Inductive Inference of Languages from Samplings

Sanjay Jain*[1] and Efim Kinber[2]

[1] School of Computing, National University of Singapore, Singapore 117417, Republic of Singapore. Email: sanjay@comp.nus.edu.sg

[2] Department of Computer Science, Sacred Heart University, Fairfield, CT 06825-1000, U.S.A. Email: kinbere@sacredheart.edu

Abstract. We introduce, discuss, and study a model for inductive inference from samplings, formalizing an idea of learning different "projections" of languages. One set of our results addresses the problem of finding a uniform learner for all samplings of a language from a certain set when learners for particular samplings are available. Another set of results deals with extending learnability from a large natural set of samplings to larger sets. A number of open problems is formulated.

Keywords: Inductive inference, samplings, sublanguages.

1 Introduction

Consider the following model of learning. A learner gets data about a target concept, one piece at a time. As the learner is receiving the data, it outputs its conjectures on what the target might be. If the sequence of conjectures of the learner converges to a correct hypothesis about the target concept, then one might say that the learner has successfully learnt the concept. This is essentially the model of **TxtEx**-learning considered by Gold [Gol67].

In our paper, the target concept is a language L from a class $\mathcal{L}$ of possible languages. The learner gets as input, one element at a time, in arbitrary order with repetitions allowed, members of the target language (such a presentation is called a *text* of the language; note that negative data is not presented to the learner in this model). The conjectures made by the learner take the form of a grammar or acceptor in some acceptable programming system [Rog67]. The learner is then successful (that is, the learner **TxtEx**-identifies the target language) if the sequence of conjectures converges to a grammar which generates/accepts the target language L.

Expecting that the learner gets all elements of the target language is unrealistic. Often it is difficult to obtain full data, and a learner gets actually elements of only a subset X of the target language L. Of course, then it may be unrealistic to expect the learner to learn the full language. Thus, in such a situation one says that the learner is successful if it converges to a grammar for a language L' such that $X \subseteq L' \subseteq L$. In [JK08], the authors considered this model of learning.

* Supported in part by NUS grant number R252-000-308-112 and C-252-000-087-001.

Note that here, the learner is supposed to be successful in the above sense for all subsets X of the target language L.

In this paper, we consider a different variant of the model explored in [JK08]: every language from a target class is being learnt from inputs defined by the same *sampling*. As in the model in [JK08], the learner must produce in the limit a grammar covering the input sampling of the target language, and the final grammar must represent a subset of the target language containing all data from the input sampling. The difference of our variant from the model in [JK08] is that, in [JK08], any *arbitrary* subset of the target language may appear on the input, whereas in our model the input is defined by a certain (fixed) sampling.

Since positive data in the process of inductive inference can be viewed as being supplied by a teacher, it is natural to assume that, in some situations, the teacher may have difficulty providing full positive data — because, for example, it is too time-consuming, or it is too lengthy. In such a situation, the teacher may provide just a part of the target language using some natural representation of the language known to the teacher. For example, the teacher may provide, say, the 2nd, 4th, 6th, etc. elements of this representation, or some other sequence of positive data that is representative of the target language as a whole and/or represents one or another salient aspect of it. Say, if the target language consists of all primes, the teacher may provide the sampling containing just every other prime (or another sampling of a similar sort) to give the learner a good idea of what the language is about. Or, if all languages in the target class are infinite, the teacher may omit some elements of the target language as being of lesser importance for the overall understanding of the concept. (Our idea of sampling can be traced to Trakhtenbrot's paper [Tra73]).

The learner is required to be able to produce in the limit a grammar that covers the input part (representing a given sampling) of the target language and does not exceed the target language (in set containment sense). Now, obviously, there are many different issues of interest: for example, does learnability of every specific sampling (from some class of samplings) imply overall (uniform) learnability for all samplings, how does knowledge of the sampling choice (provided by the teacher) affect learning capabilities, etc.

The idea of considering inductive inference from inputs defined by different samplings was first suggested by R. Freivalds in [Fre74], where he studied how learnability on different specific orderings of the input function graph (provided by the teacher) can affect overall learnability of the function regardless of the order of input. When one assumes that languages are sets of positive integers (as, following the tradition of classic inductive inference, it is done in our paper), and considers learning languages from positive data only, the teacher may use samplings based on many different natural representations of all positive data. For example, one can fix some standard way of enumerating all recursively enumerable sets, and then, for a given enumeration $a_0, a_1, a_2, a_3, \ldots$ of a (recursively enumerable) language L, the sampling, say, $A = \{1, 3, 5, \ldots\}$ will define the sublanguage $a_1, a_3, a_5, \ldots$. However, it turns out that this approach makes learnability dependent on the choice of enumerating mechanism, and, thus, rather unnatural.

We have chosen a formalization of the concept of sampling based on the representation of target languages in the increasing order. Namely, for any language L, consider the increasing order of all elements $a_0, a_1, a_2, a_3, \ldots$. Then a *sampling* A is a set of positive integers, say, $i_1, i_2, \ldots$ and the corresponding A-*sampling* (sublanguage) of L is the language $\{a_{i_1}, a_{i_2}, a_{i_3}, \ldots\}$. For example, for the sampling $A = \{5, 6, 7, \ldots\}$, one gets the A-sampling of L containing all the elements of L except the first five smallest ones.

Given the aforementioned formalization, we study several natural questions. The first question is if, given a class of languages $\mathcal{L}$ and a set of samplings $\mathcal{A}$, learnability of A-samplings of languages from $\mathcal{L}$ for every specific $A \in \mathcal{A}$ implies uniform learnability (that is, by one learner) of all A-samplings of languages from $\mathcal{L}$ for all $A \in \mathcal{A}$. We answer this question in the negative in Theorem 7, using the set of all possible samplings. For some special set of samplings, we also show that a class witnessing separation of non-uniform and uniform learnability for this set of samplings can be uniformly learned with just one error in the final conjecture (Theorem 8). A related result addresses the question whether uniform learnability on each of the two different sets of samplings implies uniform learnability on all samplings from the union of these two sets: as we show in Theorem 10, the answer is negative even if each set contains just one recursive sampling. On the other hand, we suggest a simple sufficient condition for learnability on the union of two sets of samplings when the learner gets access to the sampling A (from the oracle, or as a separate input), see Proposition 12.

We also studied the following problem: what are the circumstances when learnability of a class from some natural set of samplings ensures learnability of the class from a larger natural set of samplings? For example, is learnability of a class on the set of A-samplings for all infinite recursively enumerable A-s powerful enough to ensure learnability of the class from all infinite samplings? We were able to get only some negative results so far. In particular, we have shown that learnability of a class from all A-samplings for all infinite recursively enumerable A-s does not imply the learnability of the class from all infinite samplings (Theorem 16). Similarly, it turns out that learnability of a class from all simple recursively enumerable samplings does not imply learnability of the class from all infinite recursively enumerable samplings (Theorem 17). Moreover, it turns out that the learnability of a class from all samplings but some recursive sampling A (and all its subsets), does not imply that the class is learnable from all samplings (Theorem 18).

2 Preliminaries

2.1 Notations

Any unexplained recursion theoretic notation is from [Rog67]. Let N denote the set of natural numbers, $\{0, 1, 2, 3, \ldots\}$. Symbols $\emptyset$, $\subseteq$, $\subset$, $\supseteq$, and $\supset$ denote the empty set, subset, proper subset, superset, and proper superset, respectively. Symmetric difference of A and B is denoted by $A\mathbf{\Delta}B$. That is, $A\mathbf{\Delta}B = (A - B) \cup (B - A)$. $\mathcal{P}(A)$ denotes the power set of A, that is, $\mathcal{P}(A) = \{B \mid B \subseteq A\}$. The maximum and minimum of a set are respectively denoted by $\max(\cdot), \min(\cdot)$, where we take $\max(\emptyset) = 0$ and $\min(\emptyset) = \infty$. For a set $S = \{x_0, x_1, \ldots\}$, where $x_0 < x_1 < \ldots$, we call x_i the i-th minimal element (or just the i-th element) of S (thus, the 0-th (minimal) element is the minimal element of a set). The cardinality of a set S is denoted by $\mathrm{card}(S)$. We use $\mathrm{card}(S) \leq *$ to denote that the cardinality of S is finite. For $a \in N \cup \{*\}$, $A =^a B$ denotes that $\mathrm{card}(A\mathbf{\Delta}B) \leq a$. Quantifiers $\forall^\infty$ and $\exists^\infty$ respectively denote 'for all but finitely many' and 'there exist infinitely many'.

We let $\langle \cdot, \cdot \rangle$ denote a computable 1–1 and onto mapping from $N \times N$ to N (see [Rog67]). We assume without loss of generality that $\langle \cdot, \cdot \rangle$ is increasing in both its arguments. Let $\pi_1^2(\langle x, y \rangle) = x$ and $\pi_2^2(\langle x, y \rangle) = y$. The pairing function can be extended to coding of n-tuples in a natural way by taking $\langle x_1, x_2, \ldots, x_n \rangle = \langle x_1, \langle x_2, x_3, \ldots, x_n \rangle \rangle$, for $n > 2$. The corresponding projection functions are $\pi_i^n(\langle x_1, x_2, \ldots, x_n \rangle) = x_i$.

For a partial function η, $\eta(x)\downarrow$ denotes that $\eta(x)$ is defined. $\eta(x)\uparrow$ denotes that $\eta(x)$ is undefined. We let $\eta[n]$ denote the partial function, $\{(x, \eta(x)) \mid x < n\}$. By φ we denote a fixed *acceptable* programming system for the partial computable functions from N to N [Rog67,HU79]. Then, φ_i denotes the i-th partial computable function in this programming system, and i is called a program for the partial function φ_i. By Φ we denote a fixed Blum complexity measure [Blu67,HU79] for the φ-system. Intuitively, $\Phi_i(x)$ denotes the resources (say time or space) needed to compute $\varphi_i(x)$.

Languages are subsets of N. By W_i we denote $\mathrm{domain}(\varphi_i)$. Thus, W_i is the recursively enumerable (r.e.) set/language accepted by φ_i. We also say that i is a grammar for W_i. Symbol $\mathcal{E}$ denotes the set of all r.e. languages. By $W_{i,s}$ we denote the set $\{x < s \mid \Phi_i(x) < s\}$. L, with or without decorations, ranges over $\mathcal{E}$. We let χ_L denote the characteristic function of L. We let $\overline{L} = N - L$, that is the complement of L. Symbol $\mathcal{L}$, with or without decorations, ranges over subsets of $\mathcal{E}$.

A set S is called *immune* [Rog67] iff S is infinite, and for all infinite r.e. sets X, $X \not\subseteq S$. A set S is called *simple* [Rog67] iff S is recursively enumerable and $\overline{S}$ is immune.

For a total function f, let $L_f = \{\langle x, f(x) \rangle \mid x \in N\}$. For any, possibly partial, function g, let Zext_g be the function defined as follows: $\mathrm{Zext}_g(x) = g(x)$, if $x \in \mathrm{domain}(g)$; $\mathrm{Zext}_g(x) = 0$, otherwise.

We will consider the following classes of languages and functions:

- INF is the class of all infinite sets.
- REinf is the class of all infinite recursively enumerable sets.
- $\mathrm{INIT} = \{L \mid (\exists n)[L = \{x \mid x < n\}]\}$, the class of initial segments of N.
- $\mathrm{COINIT} = \{L \mid (\exists n)[L = \{x \mid x \geq n\}]\}$, the class of coinitial segments of N.
- $\mathrm{SD} = \{f \in \mathcal{R} \mid \varphi_{f(0)} = f\}$.
- $\mathrm{AZext} = \{f \in \mathcal{R} \mid \mathrm{domain}(\varphi_{f(0)}) \in \mathrm{INIT}$, and $f = \mathrm{Zext}_{\varphi_{f(0)}}\}$.

2.2 Preliminaries for Learning

A *text* T is a mapping from N into $(N \cup \{\#\})$. Thus, $T(i)$ represents the $(i+1)$-st element in the text. Intuitively, a text denotes the presentation of elements of a language, with #s representing pauses in the presentation. We let T, with or without decorations, range over texts. Content of a text T, denoted content(T), is the set of natural numbers in the range of T. A text T is for a language L iff content$(T) = L$. $T[n]$ denotes the initial sequence of T of length n, that is $T[n] = T(0)T(1)\ldots T(n-1)$.

A finite sequence is a mapping from an initial segment of N into $(N \cup \{\#\})$. The empty sequence is denoted by Λ. Content of σ, denoted content(σ), is the set of natural numbers in the range of σ. The *length* of σ, denoted by $|\sigma|$, is the number of elements in σ. For $n < |\sigma|$, $\sigma(i)$ denotes the $(i+1)$-th element in σ. For $n \leq |\sigma|$, $\sigma[n]$ denotes the initial sequence of σ of length n. SEQ denotes the set of all finite sequences. Thus, SEQ $= \{T[n] \mid n \in N, T \text{ is a text}\}$. We let σ and τ, with or without decorations, range over SEQ. We denote the sequence formed by the concatenation of τ at the end of σ by $\sigma\tau$.

An *inductive inference machine* (IIM) [Gol67] is an algorithmic mapping from SEQ to N. We also use the term *learner* or *learning machine* for IIM. We let $\mathbf{M}$, with or without decorations, range over IIMs. We say that $\mathbf{M}$ converges on T to i, (written: $\mathbf{M}(T)\!\downarrow = i$) iff $(\forall^\infty n)[\mathbf{M}(T[n]) = i]$.

The following define some of the notions of learning.

Definition 1. [Gol67,CL82]

(a) $\mathbf{M}$ $\mathbf{TxtEx}$-*identifies an r.e. language* L (written: $L \in \mathbf{TxtEx}(\mathbf{M})$) just in case, for all texts T for L, $\mathbf{M}(T[n])$ is defined for all n and $(\exists i \mid W_i = L)(\forall^\infty n)[\mathbf{M}(T[n]) = i]$.

(b) $\mathbf{M}$ $\mathbf{TxtEx}$-*identifies a class* $\mathcal{L}$ *of r.e. languages* (written: $\mathcal{L} \subseteq \mathbf{TxtEx}(\mathbf{M})$) just in case $\mathbf{M}$ $\mathbf{TxtEx}$-identifies each language from $\mathcal{L}$.

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

Definition 2. [OW82,Gol67,CL82]

(a) $\mathbf{M}$ $\mathbf{TxtBc}$-*identifies an r.e. language* L (written: $L \in \mathbf{TxtBc}(\mathbf{M})$) just in case, for all texts T for L, for all but finitely many n, $W_{\mathbf{M}(T[n])} = L$.

(b) $\mathbf{M}$ $\mathbf{TxtBc}$-*identifies a class* $\mathcal{L}$ *of r.e. languages* (written: $\mathcal{L} \subseteq \mathbf{TxtBc}(\mathbf{M})$) just in case $\mathbf{M}$ $\mathbf{TxtBc}$-identifies each language from $\mathcal{L}$.

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

There exists a recursive sequence of total IIMs, $\mathbf{M}_0, \mathbf{M}_1, \ldots$ such that, for the criteria of learning $\mathbf{I}$ discussed in this paper, for each $\mathcal{L} \in \mathbf{I}$, some $\mathbf{M}_i$ witnesses that $\mathcal{L} \in \mathbf{I}$. This can be shown essentially along the same lines as done for $\mathbf{TxtEx}$-learning in [OSW86]. Thus, any learner $\mathbf{M}$ can be considered to be equivalent to some $\mathbf{M}_j$ from such an enumeration (with respect to being able to learn a class of languages under a criterion of inference considered in this paper). We fix one such enumeration $\mathbf{M}_0, \mathbf{M}_1, \ldots$, and will from now on consider only learners from this list.

Definition 3. (a) [Ful90] σ is said to be a **TxtEx**-*stabilizing sequence* for $\mathbf{M}$ on L, iff (i) content$(\sigma) \subseteq L$, and (ii) for all τ such that content$(\tau) \subseteq L$, $\mathbf{M}(\sigma\tau) = \mathbf{M}(\sigma)$.

(b) [BB75,Ful90] σ is said to be a **TxtEx**-*locking sequence* for $\mathbf{M}$ on L, iff (i) σ is a **TxtEx**-stabilizing sequence for $\mathbf{M}$ on L and (ii) $W_{\mathbf{M}(\sigma)} = L$.

If $\mathbf{M}$ **TxtEx**-identifies L, then every **TxtEx**-stabilizing sequence for $\mathbf{M}$ on L is a **TxtEx**-locking sequence for $\mathbf{M}$ on L. Furthermore, one can show that if $\mathbf{M}$ **TxtEx**-identifies L, then for every σ such that content$(\sigma) \subseteq L$, there exists a **TxtEx**-locking sequence, which extends σ, for $\mathbf{M}$ on L (see [BB75,Ful90]).

Similar locking sequence results can be proved for other criteria of inference considered in this paper.

3 Definitions for Learning from Samplings

Suppose $S = \{y_0, y_1, y_2, \ldots\}$, where $y_0 < y_1 < \ldots$, and $R \subseteq S$. Then, define Order$(R, S) = \{i \mid y_i \in R\}$. We call any subset of N a *sampling*. For a sampling A and a set S, we call X an A-sampling of S iff Order$(X, S) = A$.

Note that, if A is infinite and S is finite, then there is no A-sampling of S. On the other hand, for every infinite set S, and every sampling A, there is a (unique) A-sampling of S. Thus, for ease of notation, when learning from samplings, we will only consider infinite languages, without always explicitly mentioning so. This does not effect our results, as all the diagonalizations in this paper can be achieved using classes of infinite languages. Note that the samplings themselves may or may not be infinite.

The following definition now formalizes our notion of learning from samplings. Note that the model for learning sublanguages of a target language in [JK08] is different from the one formalized in the definition below, as the model in [JK08] requires learners to learn arbitrary input sublanguages, rather than the ones defined by specific samplings as considered in this paper. In other words, the model in [JK08] is $\mathbf{UniSublang}_{\mathcal{P}(N)}$, a special case of the models considered in this paper.

Definition 4. (a) Suppose a sampling $A \subseteq N$ is given. $\mathbf{M}$ $\mathbf{SublangEx}_A$-identifies an infinite language L iff, for A-sampling X of L, for any text T for X, there exists an n such that, (i) for all $m \geq n$, $\mathbf{M}(T[m]) = \mathbf{M}(T[n])$, and (ii) $X \subseteq W_{\mathbf{M}(T[n])} \subseteq L$.

$\mathbf{M}$ $\mathbf{SublangEx}_A$-identifies a class $\mathcal{L}$ of infinite languages iff $\mathbf{M}$ $\mathbf{SublangEx}_A$-identifies each language in $\mathcal{L}$.

$\mathbf{SublangEx}_A$ denotes the collection of all $\mathcal{L}$ which can be $\mathbf{SublangEx}_A$-identified by some learner $\mathbf{M}$.

(b) Let $\mathcal{A} \subseteq \mathcal{P}(N)$ be a set of samplings.

(b.1) $\mathbf{SublangEx}_{\mathcal{A}}$ denotes the collection of all $\mathcal{L}$ such that, for each $A \in \mathcal{A}$, $\mathcal{L} \in \mathbf{SublangEx}_A$ (this is the non-uniform version).

(b.2) **UniSublangEx**$_{\mathcal{A}}$ denotes the collection of all $\mathcal{L}$ for which there exists a learner **M** such that, for each $A \in \mathcal{A}$, **M SublangEx**$_A$-identifies $\mathcal{L}$ (this is the uniform version).

(b.3) **PUniSublangEx**$_{\mathcal{A}}$ denotes the collection of all $\mathcal{L}$ for which there exists a learner **M** such that, for each $A \in \mathcal{A}$, **M** using an oracle for A, **SublangEx**$_A$-identifies $\mathcal{L}$ (this is the pseudo-uniform version, where the learner has access to the sampling A in oracle form).

One can similarly define the criteria for **SublangBc**-learning. We say that **M SublangBc**$_A$-identifies $\mathcal{L}$ iff for all $L \in \mathcal{L}$, for $X \subseteq L$ such that $\mathrm{Order}(X, L) = A$, for any text T for X, for all but finitely many n, $X \subseteq W_{\mathbf{M}(T[n])} \subseteq L$.

In this paper, we will be mainly concentrating on **Sublang** and **UniSublang**-learning paradigms. The criterion **PUniSublangEx** is used more for emphasizing what happens if a uniform learner "knows", in some way, the sampling A which it is getting. The usage of an oracle here is more for convenience, and the results presented in the paper will hold even if one gives the set A to the learner in the form of a separate text containing exactly the elements of A.

4 Results

Our first goal is to show that there are classes of languages non-uniformly learnable on all samplings, but not uniformly learnable, even just on samplings from COINIT. We begin with the following useful proposition and a corollary from it.

Proposition 5. $\mathcal{L} = \{L_f \mid f \in SD \cup AZext\} \notin$ **TxtBc**, *even when the texts given to the learner are increasing texts.*

Proof. This proposition can be proved essentially along the same lines as the proof of the non-union theorem [BB75]. For ease of presentation, we will give the proof for learning from arbitrary texts. As the class $\mathcal{L}$ used consists only of L_f such that $f \in \mathcal{R}$, such texts can be effectively converted to increasing texts. Suppose, by way of contradiction, that **M TxtBc**-identifies $\mathcal{L}$. Then, by implicit use of Kleene's recursion theorem [Rog67], there exists an e such that φ_e may be defined as follows. Let $\varphi_e(0) = e$. φ_e^s denotes φ_e defined before stage s, and x_s denotes the largest x such that $\varphi_e(x)$ is defined before stage s. Let σ_0 contain just one element: $\langle 0, e\rangle$. It will be the case that $\mathrm{content}(\sigma_s) = \{\langle x, \varphi_e(x)\rangle \mid x \leq x_s\}$. Go to stage 0.

Stage s:
1. Search for a $\sigma \supseteq \sigma_s$ and $y \in N$ such that: $\mathrm{content}(\sigma) \subseteq L_{\mathrm{Zext}_{\varphi_e^s}}$, $y > x_s$, $\langle y, 0\rangle \notin \mathrm{content}(\sigma)$ and $\langle y, 0\rangle \in W_{\mathbf{M}(\sigma)}$.
2. If and when such a σ and y are found,
   let z be maximum such that $\langle z, 0\rangle \in \mathrm{content}(\sigma)$,
   let $\varphi_e(y) = 1$,
   let $\varphi_e(x) = 0$, for $x \neq y$ such that $x_s < x \leq y + z + 1$.
   let $x_{s+1} = y + z + 1$.

let σ_{s+1} be an extension of σ such that $\text{content}(\sigma_{s+1}) = \{\langle x, \varphi_e(x)\rangle \mid x \le x_{s+1}\}$.

Go to stage $s+1$

End stage s

It is easy to verify that, if all stages terminate, then φ_e is total, $\varphi_e \in$ SD, $T = \bigcup_{s\in N} \sigma_s$ is a text for L_{φ_e}, and $W_{\mathbf{M}(T[n])} \neq L_{\varphi_e}$, for infinitely many n (as $W_{\mathbf{M}(\sigma)} \neq L_{\varphi_e}$, for each of σ found in the stages).

On the other hand, if stage s starts but does not finish, then for $g = \text{Zext}_{\varphi_e} \in$ AZext, we have that for any text T for L_g which extends σ_s, $\mathbf{M}(T[n])$ is not a grammar for L_g for any $n > |\sigma_s|$ (otherwise, the search in step 1 would succeed).

Thus, we have that $\mathbf{M}$ cannot $\mathbf{TxtBc}$-identify $\mathcal{L}$. ∎

Corollary 6. *Let* $SD_j = \{f \mid (\exists g \in SD)[(\forall x < j)[f(x) = \langle j, g(0), 1\rangle]$ *and* $(\forall x \ge j)[f(x) = \langle j, g(0), g(x-j)+2\rangle]]\}$.

Let $AZext_j = \{f \mid (\exists g \in AZext)[(\forall x < j)[f(x) = \langle j, g(0), 0\rangle]$ *and* $(\forall x \ge j)[f(x) = \langle j, g(0), g(x-j)+2\rangle]]\}$.

Let $\mathcal{L}_j = \{L_f \mid f \in SD_j \cup AZext_j\}$.

Then $\mathcal{L}_j \notin \mathbf{UniSublangBc}_A$*, where* $A = \{j, j+1, j+2, \ldots\}$*. Thus,* $\mathcal{L}_j \notin \mathbf{UniSublangBc}_{COINIT}$.

Note: We used $+2$ above just to make sure that $\langle j, e, f(x)\rangle$ for $x < j$ are smaller than $\langle j, e, f(x)\rangle$, for $x \ge j$.

Now we can prove the separation result for the uniform and non-uniform learnability from samplings. The following Theorem holds even if we replace **Uni** by **PUni**.

Theorem 7. $\mathbf{SublangEx}_{\mathcal{P}(N)} - \mathbf{UniSublangBc}_{COINIT} \neq \emptyset$.

Proof. Let $\mathcal{L}_j$ be as in the Corollary 6. Fix $L_j \in \mathcal{L}_j$ that witnesses that $\mathbf{M}_j$ does not $\mathbf{UniSublangBc}_{\text{COINIT}}$-identify $\mathcal{L}_j$. Let $\mathcal{L} = \{L_j \mid j \in N\}$. Then, clearly $\mathcal{L} \notin \mathbf{UniSublangBc}_{\text{COINIT}}$.

To see that $\mathcal{L} \in \mathbf{SublangEx}_{\mathcal{P}(N)}$, let $A \in \mathcal{P}(N)$ be given. If $A = \emptyset$, then $\mathcal{L}$ is trivially in $\mathbf{SublangEx}_A$, as the learner can just output $\emptyset$.

If $A \neq \emptyset$, then let $k \in A$. Now, the learner can $\mathbf{SublangEx}_A$-learn $\mathcal{L}$ as follows. If the input text contains an element $\langle x, \langle j, e, y\rangle\rangle$, for some $j \le k$, then output a grammar for L_j (as there are only finitely many such j, the learner can "code" these finitely many cases).

Otherwise, the input text will contain $\langle k, \langle j, e, y\rangle\rangle$ for some $j > k$, for some $y \in \{0, 1\}$ (as the k-th least element in L_j is $\langle k, \langle j, e, y\rangle\rangle$, for some $y \in \{0, 1\}$). If $y = 1$, then output a grammar for L_f, where $f(x) = \langle j, e, 1\rangle$, if $x < j$, and $f(x) = \langle j, e, \varphi_e(x-j)+2\rangle$, if $x \ge j$. If $y = 0$, then, in the limit, search for the least x_0 such that $\varphi_e(x_0)$ is not defined. Then, output a grammar for L_f, where $f(x) = \langle j, e, 0\rangle$, if $x < j$, and $f(x) = \langle j, e, g(x-j)+2\rangle$, if $x \ge j$, where $g = \text{Zext}_{\varphi_e[x_0]}$. It is easy to verify that the above learner will $\mathbf{SublangEx}_A$-identify $\mathcal{L}$. ∎

Now we exhibit a different result on separation of non-uniform and uniform learnability from samplings: for a certain set $\mathcal{A}$ of samplings, there exists a class of languages, which is non-uniformly learnable on all samplings from the given set $\mathcal{A}$, uniformly learnable with just one error in the final conjecture, but not uniformly learnable without errors in conjectures even in **Bc** style.

Here a learner $\mathbf{M}$ $\mathbf{SublangEx}^a_A$-identifies $\mathcal{L}$ iff, for all $L \in \mathcal{L}$, if $X \subseteq L$ is an A-sampling of L, then for any text T for X, there exists an n and a set Z such that, (i) for all $m \geq n$, $\mathbf{M}(T[m]) = \mathbf{M}(T[n])$, (ii) $X \subseteq Z \subseteq L$, and (iii) $Z =^a W_{\mathbf{M}(T[n])}$.

Theorem 8. *There exists a class $\mathcal{L}$ such that for $\mathcal{A} = \{N\} \cup \{N - \{k\} \mid k \in N\}$, $\mathcal{L} \in \mathbf{SublangEx}_{\mathcal{A}}$, $\mathbf{UniSublangEx}^1_{\mathcal{A}}$, but not in $\mathbf{UniSublangBc}_{\mathcal{A}}$.*

Proof. Let $\mathrm{cyle}_j = \{\langle j, 2x\rangle \mid x \in N\}$.

Let $\mathrm{cyle}^k_j = \{\langle j, 2x\rangle \mid x \in N, x \neq k\} \cup \{\langle j, 2k+1\rangle\}$

Let $\mathcal{L}_j = \{\mathrm{cyle}_j\} \cup \{\mathrm{cyle}^k_j \mid k \in N\}$.

Let $L_j = \mathrm{cyle}_j$, if $\mathbf{M}_j$ does not **TxtBc**-learn cyle_j. Otherwise, let $L_j = \mathrm{cyle}^{k_j}_j$, where k_j is the least number such that $\langle j, 2k_j\rangle$ does not belong to the least **TxtBc**-locking sequence (say τ_j) for $\mathbf{M}_j$ on cyle_j. Note that on any text extending τ_j for the set $\mathrm{cyle}^{k_j}_j - \{\langle j, 2k_j + 1\rangle\}$, $\mathbf{M}_j$ almost always outputs a grammar for cyle_j.

Let $\mathcal{L} = \{L_j \mid j \in N\}$. It follows immediately by definition of L_j above that $\mathcal{L} \notin \mathbf{UniSublangBc}_{\mathcal{A}}$.

On the other hand, $\mathcal{L}$ is easily seen to be in $\mathbf{SublangEx}_{\mathcal{A}}$. To see, this suppose $A \in \mathcal{A}$ is given. A learner can obtain the unique j such that the input text contains only elements of the form $\langle j, \cdot\rangle$. Now, if the input text contains $\langle j, 2k+1\rangle$, for some k, then the target language must be cyle^k_j and the learner can appropriately converge to a grammar for cyle^k_j; otherwise the target language is either cyle_j or cyle^k_j, for the unique k, if any, which is missing from A, and the learner can converge to a grammar for cyle_j or $\mathrm{cyle}_j - \{\langle j, 2k\rangle\}$ depending on whether $\langle j, 2k\rangle$ belongs to the given input text or not.

Also, $\mathcal{L} \in \mathbf{UniSublangEx}^1_{\mathcal{A}}$, as witnessed by a learner which converges to a grammar for cyle^k_j, if it sees an element of the form $\langle j, 2k+1\rangle$ in the input text; otherwise, the learner converges to a grammar for cyle_j, where the input text only contains elements from cyle_j. ∎

Next, we establish a non-union result: learnability of a class of languages on each of two recursive samplings does not imply uniform learnability of the class on the set consisting of the two given samplings. First we establish a useful proposition.

Proposition 9. *Fix $j, m \in N$ and an IIM $\mathbf{M}$. Let S_j be a subset of $\{\langle j, 1, x\rangle \mid x \leq m\}$.*

Let $Z^f = S_j \cup \{\langle j, 2, \langle x, f(0), f(x)\rangle + m\rangle \mid x \in N\}$.

Then one of the following holds:

(a) there exists an $f \in SD$ such that for an increasing text T for $X_j = Z^f$, for infinitely many n, $W_{\mathbf{M}(T[n])} \cap \{\langle j, 2, y\rangle \mid y \in N\} \neq X_j - S_j$;

(b) there exists an $f \in AZext$ such that for an increasing text T for $X_j = Z^f$, for infinitely many n, $W_{\mathbf{M}(T[n])} \cap \{\langle j, 2, y\rangle \mid y \in N\} \neq X_j - S_j$.

Proof. Follows from Proposition 5. ∎

Now we establish the desired non-union result.

Theorem 10. *Suppose $\mathcal{A}_1 = \{A_1\}$ and $\mathcal{A}_2 = \{A_2\}$, where A_1 and A_2 are two distinct infinite recursive sets. Then,* $\mathbf{SublangEx}_{\mathcal{A}_1} \cap \mathbf{SublangEx}_{\mathcal{A}_2} - \mathbf{UniSublangBc}_{\mathcal{A}_1 \cup \mathcal{A}_2} \neq \emptyset$.

Proof. Without loss of generality we assume that the pairing function is increasing in all its arguments and that, for each j, if $\{w_0, w_1, \ldots\} = \{\langle j, r, x\rangle \mid r \in \{1, 2\}\}$, where for all i, $w_i < w_{i+1}$, then $\text{card}(\{\langle j, 0, x\rangle \mid x \in N\} \cap \{x \mid w_i < x < w_{i+1}\}) \geq (i+2)$-th minimal elements of both A_1 and A_2. This ensures that there are enough gaps between elements of the form $\langle j, 1, \cdot\rangle$ and $\langle j, 2, \cdot\rangle$ to insert several elements of the form $\langle j, 0, \cdot\rangle$ as needed in the construction of L_j from X_j below.

Let $i_0 = \min(A_1 \boldsymbol{\Delta} A_2)$. Without loss of generality assume that $i_0 \in A_1$. Let i_1 be the smallest element in $A_2 - \{x \mid x \leq i_0\}$.

Let $S_j = \{\langle j, 1, x\rangle \mid x < i_0, x \in A_1\} \cup \{\langle j, 1, 2i_0 + 2\rangle\} \cup \{\langle j, 1, 2i_0 + 2 + 2y + 1\rangle \mid i_0 < y \leq i_1, y \in A_1\}$.

Let $m_j = 2i_0 + 2 + 2i_1 + 1$.

Now, if in Proposition 9(a) holds for $S = S_j$, $m = m_j$ and $\mathbf{M} = \mathbf{M}_j$, then let X_j be as in Proposition 9 and let L_j be an r.e. set formed by adding elements of the form $\langle j, 0, \cdot\rangle$ to $X_j \cup \{\langle j, 1, 2i_0 + 2 + 2i_1 + 1\rangle\}$ such that $\text{Order}(X_j \cup \{\langle j, 1, 2i_0 + 2 + 2i_1 + 1\rangle\}, L_j) = A_1 \cup \{i_1\}$ (here we assume that the elements of the form $\langle j, 0, \cdot\rangle$ which are added are the least ones possible so that one can, effectively from X_j, determine the elements which are added). We added $\langle j, 1, 2i_0 + 2 + 2i_1 + 1\rangle$ above just to make sure that the i_1-th element of L_j is $\langle j, 1, 2i_0 + 2 + 2i_1 + 1\rangle$ in case $i_1 \notin A_1$; in case $i_1 \in A_1$, $\langle j, 1, 2i_0 + 2 + 2i_1 + 1\rangle$ would already be in X_j. Note that the i_0-th element of L_j is $\langle j, 1, 2i_0 + 2\rangle$ and the i_1-th element of L_j is $\langle j, 1, 2i_0 + 2 + 2i_1 + 1\rangle$ (that is, the i_0-th element of L_j is of the form $\langle j, 1, 2x\rangle$ for some x, and the i_1-th element of L_j is of the form $\langle j, 1, 2x + 1\rangle$ for some x).

Otherwise, if Proposition 9(b) holds for $S = S_j$, $m = m_j$, and $\mathbf{M} = \mathbf{M}_j$, then let X_j be as in Proposition 9 and let L_j be an r.e. set formed by adding elements of the form $\langle j, 0, \cdot\rangle$ to $X_j \cup \{\langle j, 1, 2i_0 + 1\rangle\}$ such that $\text{Order}(X_j \cup \{\langle j, 1, 2i_0 + 1\rangle\}, L_j) = A_2 \cup \{i_0\}$ (here we assume that the elements of the form $\langle j, 0, \cdot\rangle$ which are added are the least ones possible so that one can, effectively from X_j, determine the elements which are added). Note that the i_0-th element of L_j is $\langle j, 1, 2i_0 + 1\rangle$ and the i_1-th element of L_j is $\langle j, 1, 2i_0 + 2\rangle$ (that is, the i_0-th element of L_j is of the form $\langle j, 1, 2x + 1\rangle$ for some x, and the i_1-th element of L_j is of the form $\langle j, 1, 2x\rangle$ for some x).

Let $\mathcal{L} = \{L_j \mid j \in N\}$. Note that for all j, if one can determine whether (a) or (b) holds in Proposition 9 for $S = S_j$, $m = m_j$, $\mathbf{M} = \mathbf{M}_j$, then from any element for $X_j - S_j$, one can determine X_j and thus L_j. Note that one can determine whether (a) or (b) holds in Proposition 9 for $S = S_j$, $m = m_j$, $\mathbf{M} = \mathbf{M}_j$, from the i_0-th element of L_j. Similarly, one can determine whether (a) or (b) holds in Proposition 9 for $S = S_j$, $m = m_j$, $\mathbf{M} = \mathbf{M}_j$, from the i_1-th element of L_j. Thus, $\mathcal{L} \in \mathbf{SublangEx}_{\mathcal{A}_1}$ and $\mathcal{L} \in \mathbf{SublangEx}_{\mathcal{A}_2}$.

On the other hand, it follows from the construction of X_j and L_j that $\mathbf{M}_j$ cannot **SublangBc**-identify L_j from a text for X_j. It follows that, $\mathcal{L} \notin \mathbf{UniSublangBc}_{\mathcal{A}_1 \cup \mathcal{A}_2}$. ∎

Corollary 11. *There exists a class $\mathcal{A}$ of samplings such that* $\mathbf{PUniSublangEx}_{\mathcal{A}} - \mathbf{UniSublangBc}_{\mathcal{A}} \neq \emptyset$.

The next proposition establishes a sufficient condition for the existence of a uniform learner on the union of the two sets of samplings when a class is learnable on each of the two sets of samplings separately, provided that all learners have access to samplings A (using an oracle).

Proposition 12. *Let $\mathcal{A}_1, \mathcal{A}_2$ be two sets. Suppose there exists a computable F such that F on any text T for A converges to* 1*, if $A \in \mathcal{A}_1 - \mathcal{A}_2$; F on any text T for A converges to* 2*, if $A \in \mathcal{A}_2 - \mathcal{A}_1$; and F on any text T for A converges to either* 1 *or* 2*, if $A \in \mathcal{A}_2 \cap \mathcal{A}_1$. Then,*
$\mathbf{PUniSublangEx}_{\mathcal{A}_1} \cap \mathbf{PUniSublangEx}_{\mathcal{A}_2} \subseteq \mathbf{PUniSublangEx}_{\mathcal{A}_1 \cup \mathcal{A}_2}$.

Now we turn our attention to the problem of extending learners from large natural sets of samplings to larger sets of samplings. For example, can a learner inferring correct grammars on all recursive samplings be extended to a learner on all recursively enumberable samplings? Or, can a learner on all recursively enumerable samplings be extended to a learner on all infinite samplings? So far, we have been able to establish only negative results. Our first result demonstrates that there is a class of languages uniformly learnable from all infinite recursively enumerable samplings, but not learnable from all infinite samplings, even non-uniformly.

We begin with a number of technical propositions.

Proposition 13. *Let $R = \{x_0, x_1, x_2, \ldots\}$, be any infinite recursive set such that, for all i, $x_{i+1} - x_i \geq i + 2$. Then, there exists a recursively enumerable set $S \supseteq R$ such that Order(R, S) is immune, x_0 is the least element of S, and, for all i,* $\mathrm{card}(S \cap \{x \mid x_i < x < x_{i+1}\}) \leq i + 1$.

Proof. Let $R = \{x_0, x_1, \ldots\}$ be as given in the hypothesis of the proposition. By implicit use of Kleene's recursion theorem [Rog67] one can define $S = W_e$ in stages as follows. Initially, W_e contains all of R, and for all k, sat_k is false and $b_k = 0$. Intuitively, if sat_k is true at the beginning of any stage, then

$$Req_k : W_k \text{ intersects with } \mathrm{Order}(S - R, S)$$

is satisfied at the beginning of stage s, and this holds as long as we do not change membership in W_e for elements $\leq b_k$. We will also enumerate at most $r + 1$ elements x in W_e such that $x_r < x < x_{r+1}$.

Go to stage 0.

Stage s

1. If there exists a $k \leq s$ such that sat_k is currently false and $W_{k,s}$ enumerates an element $w > \{b_r \mid r \leq k\}$ and the w-th element of W_e (as of now) is $> x_k$, then pick the least such k and go to step 2. Otherwise, go to stage $s+1$.
2. Let z be the w-th element in W_e as of now.
3. If $z = x_{r+1}$ for some r, then insert a new element $(x_r + k + 1)$ in W_e.
4. Set $sat_k = true$ and set $b_k = z$.
5. Set $sat_{k'} = false$, for all $k' > k$.
6. Go to stage $s+1$.

End stage s

By induction on k, one can show that sat_k eventually takes a fixed value: once $sat_{k'}$ get stabilized for $k' < k$, then sat_k can change at most once, from false to true. Furthermore, if W_k is infinite then sat_k is eventually always true. Thus, Req_k eventually holds for all infintie W_k. Also clearly, the algorithm adds at most $r+1$ elements in between x_r and x_{r+1} to W_e (see the requirement on the w-th element of W_e being $> x_k$ in step 1). It follows that $S = W_e$ satisfies the requirements of the Proposition. ∎

Proposition 14. *Let R, S be as in Proposition 13. Fix $j \in N$. Let $\{z_0, z_1, \ldots\}$, where $z_0 < z_1 < \ldots$, be an infinite recursive set such that, for all i and for all $e \leq i$, number of elements in $\{\langle j, e+1, x\rangle \mid \langle j, 0, z_i\rangle < \langle j, e+1, x\rangle < \langle j, 0, z_{i+1}\rangle\} \geq i+1$.*

Then, for $R_j = \{\langle j, 0, z_i\rangle \mid i \in N\}$, there exists an e and a set $S_j \supseteq R_j$ such that

(a) $S_j - R_j$ is an infinite subset of $\{\langle j, r+1, x\rangle \mid r, x \in N\}$.
(b) for all but finitely many $\langle j, r+1, x\rangle \in S_j - R_j, r = e$,
(c) $W_e = S_j$, and
(d) Order$(R_j, S_j) =$ Order(R, S).

Proof. Let $R_j = \{z_0, z_1, \ldots\}$ be as in the hypothesis of the proposition. Let $R = \{x_0, x_1, \ldots\}$ and S be as in Proposition 13. Then, by implicit use of Kleene's recursion theorem [Rog67] there exists an e such that W_e may be defined as follows.

W_e contains R_j, $\langle j, 0, z_0\rangle$ is the minimal element of W_e and for each i, W_e contains exactly card$(S \cap \{x \mid x_i < x < x_{i+1}\})$ elements from the set:

(i) $\{\langle j, e+1, x\rangle \mid \langle j, 0, z_i\rangle < \langle j, e+1, x\rangle < \langle j, 0, z_{i+1}\rangle\}$, if $i > e$ and
(ii) $\{\langle j, 1, x\rangle \mid \langle j, 0, z_i\rangle < \langle j, 1, x\rangle < \langle j, 0, z_{i+1}\rangle\}$, if $i \leq e$.

Note that, for each i, the gap between $\langle j, 0, z_i\rangle$ and $\langle j, 0, z_{i+1}\rangle$ is large enough to allow the above, and thus, W_e can be so defined. It is easy to verify that $S_j = W_e$ satisfies the requirements of the Proposition. ∎

Proposition 15. *Suppose X is an infinite recursive set and* **M** *is an IIM. Then for all finite $Y \subseteq X$, there exists an infinite recursive Z such that $Y \subseteq Z \subseteq X$, and for an increasing text T for Z, for infinitely many n, $W_{\mathbf{M}(T[n])} \cap X \neq Z$.*

Proof. The proof of $\{L_f \mid f \in \mathcal{R}\}$ not being **TxtBc**-learnable (see [Gol67]), can be easily modified to show this proposition. ∎

Now we can prove the desired result.

Theorem 16. $\mathbf{UniSublangEx}_{REinf} - \mathbf{SublangBc}_{INF} \neq \emptyset$.

Proof. For each j, let $y_0^j < y_1^j < \ldots$ be such that $\{y_0^j, y_1^j, \ldots\}$ is recursive, and for all i and for all $e \leq i$, number of elements in $\{\langle j, e+1, x\rangle \mid \langle j, 0, y_i^j\rangle < \langle j, e+1, x\rangle < \langle j, 0, y_{i+1}^j\rangle\} \geq i+1$.

Let R_j be Z as obtained by Proposition 15, when $Y = \emptyset$, $X = X_j = \{\langle j, 0, y_r^j\rangle \mid r \in N\}$ and $\mathbf{M} = \mathbf{M}_j$. Let S_j be as obtained in Proposition 14 for this R_j.

Let $\mathcal{L} = \{S_j \mid j \in N\}$.

Then, S_j witnesses that $\mathcal{L}$ is not $\mathbf{SublangBc}_{\mathrm{Order}(R,S)}$-identified by $\mathbf{M}_j$. Thus, $\mathcal{L} \notin \mathbf{SublangBc}_{\mathrm{Order}(R,S)}$ and, therefore, $\mathcal{L} \notin \mathbf{SublangBc}_{\mathrm{INF}}$.

On the other hand, $\mathcal{L}$ is clearly in $\mathbf{UniSublangEx}_{\mathrm{REinf}}$, as for any $S_j \in \mathcal{L}$, for any Y such that $\mathrm{Order}(Y, S_j)$ is an infinite r.e. set, we have that Y contains infinitely many elements in $S_j - R_j$. It follows that Y contains infinitely many elements of the form $\langle j, r+1, x\rangle$, and all but finitely many such r are equal to some grammar e for S_j. ∎

The next result shows that the learnability of a class from all simple recursively enumerable samplings does not imply its learnability from all recursively enumerable samplings.

Theorem 17. *Let* $\mathcal{A}_1 = \{A \mid A$ *is an infinite simple set* $\}$. *Let* $A_2 = \{2y \mid y \in N\}$.

Then, $\mathbf{UniSublangEx}_{\mathcal{A}_1} - \mathbf{SublangBc}_{A_2} \neq \emptyset$.

Proof. For each j, let $y_0^j < y_1^j < \ldots$ be such that $\{y_0^j, y_1^j, \ldots\}$ is recursive, and for all i and for all $e \leq i$, number of elements in $\{\langle j, e+1, x\rangle \mid \langle j, 0, y_i^j\rangle < \langle j, e+1, x\rangle < \langle j, 0, y_{i+1}^j\rangle\} \geq 1$.

Let $R_j = Z$, as in Proposition 15 for $\mathbf{M} = \mathbf{M}_j$, $X = X_j = \{\langle j, 0, y_r^j\rangle \mid r \in N\}$ and $Y = \emptyset$.

By implicit use of Kleene's recursion theorem [Rog67], there exists an e such that $W_e = S_j$ satisfies (i) $S_j - R_j \subseteq \{\langle j, y+1, x\rangle \mid x, y \in N\}$, (ii) $\mathrm{Order}(R_j, S_j) = \{2y \mid y \in N\}$, and (iii) for all but finitely many $\langle j, y+1, x\rangle \in S_j$, $y = e$.

Let $\mathcal{L} = \{S_j \mid j \in N\}$. It is easy to verify that $\mathcal{L} \in \mathbf{UniSublangEx}_{\mathcal{A}_1}$, as any simple set intersects with $\{2y+1 \mid y \in N\}$ infinitely often. Also, by Proposition 15, it follows that $\mathbf{M}_j$ does not $\mathbf{SublangBc}_{\mathrm{Order}(R_j,S_j)}$-identify S_j. As $\mathrm{Order}(R_j, S_j)$ is $\{2y \mid y \in N\}$, we have that $\mathcal{L} \notin \mathbf{SublangBc}_{A_2}$. ∎

Our next result shows that, even learnability of a class from all samplings but subsets of one sampling does not imply its learnability from all samplings.

Theorem 18. *Suppose A is an infinite recursive set different from N. Then,* $\mathbf{SublangEx}_{\mathcal{P}(N)-\mathcal{P}(A)} - \mathbf{SublangBc}_A \neq \emptyset$.

Proof. Without loss of generality assume $0 \in A$. (Proof can be easily modified for any other element in A, by considering appropriate modification of SD and AZext).

Let $p_0, p_1, p_2, \ldots$ be a recursive sequence of increasing prime numbers. Let $A = \{x_0, x_1, \ldots\}$, where $x_0 < x_1 < \ldots$.
For $g \in$ SD, let $h_g(x) = p_0^{g(0)} \cdot p_{x+1}^{2g(x)} \cdot \Pi_{y<x}(p_{y+1}^{2g(y)+1})$.
For $g \in$ AZext, let
$h_g(x) = p_0^{g(0)} \cdot p_{x+1}^{2g(x)} \cdot \Pi_{y<x}(p_{y+1}^{2g(y)+1})$, if $x \in A$, and
$h_g(x) = p_0^{g(0)} \cdot p_{x+1}^{2g(x)+1} \cdot \Pi_{y<x}(p_{y+1}^{2g(y)+1})$, if $x \notin A$.

Note that h_g is an increasing function in both the above cases (thus the r-th element of L_{h_g} is $\langle r, h_g(r)\rangle$). Also, every $h_g(x)$ gives away the value of $g(0)$. Furthermore, the only difference in the two cases (of $g \in$ SD or $g \in$ AZext) is how $h_g(x)$ is defined for $x \notin A$.

Let $\mathcal{L} = \{L_{h_g} \mid g \in \text{SD} \cup \text{AZext}\}$.

It is easy to verify that $\mathcal{L} \in \mathbf{SublangEx}_{\mathcal{P}(N)-\mathcal{P}(A)}$, as from any input element $\langle x, h_g(x)\rangle$, for $x \notin A$, one can determine $g(0)$ and whether $g \in$ SD or $g \in$ AZext.

Now for $g, g' \in \text{SD} \cup \text{AZext}$,

(i) one can effectively convert an infinite subgraph for h_g into a graph for g,

(ii) one can effectively convert a graph for g into a graph for h_g restricted to the domain A,

(iii) if $g \neq g'$, then $h_g(x) = h_{g'}(x)$ only for finitely many x.

Thus, $\mathcal{L} \in \mathbf{SublangBc}_A$ implies that $\{L_f \mid f \in \text{SD} \cup \text{AZext}\} \in \mathbf{TxtBc}$, contradicting Theorem 5. It follows that $\mathcal{L} \notin \mathbf{SublangBc}_A$. ∎

A corollary from the above theorem shows that **Ex**-learners from all positive data may sometimes be more powerful than any **Bc**-learner from an A-sampling defined by any infinite recursive sampling $A \neq N$.

Corollary 19. $\mathbf{TxtEx} - \mathbf{SublangBc}_A \neq \emptyset$, *for all infinite recursive* $A \neq N$.

On the other hand, uniform **Bc**-learners from all infinite samplings can sometimes be more powerful than **Ex**-learners from all positive data.

Theorem 20. $\mathbf{UniSublangBc}_{INF} - \mathbf{TxtEx} \neq \emptyset$.

Proof. Let $\mathcal{C} = \{f \mid (\forall^\infty x)[\varphi_{f(x)} = f]\}$. Let $\mathcal{L} = \{L_f \mid f \in \mathcal{C}\}$. Then, it is easy to verify that $\mathcal{L} \in \mathbf{UniSublangBc}_{\text{INF}}$. On the other hand [CS83] showed that $\mathcal{L} \notin \mathbf{TxtEx}$ ([CS83] actually showed this for function learning, which implies the result for language learning). ∎

5 Conclusion

This paper can be viewed as the first step in the study of learnability of various projections of target languages. Firstly, as we mentioned in the Introduction, several different formalizations of the concept of samplings are possible, and the notion of projection itself can be formalized in several ways. For example, a projection of a language may be defined as the set of examples satisfying some predicate — say, when one considers a language of strings over the alphabet $\{a, b, c\}$, a projection may be its sublanguage consisting of all strings over the alphabet $\{a, b\}$.

Yet, even within the framework of formalization suggested in our paper, many interesting questions remain open. We obtained some non-union type results, however, we have not been able to establish more general non-union results for large (and natural, say, containing all recursive samplings) sets of samplings, or, alternatively, find situations when learnability on two different sets of samplings implies uniform learnability on the union of the given sets. The problem of expanding learnability from all samplings from some natural large class to another class of samplings embracing it is also far from being fully explored. For example, we have not been able to find out if learnability from all recursive infinite samplings implies learnability on all recursively enumerable samplings.

References

[BB75] L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.

[Blu67] M. Blum. A machine-independent theory of the complexity of recursive functions. *Journal of the ACM*, 14:322–336, 1967.

[CL82] 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 of *Lecture Notes in Computer Science*, pages 107–115. Springer-Verlag, 1982.

[CS83] J. Case and C. Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science*, 25:193–220, 1983.

[Fre74] R. Freivalds. Uniform and non-uniform predictability. In *Theory of Algorithms and Programs, vol. 1*, pages 89–100. Latvian State University, Riga, Latvia, 1974.

[Ful90] M. Fulk. Prudence and other conditions on formal language learning. *Information and Computation*, 85:1–11, 1990.

[Gol67] E. M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.

[HU79] J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1979.

[JK08] S. Jain and E. Kinber. Learning and extending sublanguages. *Theoretical Computer Science A*, 397(1–3):233–246, 2008. Special Issue on Forty Years of Inductive Inference. Dedicated to the 60th Birthday of Rolf Wiehagen.

[OSW86] D. Osherson, M. Stob, and S. Weinstein. *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists*. MIT Press, 1986.

[OW82] D. Osherson and S. Weinstein. Criteria of language learning. *Information and Control*, 52:123–138, 1982.

[Rog67] H. Rogers. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, 1967. Reprinted by MIT Press in 1987.

[Tra73] B. A. Trakhtenbrot. Frequency computations. In *Proceedings of Steklov Mathematical Institute*, volume 133, pages 221–232. Academy of Sciences of USSR, 1973. In Russian.