ can be easily constructed. Let Ψ be defined as follows. Let $f(i)$ denote a grammar (effectively obtained from i) such that

$$W_{f(i)} = \begin{cases} \emptyset & \text{if } i \notin \text{range}(p); \\ \{x \mid x \geq p^{-1}(i)\} & \text{otherwise.} \end{cases}$$

Suppose G is a sequence of grammars, $g_0, g_1, \ldots$. Then $\Psi(G)$ denotes the sequence of grammars $g'_0, g'_1, \ldots$, where, for $n \in N$, $g'_n = f(\min(\{n\} \cup W_{g_n,n}))$. It is easy to see that Θ and Ψ witness $COINIT \leq^{\mathbf{TxtEx}}_{\text{strong}} WIEHAGEN$; we omit the details. ∎

Corollary 1 $COINIT \equiv^{\mathbf{TxtEx}}_{\text{strong}} WIEHAGEN$.

We next consider the class, $PATTERN$, of pattern languages introduced by Angluin [1].

Suppose V is a countably infinite set of variables and C is a nonempty *finite* set of constants, such that $V \cap C = \emptyset$. *Notation*: For a set X over variables and constants, X^* denotes the set of strings over X, and X^+ denotes the set of non-empty strings over X. Any $w \in (V \cup C)^+$ is called a *pattern*. Suppose f is a mapping from $(V \cup C)^+$ to C^+, such that, for all $a \in C$, $f(a) = a$ and, for each $w_1, w_2 \in (V \cup C)^+$, $f(w_1 \cdot w_2) = f(w_1) \cdot f(w_2)$, where $\cdot$ denotes concatenation of strings. Let PatMap denote the collection of all such mappings f.

Let code denote a 1-1 onto mapping from strings in C^* to N.

The language associated with the pattern w is defined as $L(w) = \{\text{code}(f(w)) \mid f \in \text{PatMap}\}$. Then, $PATTERN = \{L(w) \mid w \text{ is a pattern}\}$.

Angluin [2] showed that $PATTERN \in \mathbf{TxtEx}$. Our first result about $PATTERN$ is that it is not $\leq^{\mathbf{TxtEx}}_{\text{weak}}$-complete.

Corollary 2 $FIN \not\leq^{\mathbf{TxtEx}}_{\text{weak}} PATTERN$.

The above Corollary follows directly from Theorem 3, since for any string x, there are only finitely many patterns w such that $x \in L(w)$.

Actually, we are also able to establish the following result.

Theorem 6 $COINIT \equiv^{\mathbf{TxtEx}}_{\text{strong}} PATTERN$.

PROOF. We first show that $COINIT \leq^{\mathbf{TxtEx}}_{\text{strong}} PATTERN$. Let $L_i = L(a^i x)$, where $a \in C$ and $x \in V$. Let Θ be such that $\Theta(L) = \{\text{code}(a^l w) \mid w \in C^+ \; \wedge \; l \in L\}$. Note that such a Θ can be easily constructed. Note that $\text{code}(a^{l+1}) \in \text{content}(\Theta(L)) \Leftrightarrow l \geq \min(L)$.

Let $f(i)$ denote an index of a grammar (obtained effectively from i) for $\{x \mid x \geq i\}$. Let Ψ be defined as follows. Suppose $G = g_0, g_1, \ldots$. Then $\Psi(G) = g'_0, g'_1, \ldots$, such that, for $n \in N$, $g'_n = f(\min(\{l \mid \text{code}(a^{l+1}) \in W_{g_n,n}\}))$. It is easy to see that Θ and Ψ witness that $COINIT \leq^{\mathbf{TxtEx}}_{\text{strong}} PATTERN$.

We now show that $PATTERN \leq^{\mathbf{TxtEx}}_{\text{strong}} COINIT$. Note that there exists a recursive indexing $L_0, L_1, \ldots$ of pattern languages such that
    (1) $L_i = L_j \Leftrightarrow i = j$.
    (2) $L_i \subset L_j \Rightarrow i > j$.

(One such indexing can be obtained as follows. First note that for patterns w_1 and w_2, if $L(w_1) \subseteq L(w_2)$ then length of w_1 is at least as large as that of w_2. Also for patterns of the same length $\subseteq$ relation is decidable [1]. Thus we can form the indexing as required using the following method. We consider only canonical patterns [1]. We place w_1 before w_2 if (a) length of w_1 is smaller than that of w_2 or (b) length of w_1 and w_2 are same, but $L(w_1) \supseteq L(w_2)$ or (c) length of w_1 and w_2 are same, $L(w_1) \not\subseteq L(w_2)$ and w_1 is lexicographically smaller than w_2.)

Moreover, there exists a machine, $\mathbf{M}$, such that

(a) For all $\sigma \subseteq \tau$, such that content$(\sigma) \neq \emptyset$, $\mathbf{M}(\sigma) \geq \mathbf{M}(\tau)$.

(b) For all texts T for pattern languages, $\mathbf{M}(T)\downarrow = i$, such that $L_i = \text{content}(T)$.

(Angluin's method of identification of pattern languages essentially achieves this property).

Let $\tau_{m,n}$ be the lexicographically least sequence of length n, such that content$(\tau_{m,n}) = \{x \mid m \leq x \leq n\}$. If content$(\sigma) = \emptyset$, then $\Theta(\sigma) = \sigma$, else $\Theta(\sigma) = \Theta(\text{prev}(\sigma)) \diamond \tau_{\mathbf{M}(\sigma),|\sigma|}$.

Let $f(i)$ denote a grammar effectively obtained from i for L_i. Let Ψ be defined as follows. Suppose $G = g_0, g_1, \ldots$. Then $\Psi(G) = g'_0, g'_1, \ldots$, such that, for $n \in N$, $g'_n = f(\min(\{n\} \cup W_{g_n,n}))$. It is easy to see that Θ and Ψ witness that $PATTERN \leq^{\mathbf{TxtEx}}_{\text{strong}}$ $COINIT$. ∎

4.3 Complete Classes for Weak Reduction

Consider the following collections of languages.

$INIT = \{L \mid (\exists n)[L = \{x \mid x < n\}]\}$.

$COSINGLE = \{L \mid \text{card}(N - L) = 1\}$.

$COFIN = \{L \mid L \text{ is cofinite}\}$.

For $n \in N$, $CONTON_n = \{L \mid \text{card}(N - L) = n\}$.

We first show that $INIT$ and FIN are equivalent in the strong sense.

Theorem 7 $INIT \equiv^{\mathbf{TxtEx}}_{\text{strong}} FIN$.

PROOF. Since $INIT \subseteq FIN$, we trivially have $INIT \leq^{\mathbf{TxtEx}}_{\text{strong}} FIN$. We show that $FIN \leq^{\mathbf{TxtEx}}_{\text{strong}} INIT$.

Note that our indexing $D_0, D_1, \ldots$ of finite sets satisfies the property that if $D_i \subseteq D_j$, then $i \leq j$. Let Θ be such that, $\Theta(D_i) = \{x \mid x \leq i\}$. Note that it is easy to construct such a Θ (since $D_i \subset D_j \Rightarrow i < j$). Let f be a function such that $W_{f(i)} = D_i$. Let Ψ be defined as follows. Suppose G is the sequence $g_0, g_1, \ldots$. Then $\Psi(G)$ is the sequence $g'_0, g'_1, \ldots$, where, for $n \in N$, $g'_n = f(\max(W_{g_n,n}))$. It is easy to see that Θ and Ψ witness that $FIN \leq^{\mathbf{TxtEx}}_{\text{strong}} INIT$. ∎

We next show that for each n, $CONTON_n$ is equivalent to $COSINGLE$ in the strong sense.

Theorem 8 For all $n \in N^+$, $COSINGLE \equiv^{\mathbf{TxtEx}}_{\text{strong}} CONTON_n$.

PROOF. Fix $n \in N^+$. First we show that $COSINGLE \leq_{\text{strong}}^{\textbf{TxtEx}} CONTON_n$. For $L \in COSINGLE$ let $L' = \{y \mid \lfloor \frac{y}{n} \rfloor \in L\}$. Let f be such that, for all i, $W_{f(i)} = \{x \mid (\exists y \in W_i)[\lfloor \frac{y}{n} \rfloor = x]\}$. Now consider Θ such that $\Theta(L) = L'$. Note that such a Θ can easily be constructed. Ψ is defined as follows. Suppose G is the sequence $g_0, g_1, \ldots$. Then $\Psi(G)$ is the sequence $f(g_0), f(g_1), \ldots$. It is easy to see that Θ and Ψ witness that $COSINGLE \leq_{\text{strong}}^{\textbf{TxtEx}} CONTON_n$.

Now we show that $CONTON_n \leq_{\text{strong}}^{\textbf{TxtEx}} COSINGLE$. For $L \in CONTON_n$, let $L' = \{\langle x_1, x_2, x_3, \ldots, x_n \rangle \mid (\exists j \mid 1 \leq j \leq n)[x_j \in L] \lor (\exists i, j \mid 1 \leq i < j \leq n)[x_i = x_j]\}$. Let f be such that, for all $\langle x_1, x_2, \ldots, x_n \rangle$, $W_{f(\langle x_1, x_2, \ldots, x_n \rangle)} = \{x \mid (\forall j \mid 1 \leq j \leq n)[x \neq x_j]\}$. Let Θ be such that $\Theta(L) = L'$. Note that such a Θ can easily be constructed. Ψ is defined as follows. Suppose G is the sequence $g_0, g_1, \ldots$. Then $\Psi(G)$ is the sequence $g_0', g_1', \ldots$, where, for $i \in N$, $g_i' = f(\min(N - W_{g_i, i}))$. It is easy to see that Θ and Ψ witness that $CONTON_n \leq_{\text{strong}}^{\textbf{TxtEx}} COSINGLE$. ∎

Since $CONTON_n \subseteq COFIN$, we trivially have $CONTON_n \leq_{\text{strong}}^{\textbf{TxtEx}} COFIN$ (note however that $COFIN \notin \textbf{TxtEx}$ [11]).

The next theorem shows that $COSINGLE$ and $CONTON_n$, for each $n \in N$, are complete with respect to weak reduction.

Theorem 9

(a) $COSINGLE$ is $\leq_{\text{weak}}^{\textbf{TxtEx}}$-complete.

(b) $COFIN$ is $\leq_{\text{weak}}^{\textbf{TxtEx}}$-hard.

(c) For all $n \in N^+$, $CONTON_n$ is $\leq_{\text{weak}}^{\textbf{TxtEx}}$-complete.

PROOF. We prove part (a). Other parts follow as corollaries. Suppose $\mathcal{L} \subseteq \textbf{TxtEx}(\textbf{M})$. We construct Θ and Ψ which witness that $\mathcal{L} \leq_{\text{weak}}^{\textbf{TxtEx}} COSINGLE$. We define Θ inductively. It is helpful to simultaneously define a function F. $F(T[0]) = \langle \textbf{M}(T[0]), 0 \rangle$. $\Theta(T[0]) = \Lambda$. Define $F(T[n+1])$ and $\Theta(T[n+1])$ as follows.

$$F(T[n+1]) = \begin{cases} F(T[n]), & \text{if } \textbf{M}(T[n+1]) = \textbf{M}(T[n]); \\ \langle \textbf{M}(T[n]), j \rangle, & \text{otherwise; where } j \text{ is such that} \\ & \langle \textbf{M}(T[n]), j \rangle > \max(\text{content}(\Theta(T[n]))). \end{cases}$$

$\Theta(T[n+1])$ is a proper extension of $\Theta(T[n])$ such that $\text{content}(\Theta(T[n+1])) = \{x \mid x \leq n \land x \neq F(T[n+1])\}$.

We now define Ψ. Intuitively, Ψ is such that if G converges to a final grammar for a language in $COSINGLE$, then $\Psi(G)$ converges to the first component of the only element not in the language enumerated by the grammar to which G converges. We now formally define Ψ. Suppose G is a sequence of grammar $g_0, g_1, \ldots$. Then $\Psi(G)$ is the sequence of grammars $g_0', g_1', \ldots$, where, for $i \in N$, $g_i' = \pi_1(\min(N - W_{g_i, i}))$.

It is easy to verify that, for $\text{content}(T) \in \textbf{TxtEx}(\textbf{M})$, if G is a $\textbf{TxtEx}$-admissible sequence for $\Theta(T)$, then $\Psi(G)$ is a $\textbf{TxtEx}$-admissible sequence for T.

Thus Θ and Ψ witness that $\mathcal{L} \leq_{\text{weak}}^{\textbf{TxtEx}} COSINGLE$. ∎

13

Our next result establishes that *COSINGLE* is reducible to *INIT* in the strong sense. This result, together with Theorem 7, yields Corollary 3 which says that both *INIT* and *FIN* are complete with respect to weak reduction. It should be noted that each of these complete classes has the property that no learning machine that identifies these classes can provide an upper bound on the number of mind changes before the onset of convergence.

Theorem 10 *COSINGLE* $\leq_{\text{strong}}^{\textbf{TxtEx}}$ *INIT*.

PROOF. For L, let $L' = \{x \mid (\forall y \leq x)[y \in L]\}$. Let Θ be such that $\Theta(L) = L'$. Note that such a Θ can be easily constructed. Let $f(i)$ denote a grammar effectively obtained from i, for $\{x \mid x \neq i\}$. Suppose G is the sequence $g_0, g_1, \ldots$. Then $\Psi(G)$ is the sequence $g'_0, g'_1, \ldots$, where for $n \in N$, $g'_n = f(\min(N - W_{g_n,n}))$. It is easy to verify that Θ and Ψ witness that *COSINGLE* $\leq_{\text{strong}}^{\textbf{TxtEx}}$ *INIT*. ∎

Corollary 3 *INIT and FIN are* $\leq_{\text{weak}}^{\textbf{TxtEx}}$-*complete.*

4.4   A Complete Class for Strong Reduction

In this section we present a collection of languages that is complete with respect strong reduction. But first we show that the classes shown to be complete with respect to weak reduction in the previous section are not complete with respect to strong reduction. Proposition 3 and Lemma 1 are useful in proving that some classes are not strongly reducible to other class.

Proposition 3 *If* $\Theta(L)$ *is defined then, for all* σ, *such that* content$(\sigma) \subseteq L$, content$(\Theta(\sigma)) \subseteq \Theta(L)$.

PROOF. Follows from the definition of $\Theta(L)$. ∎

Lemma 1 *Suppose* $L \subseteq L'$. *Then if both* $\Theta(L)$ *and* $\Theta(L')$ *are defined then* $\Theta(L) \subseteq \Theta(L')$.

PROOF. Follows from Proposition 3. ∎

Theorem 11 *COINIT* $\not\leq_{\text{strong}}^{\textbf{TxtEx}}$ *FIN*.

PROOF. Suppose by way of contradiction that *COINIT* $\leq_{\text{strong}}^{\textbf{TxtEx}}$ *FIN*, as witnessed by Θ and Ψ. Then by Lemma 1 it follows that $(\forall L \in COINIT)[\Theta(L) \subseteq \Theta(N)]$. Since *COINIT* is an infinite collection of languages, it follows that either $\Theta(N)$ is infinite or there exist distinct L_1 and L_2 in *COINIT* such that $\Theta(L_1) = \Theta(L_2)$. It follows that *COINIT* $\not\leq_{\text{strong}}^{\textbf{TxtEx}}$ *FIN*. ∎

Corollary 4 *FIN is not* $\leq_{\text{strong}}^{\textbf{TxtEx}}$-*complete.*

Theorem 12 *Suppose $L_1 \subset L_2$, then $\{L_1, L_2\} \not\leq_{\text{strong}}^{\textbf{TxtEx}} COSINGLE$.*

PROOF. Suppose by way of contradiction that $L_1 \subset L_2$ and Θ and Ψ witness that $\{L_1, L_2\} \leq_{\text{strong}}^{\textbf{TxtEx}} COSINGLE$. Then by Lemma 1 we have that $\Theta(L_1) \subseteq \Theta(L_2)$. Since for all $L'_1, L'_2 \in COSINGLE$, $L'_1 \subseteq L'_2 \Rightarrow L'_1 = L'_2$, it must be the case that $\Theta(L_1) = \Theta(L_2)$. But then, Θ and Ψ do not witness that $\{L_1, L_2\} \leq_{\text{strong}}^{\textbf{TxtEx}} COSINGLE$. ∎

As a immediate corollary we have

Corollary 5

(a) $COINIT \not\leq_{\text{strong}}^{\textbf{TxtEx}} COSINGLE$.

(b) $INIT \not\leq_{\text{strong}}^{\textbf{TxtEx}} COSINGLE$.

Theorem 13 *$SINGLE \leq_{\text{strong}}^{\textbf{TxtEx}} COSINGLE$.*

PROOF. For n, let $L_n = \{x \mid x \neq n\}$. Let Θ be such that $\Theta(\{n\}) = L_n$. It is easy to construct such a Θ. Let $f(n)$ denote a grammar effectively obtained from n, for $\{n\}$. Let Ψ be defined as follows. If G is the sequence $g_0, g_1, \ldots$, then $\Psi(G)$ is the sequence $g'_0, g'_1, \ldots$, where, for $n \in N$, $g'_n = f(\min(N - W_{g_n, n}))$. It is easy to verify that Θ and Ψ witness that $SINGLE \leq_{\text{strong}}^{\textbf{TxtEx}} COSINGLE$. ∎

Clearly, $COINIT \leq_{\text{strong}}^{\textbf{TxtEx}} COFIN$. However,

Theorem 14 *$INIT \not\leq_{\text{strong}}^{\textbf{TxtEx}} COFIN$.*

PROOF. Suppose by way of contradiction that Θ and Ψ witness that $INIT \leq_{\text{strong}}^{\textbf{TxtEx}} COFIN$. Let $L_n = \{x \mid x \leq n\}$. Now by Lemma 1, we have that for all n, $\Theta(L_n) \subseteq \Theta(L_{n+1})$. Moreover since $\Theta(L_n) \neq \Theta(L_{n+1})$ (otherwise Θ and Ψ cannot witness that $INIT \leq_{\text{strong}}^{\textbf{TxtEx}} COFIN$), we have that $\Theta(L_n) \subset \Theta(L_{n+1})$. But since $\Theta(L_0) \in COFIN$, this is not possible (only finitely many additions can be done to $\Theta(L_0)$ before it becomes N). A contradiction. ∎

We finally present a collection of languages that is complete with respect to strong reduction.

Suppose $\mathbf{M}_0, \mathbf{M}_1, \ldots$ is an enumeration of the learning machines such that, $(\forall \mathcal{L} \in \textbf{TxtEx})(\exists i)[\mathcal{L} \subseteq \textbf{TxtEx}(\mathbf{M}_i)]$ (there exists such an enumeration, see for example [16]). For $j \in N$ and $L \in \mathcal{E}$, let $S_L^j = \{\langle x, j \rangle \mid x \in L\}$. Then, let $\mathcal{L}_{\textbf{TxtEx}} = \{S_L^j \mid L \in \mathcal{E} \; \wedge \; j \in N \; \wedge \; L \in \textbf{TxtEx}(\mathbf{M}_j)\}$. It is easy to see that $\mathcal{L}_{\textbf{TxtEx}} \in \textbf{TxtEx}$.

Theorem 15 *$\mathcal{L}_{\textbf{TxtEx}}$ is $\leq_{\text{strong}}^{\textbf{TxtEx}}$ complete for $\textbf{TxtEx}$.*

PROOF. Let $\mathcal{L}_j = \{S_L^j \mid L \in \textbf{TxtEx}(\mathbf{M}_j)\}$.

If $\mathcal{L} \subseteq \textbf{TxtEx}(\mathbf{M}_j)$, then it is easy to see that $\mathcal{L} \leq_{\text{strong}}^{\textbf{TxtEx}} \mathcal{L}_j$. Since for all j, $\mathcal{L}_j \subseteq \mathcal{L}_{\textbf{TxtEx}}$, it follows that $\mathcal{L}_{\textbf{TxtEx}}$ is $\leq_{\text{strong}}^{\textbf{TxtEx}}$-complete for $\textbf{TxtEx}$. ∎

4.5 Identification from Informants

The concepts of weak and strong reduction can be adopted to language identification from informants. Informally, informants, first introduced by Gold [11], are texts which contain both positive and negative data. Thus if I_L is an informant for L, then $\text{content}(I_L) = \{\langle x, 0\rangle \mid x \notin L\} \cup \{\langle x, 1\rangle \mid x \in L\}$.[2] Identification in the limit from informants is referred to as **InfEx**-identification (we refer the reader to [11] for details). The definition of weak and strong reduction can be adopted to language identification from informants in a straightforward way by replacing texts by informants in Definitions 5 and 7.

For any language L, an informant of special interest is the canonical informant. I is a *canonical* informant for L just in case for $n \in N$, $I(n) = \langle n, x\rangle$, where $x = 1$ if $n \in L$ and $x = 0$ if $n \notin \text{L}$.

Since a canonical informant can always be produced from any informant, we have the following:

Proposition 4 $\mathcal{L}_1 \leq_{\text{weak}}^{\textbf{InfEx}} \mathcal{L}_2 \iff \mathcal{L}_1 \leq_{\text{strong}}^{\textbf{InfEx}} \mathcal{L}_2$.

Theorem 16 *FIN is $\leq_{\text{strong}}^{\textbf{InfEx}}$ complete.*

PROOF. For a language L, let I_L be the canonical informant for L. Fix a machine $\mathbf{M}$, Let $S_L^{\mathbf{M}} = \{\langle \mathbf{M}(I_L[n+1]), n\rangle \mid \mathbf{M}(I_L[n]) \neq \mathbf{M}(I_L[n+1])\}$. Let Θ be such that for all L, and informants I for L, $\Theta(I) = I_{S_L^{\mathbf{M}}}$. Note that such a Θ can easily be constructed. Suppose F is such that, for a finite set S, $F(S) = \min(\{i \mid (\exists j)[\langle i, j\rangle \in S \wedge j = \max(\{k \mid (\exists x)[\langle x, k\rangle \in S]\})]\})$. Let Ψ be defined as follows. Suppose G is a sequence $g_0, g_1, \ldots$. Then $\Psi(G)$ is the sequence $g_0', g_1', \ldots$, where for $n \in N$, $g_n' = F(W_{g_n,n})$. It is easy to verify that Θ and Ψ witness that $\textbf{InfEx}(\mathbf{M}) \leq_{\text{strong}}^{\textbf{InfEx}} FIN$. ∎

However,

Theorem 17 *The classes SINGLE, INIT, COSINGLE, CONTON$_n$, COINIT, WIEHAGEN, and PATTERN are equivalent with respect to $\leq_{\text{strong}}^{\textbf{InfEx}}$ reduction.*[3]

PROOF. It is easy to see that $SINGLE \leq_{\text{strong}}^{\textbf{InfEx}} \mathcal{L}$, where $\mathcal{L}$ is one of $COSINGLE$, $CONTON_n$, $COINIT$, $WIEHAGEN$, $PATTERN$. We show that $COSINGLE \leq_{\text{strong}}^{\textbf{InfEx}} SINGLE$ and that $WIEHAGEN \leq_{\text{strong}}^{\textbf{InfEx}} SINGLE$. Other reduction can be done in a similar manner.

We first show $COSINGLE \leq_{\text{strong}}^{\textbf{InfEx}} SINGLE$. Consider Θ such that, for any I for $L \in COSINGLE$, $\Theta(I) = I'$, such that I' is an informant for $\{\min(\overline{L})\}$. Note that such a Θ can be easily constructed. Let Ψ, be defined as follows. Let $f(i)$ be a grammar, effectively obtained from i, for $\{x \mid x \neq i\}$. For $G = g_0, g_1, \ldots,$ $\Psi(G) = g_0', g_1', g_2', \ldots$, where $g_i' = f(\min(\{n\} \cup W_{g_i,i}))$. It is easy to see that Θ, Ψ witness that $COSINGLE \leq_{\text{strong}}^{\textbf{InfEx}} SINGLE$.

We now show $WIEHAGEN \leq_{\text{strong}}^{\textbf{InfEx}} SINGLE$. Consider Θ such that, for any I for $L \in WIEHAGEN$, $\Theta(I) = I'$, such that I' is an informant for $\{\min(L)\}$. Note that

---

[2]Alternatively, an informant for a language L may be thought of as a "tagged" text for N such that n appears in the text with tag 1 if $n \in L$; otherwise n appears in the text with tag 0.

[3]Actually, it can be shown that any collection of languages that can be finitely identified (i.e., identified with 0 mind changes) from informants is $\leq_{\text{strong}}^{\textbf{InfEx}} SINGLE$.

such a Θ can be easily constructed. Let Ψ, be defined as follows. For $G = g_0, g_1, \ldots,$ $\Psi(G) = g'_0, g'_1, g'_2, \ldots,$ where $g'_i = \min(\{n\} \cup W_{g_i,i})$. It is easy to see that Θ, Ψ witness that $WIEHAGEN \leq_{\mathrm{strong}}^{\mathbf{InfEx}} SINGLE$. ∎

5   Conclusion

A novel approach to studying the intrinsic complexity of language identification was undertaken using weak and strong reductions between classes of languages. The intrinsic complexity of several classes was considered. It was shown that the self referential class of Wiehagen [19] in which the least element of every language is a grammar for the language and the class of pattern languages introduced by Angluin [1] are equivalent in the strong sense. A number of complete classes were presented for both the reductions. It was also shown that the weak and strong reductions are distinct for learning from text.

The results presented were for the widely studied identification in the limit criterion. These techniques have also been applied to other criteria of success. Additionally, the structure of these reductions has also been studied [14]. However, it is felt that for these reductions to have an impact on the study of feasibility issues in language identification, their fidelity has to be improved.

Acknowledgements

References

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

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

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

[4] J. Case. Periodicity in generations of automata. *Mathematical Systems Theory*, 8:15–32, 1974.

[5] 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 [6].

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

[7] 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*, pages 107–115. Springer-Verlag, 1982. Lecture Notes in Computer Science 140.

[8] J. Feldman. Some decidability results on grammatical inference and complexity. *Information and Control*, 20:244–262, 1972.

[9] R. Freivalds. Inductive inference of recursive functions: Qualitative theory. In J. Barzdins and D. Bjorner, editors, *Baltic Computer Science. Lecture Notes in Computer Science 502*, pages 77–110. Springer-Verlag, 1991.

[10] R Freivalds, E. Kinber, and C. H. Smith. On the intrinsic complexity of learning. Technical Report 94-24, University of Delaware, Newark, Delaware, 1994.

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

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

[13] S. Jain and A. Sharma. On the intrinsic complexity of language identification. In *Proceedings of the Seventh Annual Conference on Computational Learning Theory, New Brunswick, New Jersey*, pages 278–286. ACM-Press, July 1994.

[14] S. Jain and A. Sharma. The structure of intrinsic complexity of learning. In *Proceedings of the Second European Conference on Computational Learning Theory*, March 1995. To Appear.

[15] M. Machtey and P. Young. *An Introduction to the General Theory of Algorithms*. North Holland, New York, 1978.

[16] 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.

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

[18] H. Rogers. *Theory of Recursive Functions and Effective Computability*. McGraw Hill, New York, 1967. Reprinted, MIT Press 1987.

[19] R. Wiehagen. Identification of formal languages. In *Mathematical Foundations of Computer Science, Proceedings, 6th Symposium, Tatranska Lomnica*, pages 571–579. Springer-Verlag, 1977. Lecture Notes in Computer Science 53.