Parallel Learning of Automatic Classes of Languages

Sanjay Jain[a,1], Efim Kinber[b]

[a] *School of Computing, National University of Singapore, Singapore 117417.*
[b] *Department of Computer Science, Sacred Heart University, Fairfield, CT 06432-1000, U.S.A.*

Abstract

We introduce and explore a model for parallel learning of families of languages computable by finite automata. In this model, an algorithmic or automatic learner takes on n different input languages and identifies at least m of them correctly. For finite parallel learning, for large enough families, we establish a full characterization of learnability in terms of characteristic samples of languages. Based on this characterization, we show that it is the difference $n-m$, the number of languages which are potentially not identified, which is crucial. Similar results are obtained also for parallel learning in the limit. We consider also parallel finite learnability by finite automata and obtain some partial results. A number of problems for automatic variant of parallel learning remain open.

Keywords: Inductive Inference, Automatic Classes, Parallel Learning.

1. Introduction

In this paper, we define and explore a model for learning *automatic* families of languages *in parallel*. A family of languages is called *automatic* if it is an indexed family, and there is a finite automaton that, given an index v of a language and a string u can solve the membership problem for u in the language indexed by v (study of learnability of automatic classes was initiated in [JLS12]). Our aim is to establish if, under what circumstances, and on what expense, learning several languages from an automatic family in parallel can be more powerful than learning one language at a time. In the past, few approaches to learning in parallel have been suggested. One of them, known as *team inference*, involves a finite team of learning machines working in parallel on the same input function or language (see, for example, [Smi82]). Our approach follows the one suggested for parallel learning recursive functions in [KSVW95]: one learning machine

Email addresses: `sanjay@comp.nus.edu.sg` (Sanjay Jain), `kinbere@sacredheart.edu` (Efim Kinber)

[1]Supported in part by NUS grant numbers C-252-000-087-001, R-146-000-181-112 and R-252-000-534-112.

is learning a finite collection of (pairwise distinct) languages (in some sense, this model is a generalization of the model introduced in [AGS89]). A similar approach has recently been utilized in a study of prediction of recursive function values in [BKF11]: one algorithm predicts next values of several different input functions.

We consider learning languages in two different, albeight related settings:

a) Finite learning [Gol67]: a learning machine, after seeing a finite amount of input data, terminates and outputs conjectures for grammars of languages being learnt.

b) Learning in the limit [Gol67]: a learning machine outputs a potentially infinite sequence of conjectures, stabilizing on a correct grammar for the target language.

For above type of learning, we also consider the case when the learner itself is a finite automaton (see [JLS12]). The learners in our model use input *texts* — potentially infinite sequences that contain full positive data in a target language, intermittent with periods of "no data". Both settings, under the name of inductive inference, have a long history, see, for example, [JORS99].

A simple example of the family of three languages, $\{0\}, \{1\}, \{0, 1\}$ (which can be trivially made automatic), shows that finite learning of three languages in parallel might be possible, whereas no learner can finitely learn languages in the family one at a time: the desired parallel learner will just wait until the three input texts contain 0, 1 and $0, 1$ respectively and then output three correct conjectures based on which texts contain which of the above elements; on the other hand, if an individual learner gets on the input a text containing all 0-s and settles on the conjecture $\{0\}$, it will be too late if 1 appears in the input.

However, interestingly, when families of languages are large, finite parallel learning of all input languages has no advantage over finite learning of individual languages: as it follows from one of our results (Theorem 9), if the number of languages in an automatic family is at least 4, and the family is learnable in parallel by a finite learner taking three different input texts, then the family is finitely learnable, one language at a time. Therefore, we consider a more general model of parallel learning, where the potential advantage of parallelism may compensate for lack of precision — so-called (m, n) or *frequency* learning: a learner gets input texts for n different languages and learns at least m of them correctly. This model of learning was first suggested and explored for algorithmic learning of recursive functions in [KSVW95]. The idea of frequency learning stems from a more general idea of (m, n)-computation, which, in the recursion-theoretic setting, means the following: to compute a function, an algorithm takes on n different inputs at a time and outputs correct values on at least m inputs. This idea can be traced to the works by G. Rose [Ros60] and B.A. Trakhtenbrot [Tra64] who suggested frequency computation as a deterministic alternative to traditional probabilistic algorithms using randomization. Since then, this idea has been applied to various settings, from computation by finite automata ([Kin76, ADHP05]) to computation with a small number of bounded queries ([BGK96]).

We explore and, whenever it has been possible, determine what makes automatic classes of languages (m, n)-learnable for various numbers n and $m \leq n$. Whereas, in our general model, it is not possible to identify which m conjectures among n are correct, we also consider the special case of finitely learning automatic classes when the learner can identify m correct conjectures.

In the theory of language learning, a prominent role belongs to the so-called *characteristic samples* ([LZ92, Muk92]), and *tell-tale sets* (see [Ang80]). A finite subset D of a language L is called a characteristic sample of L (with respect to the family of languages under consideration) if, for every language L' in the family, $D \subseteq L'$ implies $L' = L$. A finite subset D of a language L is called a tell-tale set of L (with respect to the family of languages under consideration) if, for every language L' in the family, $D \subseteq L'$ implies $\neg[L' \subset L]$.

A family of languages satisfies characteristic sample condition (tell-tale set condition) if every language in it has a characteristic sample (has a tell-tale set) with respect to the family. Several of our characterizations of (m, n)-learnability are based on suitable variants of the characteristic sample or tell-tale set condition. Since in all our settings (m, n)-learning (for $m < n$) turns out to be more powerful than learning individual languages, we study and discover interesting relationships between classes of languages (m, n)-learnable with different parameters m and/or n. In particular, we are concerned with the following questions:

a) does (m, n)-learnability imply $(m+1, n+1)$-learnability of a class? (thus, increasing frequency of correct conjectures, while keeping the number of possibly erroneous conjectures the same);

b) does $(m + 1, n + 1)$-learnability imply (m, n)-learnability? (thus, loosing in terms of frequency of correct conjectures, but allowing a smaller number of languages to be learnt in parallel, with the same number of possibly erroneous conjectures);

c) does $(m, n+1)$-learnability imply (m, n)-learnability? (thus, reducing the number of possibly erroneous conjectures and increasing frequency of correct conjectures at the same time).

For each of our variants of learnability, we obtain either full or partial answers to all the above questions, for large enough families.

The structure of our study of (m, n)-learning is as follows. In the next section, we introduce necessary mathematical preliminaries and notation. In Section 3 we formally define our learning models. In Section 4, we take on the case of finite (m, n)-learning when a learner can specify at least m of the texts which it learnt correctly — following [KSVW95], we call (m, n)-learning of this kind *superlearning*. In Theorems 8 and 9, for the classes containing at least $2n+1-m$ languages, we give a full characterization for (m, n)-superlearnability in terms of characteristic samples. For large classes of languages, this characterization provides us full positive answers to the above questions a), b), and the negative answer to c). We also address the case when the number of languages in a class to be learnt is smaller than $2n + 1 - m$, providing, in particular, a different characterization for (m, n)-superlearnability for this case.

In Section 5 we consider finite (m, n)-learning when a learner may not be able to tell which of the m texts it learns. For large classes of languages, we again obtain a full characterization of (m, n)-learnability in terms of characteristic samples — albeight somewhat different from the case of superlearnability. This characterization, as in case of superlearnability, provides us answers to the questions a), b), and c). The proofs in this section are quite involved — to obtain necessary results, we developed a technique based on bipartite graphs. We then also obtain a different characterization for finite (m, n)-learning of finite automatic classes.

In Section 6 we obtain full positive answers for the questions a) and b) and the negative answer to the question c) for (m, n)-learnability of automatic classes in the limit. We also obtain characterizations of this kind of learning in terms of presence of tell-tale sets [Ang80] for languages in the class being learnt.

In Section 7, we address finite (m, n)-learning by finite automata — automatic learning. We have not been able to come up with a characterization of this type of learnability, however, we answer positively to the question b) and negatively to the question c). The question a) remains open. For finite superlearning by finite automata, we give positive answers to questions a) and b) for large enough classes, and negative answer to question c). In Section 8, we address (m, n)-learning by finite automata in the limit.

2. Preliminaries

The set of natural numbers, $\{0, 1, 2, \ldots\}$, is denoted by N. We let Σ denote a finite alphabet. The set of all strings over the alphabet Σ is denoted by Σ^*. A language is a subset of Σ^*. The length of a string x is denoted by $|x|$. We let ϵ denote the empty string.

A string $x = x(0)x(1)\ldots x(n-1)$ is identified with the corresponding function from $\{0, 1, \ldots, n-1\}$ to Σ. We assume some canonical ordering of members of Σ. Lexicographic order is then the dictionary order over strings. A string w is length-lexicographically before (or smaller than) string w' (written $w <_{ll} w'$) iff $|w| < |w'|$ or $|w| = |w'|$ and w is lexicographically before w'. Furthermore, $w \leq_{ll} w'$ denotes that either $w = w'$ or $w <_{ll} w'$. For any set of strings S, let $\mathrm{succ}_S(w)$ denote the length-lexicographically least w' such that $w' \in S$ and $w <_{ll} w'$ — if there is no such string, then $\mathrm{succ}_S(w)$ is undefined.

We let $\emptyset, \subseteq$ and $\subset$ respectively denote empty set, subset and proper subset. The cardinality of a set S is denoted by $\mathrm{card}(S)$. $A\boldsymbol{\Delta}B$ denotes the symmetric difference of A and B, that is, $(A - B) \cup (B - A)$.

We now define the convolution of two strings $x = x(0)x(1)\ldots x(n-1)$ and $y = y(0)y(1)\ldots y(m-1)$, denoted $conv(x, y)$. Let x', y' be strings of length $\max(\{m, n\})$ such that $x'(i) = x(i)$ for $i < n$, $x'(i) = \#$ for $n \leq i < \max(\{m, n\})$, $y'(i) = y(i)$ for $i < m$, and $y'(i) = \#$ for $m \leq i < \max(\{m, n\})$, where $\# \notin \Sigma$ is a special padding symbol. Thus, x', y' are obtained from x, y by padding the smaller string with #'s. Then, $conv(x, y) = z$, where $|z| = \max(\{m, n\})$ and $z(i) = (x'(i), y'(i))$, for $i < \max(\{m, n\})$. Here, note that z is a string over the alphabet $(\Sigma \cup \{\#\}) \times (\Sigma \cup \{\#\})$. Intuitively, giving a convolution of two strings as input to a machine means giving the two strings in parallel, with the shorter string being padded with #s. The definition of convolution of two strings can be easily generalized to convolution of more than two strings. An n-ary relation R is *automatic*, if $\{conv(x_1, x_2, \ldots, x_n) : (x_1, x_2, \ldots, x_n) \in R\}$ is regular. Similarly, an n-ary function f is automatic if $\{conv(x_1, x_2, \ldots, x_n, y) : f(x_1, x_2, \ldots, x_n) = y\}$ is regular.

A family of languages, $(L_\alpha)_{\alpha \in I}$, over some finite alphabet Σ, is called an *automatic family* if (a) the index set I is regular and (b) the set $\{conv(\alpha, x) : \alpha \in I, x \in L_\alpha\}$ is regular. We often identify an automatic family $(L_\alpha)_{\alpha \in I}$ with the class $\mathcal{L} = \{L_\alpha : \alpha \in I\}$, where the indexing is implicit. An automatic family $(L_\alpha)_{\alpha \in I}$ is 1–1 (or the indexing is 1–1), if for all $\alpha, \beta \in I$, $L_\alpha = L_\beta$ implies $\alpha = \beta$.

It can be shown that any family, relation or function that is first-order definable using other automatic relations or functions is itself automatic.

Lemma 1. *[BG00, KN95] Any relation that is the first-order definable from existing automatic relations is automatic.*

We use the above lemma implicitly in our proofs, without explicitly stating so. The example below gives some well-known automatic families.

Example 2. **(a)** *For any fixed k, the class of all subsets of Σ^* having at most k elements is an automatic family.*

(b) *The class of all finite and cofinite subsets of $\{0\}^*$ is an automatic family.*

(c) *The class of closed intervals, consisting of languages $L_{Conv(\alpha,\beta)} = \{x \in \Sigma^* : \alpha \leq_{lex} x \leq_{lex} \beta\}$ where $\alpha, \beta \in \Sigma^*$, over the alphabet Σ is an automatic family.*

3. Learning Automatic Families

A *text* T is a mapping from N to $\Sigma^* \cup \{\#\}$. The content of a text T, denoted content(T), is $\{T(i) : i \in N\} - \{\#\}$. A text T is for a language L iff content$(T) = L$. Intuitively, #'s denote pauses in the presentation of data. Furthermore, $\#^\infty$ is the only text for $\emptyset$.

Let $T[n]$ denote $T(0)T(1)\ldots T(n-1)$, the initial sequence of T of the length n. We let σ and τ range over finite initial sequences of texts. The length of σ is denoted by $|\sigma|$. For $n \leq |\sigma|$, $\sigma[n]$ denotes $\sigma(0)\sigma(1)\ldots\sigma(n-1)$. The empty sequence is denoted by Λ. Let content$(\sigma) = \{\sigma(i) : i < |\sigma|\}$.

We now consider learning machines. Since we are considering parallel learning, we directly define learners which take as input n texts. Furthermore, to make it easier to define automatic learners, we define the learners as mapping from the current memory and the new datum, to the new memory and conjecture (see [JLS12]). When one does not have any memory constraints (as imposed, for example, by automatic learning requirement), these learners are equivalent to those defined by Gold [Gol67]. The learner uses some hypothesis space $\{H_\alpha : \alpha \in J\}$ to interpret its hypothesis. We always require (without explicitly stating so) that $\{H_\alpha : \alpha \in J\}$ is a uniformly r.e. class (that is, $\{(x,\alpha) : x \in H_\alpha\}$ is r.e.). Often the hypothesis space is even required to be an automatic family, with the index set J being regular. For this paper, without loss of generality, we assume that the learners are total.

Definition 3. (Based on [Gol67, JLS12]) Suppose Σ and Δ are finite alphabets used for languages and memory of learners respectively, where $\# \notin \Sigma$. Suppose J is the index set (over some finite alphabet) for the hypothesis space used by the learner. Let ? be a special symbol not in J. Suppose $0 < n$.

(a) A *learner* (from n-texts) is a recursive mapping from $\Delta^* \times (\Sigma^* \cup \{\#\})^n$ to $\Delta^* \times (J \cup \{?\})^n$.

A learner has an initial memory $mem_0 \in \Delta^*$, and an initial hypotheses $(hyp_1^0, hyp_2^0, \ldots, hyp_n^0) \in (J \cup \{?\})^n$.

(b) Suppose a learner $\mathbf{M}$ with the initial memory mem_0 and the initial hypotheses $hyp_1^0, hyp_2^0, \ldots, hyp_n^0$ is given. Suppose $T_1, T_2, \ldots, T_n$ are n texts. Then the definition of $\mathbf{M}$ is extended to sequences as follows.

$\mathbf{M}(\Lambda, \Lambda, \ldots, \Lambda) = (mem_0, hyp_1^0, hyp_2^0, \ldots, hyp_n^0)$;

$\mathbf{M}(T_1[s+1], T_2[s+1], \ldots, T_n[s+1]) = \mathbf{M}(mem, T_1(s), T_2(s), \ldots, T_n(s))$, where $\mathbf{M}(T_1[s], T_2[s], \ldots, T_n[s]) = (mem, hyp_1, hyp_2, \ldots, hyp_n)$, for some $(hyp_1, hyp_2, \ldots, hyp_n) \in (J \cup \{?\})^n$ and $mem \in \Delta^*$.

(c) We say that $\mathbf{M}$ converges on $T_1, T_2, \ldots, T_n$ to hypotheses $(\beta_1, \beta_2, \ldots, \beta_n) \in (J \cup \{?\})^n$ (written: $\mathbf{M}(T_1, T_2, \ldots, T_n)\downarrow_{hyp} = (\beta_1, \beta_2, \ldots, \beta_n)$) iff there exists a t such that, for all $t' \geq t$,

$\mathbf{M}(T_1[t'], T_2[t'], \ldots, T_n[t']) \in \Delta^* \times \{(\beta_1, \beta_2, \ldots, \beta_n)\}$.

Intuitively, $\mathbf{M}(T_1[s], T_2[s], \ldots, T_2[s]) = (mem, hyp_1, hyp_2, \ldots, hyp_n)$ means that the memory and the hypotheses of the learner $\mathbf{M}$ after having seen the initial parts $T_1[s], T_2[s], \ldots, T_n[s]$ of the n texts are mem and $hyp_1, hyp_2, \ldots, hyp_n$, respectively. We call hyp_i above the hypothesis of the learner on the text T_i.

We call a learner automatic if the corresponding graph of the learner is automatic. That is, $\{conv(mem, x_1, x_2, \ldots, x_n, newmem, \beta_1, \beta_2, \ldots, \beta_n) : \mathbf{M}(mem, x_1, x_2, \ldots, x_n) = (newmem, \beta_1, \beta_2, \ldots, \beta_n)\}$ is regular.

We can think of a learner as receiving the texts $T_1, T_2, \ldots, T_n$ one element at a time, from each of the texts. At each input, the learner updates its previous memory, and outputs a new conjecture (hypothesis) for each of the texts. If the sequence of hypotheses converges to a grammar for content(T), then we say that the learner **TxtEx**-learns the corresponding text ([Gol67]). Here **Ex** denotes "explains", and **Txt** denotes learning from text. For parallel (m,n)-learnability, we require that the learner converges to a correct grammar for at least m out of the n input texts. Now we define learnability formally.

Definition 4. (Based on [Gol67, KSVW95])

Suppose $\mathcal{L} = \{L_\alpha : \alpha \in I\}$ is a target class, and $\mathcal{H} = \{H_\beta : \beta \in J\}$ is a hypothesis space. Suppose $0 < m \leq n$.

(a) We say that $\mathbf{M}$ (m,n)-**TxtEx**-learns the class $\mathcal{L}$ (using $\mathcal{H}$ as the hypothesis space) iff for all n-texts $T_1, T_2, \ldots, T_n$ for pairwise distinct languages in $\mathcal{L}$, $\mathbf{M}(T_1, T_2, \ldots, T_n)\downarrow_{hyp} = (\beta_1, \beta_2, \ldots, \beta_n)$ such that for at least m different $i \in \{1, 2, \ldots, n\}$, $\beta_i \in J$ and $H_{\beta_i} = \text{content}(T_i)$.

(b) (m,n)-**TxtEx** $= \{\mathcal{L} : (\exists \text{ learner } \mathbf{M})[\mathbf{M}\ (m,n)\text{-}\mathbf{TxtEx}\text{-learns } \mathcal{L} \text{ using some } \mathcal{H} \text{ as the hypothesis space}]\}$.

(c) We say that $\mathbf{M}$ (m,n)-**TxtFin**-learns the class $\mathcal{L}$ (using $\mathcal{H}$ as the hypothesis space) iff for all n-texts $T_1, T_2, \ldots, T_n$ for pairwise distinct languages in $\mathcal{L}$, there exists an s and $\beta_1, \ldots, \beta_n \in J$, such that, for all $s' < s$ and $s'' \geq s$:

(i) $\mathbf{M}(T_1[s'], T_2[s'], \ldots, T_n[s']) \in \Delta^* \times (?, ?, \ldots, ?)$ (where there are n ? in the above);

(ii) $\mathbf{M}(T_1[s''], T_2[s''], \ldots, T_n[s'']) \in \Delta^* \times \{(\beta_1, \beta_2, \ldots, \beta_n)\}$;

(iii) for at least m pairwise distinct $i \in \{1, 2, \ldots, n\}$; $H_{\beta_i} = \text{content}(T_i)$.

(d) (m,n)-**TxtFin** $= \{\mathcal{L} : (\exists \text{ learner } \mathbf{M})[\mathbf{M}\ (m,n)\text{-}\mathbf{TxtFin}\text{-learns } \mathcal{L} \text{ using some } \mathcal{H} \text{ as the hypothesis space}]\}$.

Intuitively, for (m,n)-**TxtFin**-learning the first conjecture of $\mathbf{M}$, different from $(?, ?, \ldots, ?)$, is such that at least m of the conjectures are correct for the corresponding texts. Thus, we often say that the finite learner stops after outputting the first non $(?, ?, \ldots, ?)$ hypothesis and alternatively, say and assume that it just repeats its hypothesis after that.

We use the terms "learning" and "identifying" as synonyms. We often refer to $(1,1)$-**TxtFin**-learning as just **TxtFin**-learning. Similar convention applies to other criteria of learning considered in this paper.

We drop the reference to "using the hypothesis space $\mathcal{H}$", when the hypothesis space is clear from the context.

For (m,n)-**superTxtEx** or (m,n)-**superTxtFin**-learnability, we require a learner (called superlearner in this case) to also specify/mark texts which it has learnt (at least m of them). In some sense, the learner guarantees that the marked texts have been learnt correctly, where it marks at least m of the texts. This can be done as follows. The learner, along with its n conjectures $(\beta_1, \beta_2, \ldots, \beta_n)$, outputs an indicator $(i_1, i_2, \ldots, i_n)$, where $i_j \in \{0, 1\}, 1 \leq j \leq n$. Suppose the conjecture/indicator of the learner after having seen the input $T_1[s], T_2[s], \ldots, T_n[s]$ are $(\beta_1^s, \beta_2^s, \ldots, \beta_n^s)$ and $(i_1^s, i_2^s, \ldots, i_n^s)$, respectively. Then, for (m,n)-**superTxtEx**-learning, if the n texts $T_1, T_2, \ldots, T_n$ given as input are for distinct languages in the class $\mathcal{L}$ being learnt, we require that $\lim_s i_j^s$ as well as $\lim_s \beta_j^s$ converge for each $j \in \{1, 2, \ldots, n\}$, where for at least m different values of j, $\lim_s i_j^s$ converges to 1. Furthermore, whenever $\lim_s i_j^s$ converges to 1, $\lim_s \beta_j^s$ converges to some β_j such that H_{β_j} equals content(T_j). Requirements for (m,n)-**superTxtFin** learning can be stated similarly, demanding that the learner does not change its indicators/conjectures after the first time it outputs a non $(?, ?, \ldots, ?)$ conjecture.

In the sequal, we do not mention indicators explicitly. Rather, we often simply say which texts (at least m) have been specified or marked as having been learnt correctly (or simply learnt).

When we are considering automatic learners (that is, learners, whose graphs are regular [JLS12]), we prefix the learning criterion **TxtEx** or **TxtFin** by **Auto**. For this we also require the hypothesis space used to be an automatic family.

Trivially, (n,n)-**TxtFin** is the same as (n,n)-**superTxtFin** and (n,n)-**TxtEx** is the same as (n,n)-**superTxtEx**. Similar results hold for learning by automatic learners. Furthermore, every class containing $< n$ languages is trivially (m,n)-**superTxtFin** learnable (as there are no texts $T_1, T_2, \ldots, T_n$ for pairwise different languages in $\mathcal{L}$). Thus, in the results and discussion below, for (m,n)-learnability, we often implicitly or explicitly only consider the cases when $\mathcal{L}$ has at least n elements.

Except for automatic learners considered in Section 7 and Section 8, all learners considered in this paper can memorize the whole input. Thus, for such learners, for ease of notation, we usually ignore *mem* in the output of $\mathbf{M}(T_1[s], T_2[s], \ldots, T_2[s])$ and just consider $(hyp_1, hyp_2, \ldots, hyp_n)$ as the output of the learner. Furthermore, for ease of notation, rather than giving the learner as a mapping from memory and n-input elements to new memory and conjecture, we often just give an informal description of the learner, where for finite learning the learner will output only one conjecture different from $(?, ?, \ldots, ?)$. It will be clear from the context how the formal learners can be obtained from the description.

We now consider some useful concepts from the literature.

Definition 5. [LZ92, Muk92] We say that S is a *characteristic sample* for L with respect to $\mathcal{L}$ iff (a) S is a finite subset of L and (b) for all $L' \in \mathcal{L}$, $S \subseteq L'$ implies $L = L'$.

Definition 6. [Ang80] We say that S is a *tell-tale set* for L with respect to $\mathcal{L}$ iff (a) S is a finite subset of L and (b) for all $L' \in \mathcal{L}$, $S \subseteq L'$ implies $\neg[L' \subset L]$.

Note that L having characteristic sample with respect to $\mathcal{L}$ implies that L has tell-tale set with respect to $\mathcal{L}$. Also, if $L \in \mathcal{L}$ does not have a tell-tale with respect to $\mathcal{L}$, then for each finite subset S of L, there exist infinitely many $L' \in \mathcal{L}$ which contain S. To see this, suppose $L \in \mathcal{L}$, S is a finite subset of L and there are only finitely many $L' \in \mathcal{L}$ which contain S. Then, let S' be the set of minimal elements in $L - L'$, for each of these L'. Now $S \cup S'$ will be a tell-tale set for L with respect to $\mathcal{L}$.

For a given automatic family $\mathcal{L}$, using Lemma 1, it is easy to see that testing whether or not a finite set S is a characteristic sample (or a tell-tale set) of $L \in \mathcal{L}$ with respect to the automatic family $\mathcal{L}$ is decidable effectively in S and index for L. This holds, as we can express the characteristic sample (or a tell-tale set) property as a first order formula. For example, for characteristic sample: S is a characteristic sample for L_β with respect to $\mathcal{L}$ (with index set I) iff $S \subseteq L_\beta$

and $(\forall\alpha \in I)[S \subseteq L_\alpha \Rightarrow (\forall x \in \Sigma^*)[x \in L_\beta \Leftrightarrow x \in L_\alpha]]$, where it is easy to see that subset property can be expressed as a first order formula for a fixed finite set S.

The following lemma is used implicitly in several proofs, without explicitly referring to it.

Lemma 7. *Suppose $\mathcal{L}$ and a language $L \in \mathcal{L}$ are given such that L does not have a characteristic sample with respect to $\mathcal{L}$. Then, either (a) there exists $L' \in \mathcal{L}$ such that $L \subset L'$ or (b) for all n, there are $X_n \in \mathcal{L}$, such that X_n are pairwise distinct and $L \cap \{x : x \leq n\} \subseteq X_n \cap \{x : x \leq n\}$.*

Proof. If L does not have a characteristic sample with respect to $\mathcal{L}$ then, for all n, there exists a $Y_n \in \mathcal{L}$, $Y_n \neq L$, such that $L \cap \{x : x \leq n\} \subseteq Y_n \cap \{x : x \leq n\}$. Fix such Y_n. If some Y equals Y_n for infinitely many n, then clearly this $L' = Y$ satisfies (a). Otherwise, there exists a subsequence $X_1 = Y_{i_1}, X_2 = Y_{i_2}, \ldots$ of pairwise different sets, where $i_1 < i_2 < \ldots$, such that X_n's satisfy (b). ∎

4. (m, n)-superTxtFin-learnability

The next two theorems give a full characterization of (m, n)-**superTxtFin**-learnability for large automatic classes.

Theorem 8. *Suppose $0 < m \leq n$. Suppose $\mathcal{L}$ is an automatic family, and for all except at most $n - m$ $L \in \mathcal{L}$, there exists a characteristic sample for L with respect to $\mathcal{L}$. Then $\mathcal{L}$ is (m, n)-**superTxtFin**-learnable.*

Proof. Suppose $\mathcal{L} = \{L_\alpha : \alpha \in I\}$ is an automatic family, where I is a regular index set. As $\mathcal{L}$ is automatic, by Lemma 1, given a finite set S and an index $\alpha \in I$, one can effectively check if S is a characteristic sample for L_α with respect to $\mathcal{L}$.

The desired learner $\mathbf{M}$, on any input texts $T_1, \ldots, T_n$, searches for an r, a subset $X \subseteq \{1, 2, \ldots, n\}$ of size m, and α_i for $i \in X$ such that for each $i \in X$, content$(T_i[r])$, is a characteristic sample for L_{α_i} with respect to $\mathcal{L}$ (before finding such an r, $\mathbf{M}$ conjectures ? for all the texts). When $\mathbf{M}$ finds such an r, X and corresponding α_i, $i \in X$, $\mathbf{M}$ conjectures hypothesis α_i on T_i, $i \in X$, and specifies them as having been learnt (the conjectures on remaining texts T_i for $i \notin X$ are irrelevant, thus these texts are specified as not learnt). It is easy to verify that $\mathbf{M}$ (m, n)-**superTxtFin**-learns $\mathcal{L}$. ∎

The following result shows that the above result is optimal for large enough classes of languages. For small finite classes, as illustrated by Remark 15 below, such characterization does not hold.

Theorem 9. *Suppose $0 < m \leq n$. If an automatic class $\mathcal{L}$ has at least $2n+1-m$ languages, then (m, n)-**superTxtFin**-learnability of $\mathcal{L}$ implies there are at most $n - m$ languages in $\mathcal{L}$ which do not have a characteristic sample with respect to $\mathcal{L}$.*

Proof. Suppose, by way of contradiction, otherwise. Pick at least $n-m+1$ languages in the class $\mathcal{L}$ which do not have a characteristic sample with respect to $\mathcal{L}$. Let these languages be $A_1, A_1, \ldots, A_{n-m+1}$.

For $1 \leq r \leq n-m+1$, let $B_r \in \mathcal{L} - \{A_1, A_2, \ldots, A_{n-m+1}\}$ be a language in $\mathcal{L}$ which is a superset of A_r. If there is no such language, then B_r is taken to be an arbitrary member of $\mathcal{L} - \{A_1, A_2, \ldots, A_{n-m+1}\}$. The B_r's may not be different from each other.

Note that if B_r is not a superset of A_r, then by Lemma 7, there exist infinitely many pairwise distinct languages $S_r^w \in \mathcal{L}$, $w \in \Sigma^*$, such that each S_r^w contains $A_r \cap \{x : x \leq_{ll} w\}$.

Now consider the behaviour of a superlearner on the texts $T_1, T_2, \ldots, T_n$ for languages $A_1, A_2, \ldots, A_{n-m+1}, C_{n-m+2}, \ldots, C_n$, where $C_{n-m+2}, \ldots, C_n$ are pairwise distinct members of $\mathcal{L}$ which are different from A_r, B_r, $1 \leq r \leq n-m+1$. Suppose the superlearner outputs its conjecture (different from $(?, ?, \ldots, ?)$) after seeing input $T_1[s], T_2[s], \ldots, T_n[s]$. As the superlearner identifies (and specifies as identified) at least m languages, it has to specify and identify at least one $T_1, T_2, \ldots, T_{n-m+1}$, say T_r. Suppose $\text{content}(T_r[s]) \subseteq \{x : x \leq_{ll} w\}$ Then one can replace A_r by B_r or by an appropriate one of $S_r^{w'}$, $w' \geq_{ll} w$, which is not among $A_1, A_2, \ldots, A_{n-m+1}, C_{n-m+2}, \ldots, C_n$, thus making the superlearner fail. ∎

The following corollaries easily follow from the above two theorems.

Corollary 10. *Suppose $0 < n$. If an automatic class $\mathcal{L}$ contains at least $n+1$ languages and is (n, n)-**superTxtFin**-learnable, then every language in $\mathcal{L}$ has a characteristic sample with respect to $\mathcal{L}$ and, thus, the class is **TxtFin**-learnable.*

Corollary 11. *Consider an automatic class $\mathcal{L}$. $\mathcal{L}$ is **TxtFin**-learnable iff every language in $\mathcal{L}$ has a characteristic sample with respect to $\mathcal{L}$.*

The next corollary shows that superlearnability for large enough classes can be preserved if the number n increases or decreases, as long as the number of errors $n-m$ is required to remain the same.

Corollary 12. *Suppose $0 < m \leq n$. Suppose $\mathcal{L}$ is a large enough automatic class (that is, it contains at least $2n-m+1$ languages). Then, $\mathcal{L}$ is (m, n)-**superTxtFin**-learnable iff it is $(m+1, n+1)$-**superTxtFin**-learnable.*

On the other hand, decreasing the number n of superlearned languages while preserving the number m of learnt languages is not always possible, as the following Corollary shows (it follows from Corollary 31 shown in Section 5 below).

Corollary 13. *Suppose $0 < m < n$. There exists an automatic class $\mathcal{L}$ that is (m, n)-**superTxtFin**-learnable, but not $(m, n-1)$-**superTxtFin**-learnable.*

For $m = 1$, Theorem 9 can be strengthened to

Theorem 14. *Suppose $0 < n$. Suppose an automatic class $\mathcal{L}$ contains at least n languages L which do not have a characteristic sample with respect to $\mathcal{L}$. Then $\mathcal{L}$ is not $(1, n)$-**superTxtFin**-learnable.*

Proof. Suppose $A_1, A_2, \ldots, A_n$ are n languages in $\mathcal{L}$ which do not have characteristic samples with respect to $\mathcal{L}$. Let $B_1, B_2, \ldots, B_n \notin \{A_1, A_2, \ldots, A_n\}$ be such that $A_i \subseteq B_i$ — if there is no such i, then we take B_i to be arbitrary member of $\mathcal{L} - \{A_1, A_2, \ldots, A_n\}$. Note that if B_i is not a superset of A_i, then by Lemma 7, there exist infinitely many pairwise distinct $S_i^w \in \mathcal{L}$, (where $w \in \Sigma^*$) such that $S_i^w \supseteq A_i \cap \{x : x \leq_{ll} w\}$.

Now consider the behaviour of a superlearner on the texts $T_1, T_2, \ldots, T_n$ for the languages $A_1, A_2, \ldots, A_n$. Suppose the superlearner outputs its conjecture after seeing input $T_1[s], T_2[s], \ldots, T_n[s]$. Suppose the superlearner identifies (and specifies as identified) T_r. Now suppose content$(T_r[s]) \subseteq \{x : x \leq_{ll} w\}$. Then, one can replace A_r by B_r or by an appropriate one of $S_r^{w'}$, $w' \geq_{ll} w$, which is not among $A_1, A_2, \ldots, A_n$, thus making the superlearner fail. ∎

Note that the requirement of the class size being at least $2n-m+1$ is needed in some cases for the results above. This can be seen from the remark below.

Remark 15. *Let $2 \leq m \leq n$.*

Let $L_{2r} = \{a^{2r}\}, L_{2r+1} = \{a^{2r}, a^{2r+1}\}$, for $r \leq n-m$.

Let $L_i = \{a^i\}$, for $2n-2m+2 \leq i < 2n-m$.

Let $\mathcal{L} = \{L_i : i < 2n-m\}$.

*Now, $\mathcal{L}$ contains $n-m+1$ languages (L_{2r}, for $r \leq n-m$) which do not have a characteristic sample with respect to $\mathcal{L}$. However, $\mathcal{L}$ is (m,n)-**superTxtFin**-learnable. To see (m,n)-**superTxtFin**-learnability, note that in any collection of n languages from $\mathcal{L}$, there can be at most $n-m$ different $s \leq n-m$ such that the collection contains L_{2s} but not L_{2s+1}. Note that if the collection contains both L_{2s} and L_{2s+1}, then we can identify both of them, from texts, as the languages given as input to (m,n)-**superTxtFin**-learner are supposed to be different. Thus, one can easily (m,n)-**superTxtFin**-learn the class $\mathcal{L}$.*

Thus, now we consider finite classes (which covers also the case of classes of size at most $2n-m$) below. For these classes, we obtain a different characterization of (m,n)-**TxtFin**-superlearnability.

Theorem 16. *Suppose $0 < m < n$. Consider any finite automatic class $\mathcal{L}$ of cardinality at least n. Then $\mathcal{L}$ is (m,n)-**superTxtFin**-learnable iff, for all $\mathcal{S} \subseteq \mathcal{L}$ of cardinality n, there are at most $n-m$ languages in $\mathcal{S}$ which have a superset in $\mathcal{L} - \mathcal{S}$.*

Proof. Suppose there exists $\mathcal{S} \subseteq \mathcal{L}$ of cardinality n such that there are $n-m+1$ languages in $\mathcal{S}$ which have a superset in $\mathcal{L}-\mathcal{S}$. Then if the learner is given texts for the n languages in $\mathcal{S}$, it must fail to super-**TxtFin** learn at least $n-m+1$ of them (as all the languages which have supersets in $\mathcal{L}-\mathcal{S}$ cannot be learnt (along with being specified as learnt)). Thus the learner cannot (m,n)-**superTxtFin**-learn $\mathcal{L}$.

On the other hand, if all $\mathcal{S} \subseteq \mathcal{L}$ of cardinality n have at most $n-m$ languages which are contained in some language in $\mathcal{L} - \mathcal{S}$, then consider the following learner which (m,n)-**superTxtFin**-learns $\mathcal{L}$.

Let $S = \{\min(L - L') : L, L' \in \mathcal{L}, L - L' \neq \emptyset\}$. For $L \in \mathcal{L}$, let $S_L = S \cap L$. Note that S_L is different for different $L \in \mathcal{L}$.

On input text $T_1, T_2, \ldots, T_n$, the learner keeps track of $S_i = S \cap \text{content}(T_i)$, based on the input seen so far (let S_i^s denote the value of S_i after having seen the input $T_1[s], T_2[s], \ldots, T_n[s]$). If and when the learner finds, for some s, that there are n pairwise distinct languages $L_1, L_2, \ldots, L_n$ in $\mathcal{L}$ such that $S_i^s = S_{L_i}$, the learner outputs conjectures for L_i on T_i. Moreover, it outputs indicator 1 for exactly those texts T_i for which L_i is not contained in any language in $\mathcal{L} - \{L_1, L_2, \ldots, L_n\}$. It is now easy to verify that the above learner (m, n)-**superTxtFin**-learns $\mathcal{L}$, as there can be at most $n - m$ languages in $\{L_1, L_2, \ldots, L_n\}$ which have a superset in $\mathcal{L} - \{L_1, L_2, \ldots, L_n\}$. ∎

Corollary 17. *Suppose $\mathcal{L}$ has at least $n+1$ languages. Then, for $0 < m \leq n$, $\mathcal{L}$ is (m, n)-**superTxtFin** learnable implies that $\mathcal{L}$ is $(m+1, n+1)$-**superTxtFin**-learnable.*

Proof. For infinite language classes $\mathcal{L}$, Theorems 8 and 9 imply that $\mathcal{L}$ is (m, n)-**superTxtFin**-learnable iff it is $(m + 1, n + 1)$-**superTxtFin**-learnable. For a finite class $\mathcal{L}$, suppose $\mathcal{L}$ is not $(m + 1, n + 1)$-**superTxtFin**-learnable. Then, by Theorem 16 there exists a $\mathcal{S} \subseteq \mathcal{L}$ of size $n+1$ which has $n-m+1$ languages that have a superset in $\mathcal{L} - \mathcal{S}$. Let L be a language in $\mathcal{S}$ which is not among these $n - m + 1$ languages. Note that there exists such a language as $m \geq 1$. Then $\mathcal{S} - \{L\}$ has $n - m + 1$ languages which have a superset in $\mathcal{L} - (\mathcal{S} - \{L\})$. This, by Theorem 16, implies that $\mathcal{L}$ is not (m, n)-**superTxtFin**-learnable. ∎

Now we are concerned with a possibility of preserving superlearnability when the number of errors $n' - m'$ is smaller than $n - m$.

Corollary 18. *Suppose $0 < m \leq n$ and $0 < m' \leq n'$, where $n - m > n' - m'$. Suppose $r \geq \max(n, n' + 1)$ (where r can be infinity). Then, there exists an automatic class $\mathcal{L}$ having r languages such that it can be (m, n)-**superTxtFin**-learnt but not (m', n')-**superTxtFin**-learnt.*

Proof. Let $\mathcal{L}$ consist of languages

$L_\epsilon = \{a^i : i \in N\}$,

$L_{a^i} = \{a^i\}$, for $1 \leq i \leq n - m$, and

$L_{b^i} = \{b^i\}$, for $n - m < i < r$.

Then, clearly $\mathcal{L}$ is (m, n)-**superTxtFin**-learnable (as only the languages L_i, $1 \leq i \leq n - m$ do not have a characteristic sample with respect to $\mathcal{L}$). However, $\mathcal{L}$ is not (m', n')-**superTxtFin**-learnable as given input texts for $L_{a^1}, L_{a^2}, \ldots, L_{a^{n'}}$, a learner must fail to **TxtFin**-learn (along with specifying them as having been learnt) at least the texts corresponding to the languages $L_{a^1}, \ldots, L_{a^{\min(n-m,n')}}$. ∎

Note that for $r = n'$, the classes of size r are easily (m', n')-**superTxtFin**-learnable. To see this, let $L_1, L_2, \ldots, L_r$ be the distinct languages in the class, $S = \{\min(L_i - L_j) : 1 \leq i, j \leq r, L_i - L_j \neq \emptyset\}$, $S_i = S \cap L_i$. Note that

S_i is different for different values of i. Then a learner just waits until it has seen $T_1[s], T_2[s], \ldots, T_r[s]$ such that there exists a permutation $j_1, j_2, \ldots, j_r$ of $1, 2, \ldots, r$ satisfying content$(T_i[s]) \cap S = S_{j_i}$ (note that the learner can easily remember the intersection of the content of each text with S). At which point, for each $i \in \{1, 2, \ldots, r\}$, the learner can output index for L_{j_i} on the text T_i. Furthermore, for classes of size smaller than n, (m, n)-**superTxtFin**-learnability is trivial. Thus, the above corollary handles all interesting cases when $n - m > n' - m'$. Now, the only remaining case where the separation problem "(m, n)-**superTxtFin** $-$ (m', n')-**superTxtFin** $= \emptyset$?" is not solved by the above results is when $n - m \leq n' - m'$, and $n > n'$. We consider this case now.

Proposition 19. *Suppose we distribute k balls in s boxes. Suppose $t \leq s$. Here we assume k, t are finite. Let $F(k, s, t) = \lfloor \frac{k}{s} \rfloor * t + \max(\{0, t - s + k - s * \lfloor \frac{k}{s} \rfloor\})$ (where, $F(k, s, t) = 0$, for $s = \infty$).*

(a) One can select t boxes such that they contain at most $F(k, s, t)$ balls.

(b) If one distributes the balls almost equally (that is each box gets $\lfloor \frac{k}{s} \rfloor$ or $\lceil \frac{k}{s} \rceil$ balls, where $\lceil \frac{k}{s} \rceil$ is taken to be 1 for $s = \infty$), then any way of selecting t boxes contains at least $F(k, s, t)$ balls.

Proof. The claim of the proposition is easy to see for $s = \infty$. So assume s is finite. For part (a), the t boxes selected would be the ones which contain the least number of balls. Thus the worst case happens when the balls are distributed nearly equally. This, is done by placing $\lfloor \frac{k}{s} \rfloor$ balls in each of the boxes, and then placing $k - s * \lfloor \frac{k}{s} \rfloor$ balls in $k - s * \lfloor \frac{k}{s} \rfloor$ boxes. Let $w = k - s * \lfloor \frac{k}{s} \rfloor$. Hence w boxes get $\lfloor \frac{k}{s} \rfloor + 1$ balls and $s - w$ boxes get $\lfloor \frac{k}{s} \rfloor$ balls. Thus, when selecting t boxes with the least number of balls, the number of balls we get equals $F(k, s, t) = t * \lfloor \frac{k}{s} \rfloor + \max(\{0, t - (s - w)\})$, which equals $t * \lfloor \frac{k}{s} \rfloor + \max(\{0, t - s + w\})$. ∎

Here, note that $F(k, s, t)$ is monotonically non-increasing in s.

The next corollary shows that (m, n)-superlearners on automatic classes may be stronger than (m', n')-superlearners if $n - m \leq n' - m'$ and $n > n'$, and gives conditions for the cases when such separation can be established.

Corollary 20. *Suppose $0 < m \leq n$ and $0 < m' \leq n'$. Suppose, further, that $n > n'$, $n - m \leq n' - m'$ and $r \geq n$. Let F be as defined in Proposition 19.*

*(a) If $n' - m' + 1 - F(n' - m' + 1, r - n', n - n') \leq n - m$, then there exists an automatic class $\mathcal{L}$ of the size r which is (m, n)-**superTxtFin**-learnable but not (m', n')-**superTxtFin**-learnable.*

*(b) If $n' - m' + 1 - F(n' - m' + 1, r - n', n - n') > n - m$, then every automatic (m, n)-**superTxtFin**-learnable class $\mathcal{L}$ of the size r is (m', n')-**superTxtFin**-learnable.*

Proof. (a) Suppose $n' - m' + 1 - F(n' - m' + 1, r - n', n - n') \leq n - m$. Note that this implies r is finite.

Let X_i, $n' < i \leq r$ be a partition of $\{m', m' + 1, \ldots, n'\}$ such that X_i's are of size either $\lfloor \frac{n' - m' + 1}{r - n'} \rfloor$ or $\lceil \frac{n' - m' + 1}{r - n'} \rceil$.

Let $L_i = \{a^i\}$, for $1 \leq i \leq n'$.

Let $L_i = \{b^i\} \cup \bigcup_{j \in X_i} L_j$, for $n' < i \leq r$.

Let $\mathcal{L} = \{L_i : 1 \leq i \leq r\}$.

Intuitively, think of languages $L_{m'}, L_{m'+1}, \ldots, L_{n'}$ as balls which are placed into the $r - n'$ boxes L_i, $n' < i \leq r$. The balls are distributed nearly equally in this analogy. The b^i are added just to make sure that these languages are different from L_i, $1 \leq i \leq n'$, in case some X_i is of size 1.

Clearly, $\mathcal{L}$ is not (m', n')-**superTxtFin**-learnable, since, when given texts for L_i, $1 \leq i \leq n'$ as input, the learner cannot learn (along with specifying that they are learnt) the texts for $L_{m'}, L_{m'+1}, \ldots, L_{n'}$.

To see that $\mathcal{L}$ is (m, n)-**superTxtFin**-learnable we proceed as follows.

Note that, if texts for any subset $\mathcal{S}$ of $\mathcal{L}$ of size n are given as input to the learner, the worst case for learning would be when languages in $\mathcal{L} - \mathcal{S}$ are supersets of as many as possible languages in $\mathcal{S}$. For this worst case, languages in $\mathcal{L} - \mathcal{S}$ would contain only languages of type L_i, $n' < i \leq r$. Thus, in this worst case, languages in $\mathcal{S}$ are all the languages L_i, $1 \leq i \leq n'$ plus $(n - n')$ of the languages among L_i, $n' < i \leq r$.

Now, no matter which $n - n'$ languages L_i, $n' < i \leq r$, are in $\mathcal{S}$, using the aforementioned balls/boxes analogy and applying Proposition 19 (b), we can conclude that these $n - n'$ languages/boxes contain at least $F(n' - m' + 1, r - n', n - n')$ languages/balls in $\mathcal{S}$. Thus, as the total number of languages/balls contained in some other languages in $\mathcal{L}$ is $n' - m' + 1$, there can be at most $n' - m' + 1 - F(n' - m' + 1, r - n', n - n')$ languages in $\mathcal{S}$ which are properly contained in some language in $\mathcal{L} - \mathcal{S}$. Therefore, by Theorem 16, $\mathcal{L}$ is (m, n)-**superTxtFin**-learnable.

(b) Suppose $n' - m' + 1 - F(n' - m' + 1, r - n', n - n') > n - m$. Suppose $\mathcal{L}$ of size r is not (m', n')-**superTxtFin**-learnable. Then, by Theorem 16, there exists a subset $\mathcal{S}$ of $\mathcal{L}$ of size n' such that at least $n' - m' + 1$ languages in $\mathcal{S}$ have a superset in $\mathcal{L} - \mathcal{S}$. Consider the languages in $\mathcal{S}$ which are contained in some language in $\mathcal{L} - \mathcal{S}$ as balls, and the languages in $\mathcal{L} - \mathcal{S}$ as boxes. But then, by Proposition 19(a) one can select $n - n'$ languages $A_1, A_2, \ldots, A_{n-n'}$ in $\mathcal{L} - \mathcal{S}$ such that they together contain at most $F(n' - m' + 1, r - n', n - n')$ languages/balls from $\mathcal{S}$. Thus, at least $n' - m' + 1 - F(n' - m' + 1, r - n', n - n') > n - m$ many languages in $\mathcal{S}$ have a superset in $\mathcal{L} - \mathcal{S} - \{A_1, A_2, \ldots, A_{n-n'}\}$. This would imply by Theorem 16 that $\mathcal{L}$ is not (m, n)-**superTxtFin**-learnable. ∎

5. (m, n)-TxtFin-learnability

Our first goal is to find a necessary condition for finite (m, n)-learnability of large automatic classes in terms of characteristic samples. For this, we introduce the concept of a *cut* and *matching* in a bipartite graph and an important Lemma 23.

Definition 21. Suppose $G = (V, E)$ is a bipartite graph, where V_1, V_2 are the two partitions of the vertices. Then,

(a) Suppose $V_1' \subseteq V_1$ and $V_2' \subseteq V_2$. (V_1', V_2') is called a *cut* of G if G does not contain any edges between $V_1 - V_1'$ and $V_2 - V_2'$. (V_1', V_2') is called a *minimum cut*, if it is a cut which minimizes $\mathrm{card}(V_1' \cup V_2')$.

(b) $E' \subseteq E$ is called a *matching* if for all pairwise distinct edges $(v, w), (v', w')$ in E', $v \neq v'$ and $w \neq w'$. E' is called a *maximum matching* if E' is a matching with maximum cardinality.

Note that cuts are usually defined using edges rather than vertices, however for our purposes it is convenient to define cut sets using vertices. We often write a bipartite graph (V, E) as (V_1, V_2, E), where V_1, V_2 are the two partitions. For example, consider the graph with vertices $V = \{a, b, c_1, c_2, \ldots, c_r, d_1, d_2, \ldots, d_k\}$ and edges $E = \{(a, d_1), (a, d_2), \ldots, (a, d_k), (c_1, b), (c_2, b), \ldots, (c_r, b)\}$. The minimum cut in the graph would be $(\{a\}, \{b\})$. Note also that $\{(a, d_1), (c_1, b)\}$ forms a maximum matching in the graph. Both minimum cut and maximum matching have same cardinality. This is not an accident, and Lemma 23 can be proven using the Max-Flow-Min-Cut Theorem (by adding a source node, with edge to each vertex in V_1, and a sink node, with edge from each vertex in V_2). For Max-Flow-Min-Cut Theorem and related concepts see, for example, [PS98]. We give a proof Lemma 23 for completeness. For this we need the following technical lemma.

Lemma 22. *Suppose G is a bipartite graph, where V_1, V_2 are two partitions of the vertices of G. Suppose $V_1' \subseteq V_1$ is a finite set such that for every subset V_1'' of V_1', the set of neighbours of V_1'' is at least* $\mathrm{card}(V_1'')$. *Then, there exists a matching of size* $\mathrm{card}(V_1')$.

Proof. We prove this by induction on size of V_1', and on the number of neighbours of V_1' (for every possible graph). If $\mathrm{card}(V_1') = 1$, then the lemma is obvious. So, for the following cases, assume $\mathrm{card}(V_1') > 1$.

Case 1: There exists $V_1'' \subset V_1'$ such that the number of neighbours of V_1'' is exactly $\mathrm{card}(V_1'')$.

In this case, clearly (a) for each subset V_1''' of V_1'', the number of neighbours of V_1''' is at least $\mathrm{card}(V_1''')$, and (b) for each subset V_1''' of $V_1' - V_1''$, there are at least $\mathrm{card}(V_1''')$ neighbours of V_1''' which are different from the neighbours of V_1''.

Thus, using induction for the two subgraphs $(V_1'', \mathit{neighbours}(V_1''), E_1)$ and $(V_1' - V_1'', V_2 - \mathit{neighbours}(V_1''), E_2)$ (where E_1, E_2 are the corresponding edges for the restricted subgraphs) we can get two disjoint matchings of size $\mathrm{card}(V_1'')$ and $\mathrm{card}(V_1' - V_1'')$ respectively, and thus we have a matching of size $\mathrm{card}(V_1')$.

Case 2: Not Case 1.

Without loss of generality, assume that every vertex in V_1' has a finite number of neighbours (otherwise, we can just keep arbitrary first $\mathrm{card}(V_1')$ neighbours of any vertex in V_1', and ignore the rest).

Now consider any vertex v of V_1'. Let w be a neighbour of v.

Case 2.1: Every subset V_1'' of V_1' has at least $\mathrm{card}(V_1'')$ neighbours in V_2, without using the edge (v, w).

In this case we are done by induction — by just deleting/ignoring the edge (v, w).

Case 2.2: Not Case 2.1.

Thus, there exists a subset V_1'' of V_1' that has less than card(V_1'') neighbours when the edge (v, w) is not considered. This implies that $v \in V_1''$ and $V_1'' - \{v\}$ has exactly card$(V_1'') - 1$ neighbours. But this violates the hypothesis of Case 2. ∎

Lemma 23. *For any bipartite graph, the size of the minimum cut is the same as the size of the maximum matching.*

Proof. Clearly, if k is the size of minimum cut, then the maximum matching can be of size at most k, as, for a cut, one needs to pick at least one vertex from each edge in the matching. Thus, we need to only show that there exists a matching of size k. For this, let (V_1', V_2') be a minimum cut. We will construct a matching of size card$(V_1' \cup V_2')$. For this, note that, for every subset V_1'' of V_1', the number of neighbours of V_1'' in $V_2 - V_2'$ is at least card(V_1'') (otherwise, we could replace V_1'' by neighbours of V_1'' in the cut, getting a cut of a smaller size). Similarly, for every subset V_2'' of V_2', the number of neighbours of V_2'' in $V_1 - V_1'$ is at least card(V_2''). By Lemma 22, this is enough to get a matching of size card$(V_1' \cup V_2')$. ∎

Lemma 24. *Fix some hypothesis space. Suppose pairwise distinct languages $X_1, X_2, \ldots, X_r$ and pairwise distinct languages $Y_1, Y_2, \ldots, Y_r$ are given (where there may be some common languages between X_i's and Y_j's) such that $X_i \subset Y_i$, for $1 \leq i \leq r$. Furthermore suppose $p_1, \ldots, p_r$ are given. Then, one can define pairwise distinct languages E_i, $1 \leq i \leq r$ such that*

(a) $X_i \subseteq E_i$

(b) p_i is not a grammar/index for E_i

(c) $E_i \in \{X_1, X_2, \ldots, X_r, Y_1, Y_2, \ldots, Y_r\}$.

Proof. Without loss of generality assume that if $i < j$, then $X_j \not\subseteq X_i$ (otherwise we can just reoder the X_i's). We define E_j by induction from $j = r$ to 1. We will also maintain languages $Y_1', Y_2', \ldots$, which change over the construction. Initially, let $Y_i' = Y_i$ for all i. We will have (by induction) the following invariants:
(i) $Y_1', \ldots, Y_j', E_{j+1}, \ldots, E_r$ are pairwise distinct and belong to $\{X_1, X_2, \ldots, X_r, Y_1, Y_2, \ldots, Y_r\}$,
(ii) $X_1, X_2, \ldots, X_j, E_{j+1}, \ldots, E_r$ are pairwise distinct,
(iii) $X_i \subset Y_i'$ for $1 \leq i \leq j$, and
(iv) p_i is not a grammar/index for E_i, for $j < i \leq r$.

Note that the above invariants imply that $Y_j' \neq X_i$, for $1 \leq i \leq j$ (as $X_i \not\supset X_j$, for $1 \leq i \leq j$).

It is easy to verify that the induction hypotheses hold when $j = r$. Suppose we have already defined $E_r, \ldots, E_{j+1}$; we then define E_j as follows.

If p_j is a grammar for X_j, then let $E_j = Y'_j$ (and other values do not change).
If p_j is not a grammar for X_j, then let $E_j = X_j$. If one of $Y'_i = X_j$, for $i < j$, then replace Y'_i by Y'_j. Other variables do not change value.

It is easy to verify that the above maintains the invariants. Now, the values of $E_1, \ldots, E_r$ satisfy the requirements of the lemma. ∎

Now we can show that the existence of characteristic samples for all the languages in the class, except at most $n-1$ ones, (where characteristic samples are relative to the class excluding the $n-1$ latter languages) is a necessary condition for $(1,n)$-**TxtFin**-learnability. Note that this characteristic sample condition is similar, but different from the one for $(1,n)$-**superTxtFin**-learning.

Theorem 25. *Suppose $\mathcal{L}$ is $(1,n)$-**TxtFin**-learnable.*
Then there exists a subset $\mathcal{S}$ of $\mathcal{L}$ of size at most $n-1$ such that every language in $\mathcal{L} - \mathcal{S}$ has a characteristic sample with respect to $\mathcal{L} - \mathcal{S}$.

Proof. Suppose $\mathbf{M}$ $(1,n)$-**TxtFin**-learns $\mathcal{L}$.
Let $\mathcal{L}' = \{L \in \mathcal{L} : (\exists$ pairwise distinct $S^L_w \in \mathcal{L}$ for each $w \in \Sigma^*)[L \cap \{x : x \leq_{ll} w\} \subseteq S^L_w]\}$.
For each L in $\mathcal{L}'$, $w \in \Sigma^*$, fix S^L_w as in the definition of $\mathcal{L}'$.
Let $\mathcal{L}'' = \{L \in \mathcal{L} - \mathcal{L}' : L$ does not have a characteristic sample with respect to $\mathcal{L} - \mathcal{L}'\}$.
Let $\mathcal{L}''' = \{A \in \mathcal{L} - \mathcal{L}' : (\exists L \in \mathcal{L} - \mathcal{L}')[L \subset A]\}$.

Claim 26. $\mathrm{card}(\mathcal{L}') < n$.

To see the claim, suppose $\mathcal{L}'$ has $\geq n$ languages. Then as input to $\mathbf{M}$, we can give texts $T_1, T_2, \ldots, T_n$ for $C_1, C_2, \ldots, C_n \in \mathcal{L}'$. Suppose $\mathbf{M}$, after seeing $T_1[m], T_2[m], \ldots, T_n[m]$, conjectures $(p_1, p_2, \ldots, p_n)$ (different from $(?, ?, \ldots, ?)$). Then consider texts T'_i extending $T_i[m]$, where T'_i is a text for $E_i = S^{C_i}_{j_i}$, for some j_i such that $S^{C_i}_{j_i} \supseteq \mathrm{content}(T_i[m])$ and (a) p_i is not a grammar for $S^{C_i}_{j_i}$ and (b) $S^{C_i}_{j_i}$ are pairwise distinct for different i. Note that this can be easily ensured. Then $\mathbf{M}$ fails on input texts being $T'_1, T'_2, \ldots, T'_n$. This completes the proof of the claim.

Suppose $\mathrm{card}(\mathcal{L}') = n - r$.

Note that every language in $\mathcal{L}''$ has a proper superset in $\mathcal{L}'''$ and every language in $\mathcal{L}'''$ has a proper subset in $\mathcal{L}''$. Consider the bipartite graph G formed by having the vertex set $V_1 = \mathcal{L}''$ and $V_2 = \mathcal{L}'''$, and edge between (L'', L''') iff $L'' \subset L'''$. (If $L \in \mathcal{L}'' \cap \mathcal{L}'''$, then for the purposes of the bipartite graph, we consider corresponding vertex in V_1 and V_2 representing L as different).

Claim 27. *There exists a cut of G of size at most $r-1$.*

Assume by way of contradiction otherwise. Then, by Lemma 23, there exists a matching of size at least r. Let this matching be $(A_1, B_1), \ldots, (A_r, B_r)$. Here, each $A_i \in \mathcal{L}''$ and each $B_i \in \mathcal{L}'''$. A_i's are pairwise distinct, B_i's are pairwise distinct, but A_i's and B_j's might coincide with each other. Assume, without loss of generality, that if $i < j$, then $A_j \not\subseteq A_i$. Now consider giving the learner input texts $T_1, T_2, \ldots, T_n$ for $A_1, A_2, \ldots, A_r, C_{r+1}, C_{r+2}, \ldots, C_n$, where $C_{r+1}, C_{r+2}, \ldots, C_n$ are pairwise distinct members of $\mathcal{L}'$. Suppose $\mathbf{M}$ outputs a conjecture $(p_1, p_2, \ldots, p_n)$ (which is different from $(?, ?, \ldots, ?)$) after seeing input $T_1[m], T_2[m], \ldots, T_n[m]$.

Then, define $E_1, \ldots, E_r$ using Lemma 24, where we take $X_1, \ldots, X_r$ to be $A_1, \ldots, A_r$ and $Y_1, \ldots, Y_r$ to be $B_1, \ldots, B_r$. Definition of $E_{r+1}, \ldots, E_n$ can be done appropriately as done in the proof of Claim 26 above when $\mathcal{L}'$ was at least n. Then, by taking T_i' to be a text extending $T_i[m]$ for E_i, we get that $\mathbf{M}$ fails to identify each T_i'. This proves the claim.

Now, it is easy to verify that taking $\mathcal{S}$ as $\mathcal{L}'$ unioned with the cut of G as in the claim, satisfies the requirements of the Theorem. ∎

Note that by appropriate modification of the above proof of Theorem 25, one can show the following more general Theorem. This can be done by first showing, as in the proof of Claim 26, that $\mathcal{L}'$ is of size less than $n - m + 1$ (say, of size $n - m + 1 - r$), where in the proof of the claim, besides the $n - m + 1$ texts for $C_1, C_2, \ldots, C_{n-m+1}$, we take texts for $m - 1$ arbitrary pairwise distinct members of $\mathcal{L} - \{C_1, C_2, \ldots, C_{n-m+1}\}$). Then, one can show, as in the proof of Claim 27, that the cut of G is of size at most $r - 1$, (where, in addition to texts for $A_1, A_2, A_r, C_{r+1}, \ldots, C_{n-m+1}$, we take texts for $m - 1$ arbitrary languages in $\mathcal{L} - \{A_1, A_2, \ldots, A_r, B_1, B_2, \ldots, B_r, C_{r+1}, \ldots, C_{n-m+1}\}$). Here note that the above construction needs the size of $\mathcal{L}$ to be at least $2n + m - 1$, as the value of r maybe upto $n - m + 1$.

Theorem 28. *Suppose $\mathcal{L}$ is (m, n)-**TxtFin**-learnable and $\mathcal{L}$ contains at least $2n - m + 1$ languages.*

Then, there exists a subset $\mathcal{S}$ of $\mathcal{L}$ of size at most $n - m$ such that every language in $\mathcal{L} - \mathcal{S}$ has a characteristic sample with respect to $\mathcal{L} - \mathcal{S}$.

Now we show that the necessary condition of the previous Theorem is sufficient for (m, n)-**TxtFin**-learning.

Theorem 29. *Suppose $\mathcal{L}$ is an automatic class. Suppose $0 < m \leq n$. Suppose there exists a subset $\mathcal{S}$ of $\mathcal{L}$ of size at most $n - m$ such that every language in $\mathcal{L} - \mathcal{S}$ has a characteristic sample with respect to $\mathcal{L} - \mathcal{S}$. Then, $\mathcal{L}$ is (m, n)-**TxtFin**-learnable.*

Proof. Let $\mathcal{L}' = \{L \in \mathcal{L} : (\exists$ infinitely many pairwise distinct $S_w^L \in \mathcal{L}$, $w \in \Sigma^*)[L \cap \{x : x \leq_{ll} w\} \subseteq S_w^L]\}$.

Note that $\mathcal{L}' \subseteq \mathcal{S}$, as the languages in $\mathcal{L}'$ cannot have a characteristic sample with respect to $\mathcal{L} - \mathcal{S}$. Thus, $\text{card}(\mathcal{L}') = n - r \leq n - m$, for some r.

Furthermore, for all $L \in \mathcal{L} - \mathcal{L}'$, there exists a finite subset X of L such that there exist at most finitely many $L' \in \mathcal{L}$ satisfying $X \subseteq L'$. Furthermore, none of the members of $\mathcal{L} - \mathcal{L}'$ are contained in any member of $\mathcal{L}'$.

So the learner $\mathbf{M}$ behaves as follows on input texts $T_1, T_2, \ldots, T_n$. It first searches for an s such that, for at least r members j of $\{1, 2, \ldots n\}$,

(a) content($T_j[s]$) is not contained in any L in $\mathcal{L}'$, and

(b) content($T_j[s]$) is contained in at most finitely many of L in $\mathcal{L}$.

Note that there exists such an s, and using Lemma 1, it can be effectively found, given an automatic indexing of $\mathcal{L}$ and indices for members of $\mathcal{L}'$. Without loss of generality, for ease of notation, from now on we assume that these r members are $\{1, 2, \ldots, r\}$.

This gives us that the corresponding r texts $T_1, T_2, \ldots, T_r$ can only be for languages from $\mathcal{L} - \mathcal{L}'$. Up to card($\mathcal{S}$) $-$ $(n - r)$ of these may be from $\mathcal{S} - \mathcal{L}'$, and thus at least $n -$ card($\mathcal{S}$) are from $\mathcal{L} - \mathcal{S}$.

Let $\mathcal{H} = \{L \in \mathcal{L} : (\exists i : 1 \leq i \leq r)[\text{content}(T_i[s]) \subseteq L]\}$. By (b) above, $\mathcal{H}$ is finite. Moreover, using Lemma 1, an index (in the automatic indexing of $\mathcal{L}$) for each of the members in $\mathcal{H}$ can be found effectively. Furthermore, subset relation among the members of $\mathcal{H}$ can be effectively determined. Arrange elements of $\mathcal{H}$ in a directed graph G, where there is an edge from L to L' iff $L \subset L'$ and no other $L'' \in \mathcal{H}$ satisfies $L \subset L'' \subset L'$. Note that the graph is acyclic. Also, note that there is no path from any $L \in \mathcal{L} - \mathcal{S}$ to another $L' \in \mathcal{L} - \mathcal{S}$ (as this would imply that $L \subset L'$, and thus no characteristic sample for L with respect to $\mathcal{L} - \mathcal{S}$ would exist).

Now let $s' > s$ be such that

(c) for each i, $1 \leq i \leq r$, there exists a (necessarily unique) $L \in \mathcal{H}$ such that content($T_i[s']$) $\subseteq L$ but content($T_i[s']$) $\not\subseteq L'$ for any other $L' \in \mathcal{H}$ which satisfies $L' \not\supseteq L$ — we assign T_i to the node L in the graph G in this case, and

(d) for each node L in G, at most one T_i is assigned to L.

Note that such s' will eventually be found as the texts $T_1, T_2, \ldots, T_r$ are for different languages from $\mathcal{H}$. Once such s' is found, the learner outputs grammar for L on T_i iff T_i is assigned to the node L. We now claim that the above learner (m, n)-**TxtFin**-learns $\mathcal{L}$. For this, it suffices to show that the learner is correct on at least m of the texts $T_1, T_2, \ldots, T_r$.

We will only count the correctness of the learner for languages in $\mathcal{L} - \mathcal{S}$.

Let G' be a graph just like G, except that the texts assigned to nodes may change — T_i is assigned to a node L iff T_i is actually a text for L. Note that each node in G and G' is assigned at most one text.

Note that if T_i is assigned to a node L in G, but to a node L' in G', then $L \subseteq L'$. Now consider the texts T_i on which the learner is wrong. These texts can be divided up into maximal chains of the form $T_{i_1}, T_{i_2}, T_{i_3}, \ldots, T_{i_{j-1}}$, where (i) T_{i_s} is assigned to A_{i_s} in G and $A_{i_{s+1}}$ (which represented content(T_{i_s})) in

G', (ii) $A_{i_s} \subset A_{i_{s+1}}$, for $1 \leq s < j$, (iii) no text is assigned to A_{i_1} in G' and no text is assigned to A_{i_j} in G, and (iv) the different maximal chains as above do not have any texts/nodes in the graph in common. Thus, we can consider such maximal chains independently for error computation: each of these chains $A_{i_1} \subset A_{i_2} \subset \ldots \subset A_{i_j}$ has at most one member from $\mathcal{L} - \mathcal{S}$ (since members of $\mathcal{L} - \mathcal{S}$ are pairwise not included in each other), and thus contain at least $j - 1$ members from $\mathcal{S} - \mathcal{L}'$. Thus, the learner fails on at most $card(\mathcal{S} - \mathcal{L}')$ texts among $T_1, T_2, \ldots, T_r$. It follows that the learner is correct on at least $r - \text{card}(\mathcal{S} - \mathcal{L}')$ many texts from $T_1, T_2, \ldots, T_r$. As the size of $\mathcal{S}$ is at most $n - m$ and $\text{card}(\mathcal{L}') = n - r$, the learner must be correct on at least m input texts. ∎

The following corollaries, solving the questions of relationships between (n, m)-**TxtFin**-learnability and (n', m')-**TxtFin**-learnability for the cases when the difference $n - m = n' - m'$ and $n - m > n' - m'$, easily follow from the above two theorems.

Corollary 30. *Suppose* $0 < m \leq n$. *Suppose* $\mathcal{L}$ *is a large enough automatic class (that is, it contains at least* $2n + m - 1$ *languages). Then,* $\mathcal{L}$ *is* (m, n)-**TxtFin**-*learnable iff it is* $(m + 1, n + 1)$-**TxtFin**-*learnable.*

Corollary 31. *Suppose* $0 < m < n$. *There exists an automatic class* $\mathcal{L}$ *that is* (m, n)-**superTxtFin**-*learnable, but not* $(m, n - 1)$-**TxtFin**-*learnable.*

Proof. Let $\mathcal{L}$ consist of the languages
$L_{a^i} = \{a^i\}$, for $i > 0$;
$L_{b^i} = \{a^i, b^i\}$, for $1 \leq i \leq n - m$.

Let $\mathcal{S} = \{L_{a^i} : 1 \leq i \leq n - m\}$. Then every language in $\mathcal{L} - \mathcal{S}$ has a characteristic sample with respect to $\mathcal{L}$. Thus, it follows from Theorem 8 that $\mathcal{L}$ is (m, n)-**superTxtFin**-learnable. However, by Theorem 28, $\mathcal{L}$ is not $(m, n-1)$-**TxtFin**-learnable. To see this, note that for any $\mathcal{S}'$, for every language in $\mathcal{L} - \mathcal{S}'$ to have a characteristic sample with respect to $\mathcal{L} - \mathcal{S}'$, $\mathcal{S}'$ must contain at least one of L_{a^i} and L_{b^i}, for each i with $1 \leq i \leq n - m$, and, thus, cannot be of the size at most $n - 1 - m$. ∎

The next corollary shows that the (m, n)-**superTxtFin**-learnability is weaker than (m, n)-**TxtFin**-learnability, even if the **TxtFin**-learner is allowed to make just one error.

Corollary 32. *There exists a class* $\mathcal{L}$ *which is* $(n - 1, n)$-**TxtFin**-*learnable for all* $n \geq 2$, *but not* (m, n)-**superTxtFin**-*learnable for any* m, n *with* $0 < m \leq n$.

Proof. Let $\Sigma = \{a\}$, and consider $L_\epsilon = \Sigma^*$, $L_{a^{i+1}} = \{a^i\}$ and $\mathcal{L} = \{L_{a^i} : i \in N\}$. Then, by Theorem 29, for all $n \geq 2$, $\mathcal{L} \in (n - 1, n)$-**TxtFin** as each $L_{a^{i+1}}$ has the characteristic sample $\{a^i\}$ with respect to $\mathcal{L} - \{L_\epsilon\}$. On the other hand, as none of $L_{a^{i+1}}$ have a characteristic sample with respect to $\mathcal{L}$, $\mathcal{L}$ is not (m, n)-**superTxtFin**-learnable, by Theorem 9. ∎

Note that Theorem 28 and thus Corollary 30 needed the classes to be large enough (of size at least $2n - m + 1$). We now address the case when the classes might be small.

The following theorem suggests certain characterization of small (m, n)-**TxtFin**-learnable classes.

Theorem 33. *Suppose $0 < m < n$. Consider any finite automatic class $\mathcal{L}$ of cardinality at least n. Then, $\mathcal{L}$ is (m, n)-**TxtFin**-learnable iff for every $\mathcal{S} \subseteq \mathcal{L}$ of size n, for every $r \geq n - m + 1$ and for every subset $\mathcal{S}' = \{X_1, X_2, \ldots, X_r\}$ of $\mathcal{S}$ of size r, there does not exist a subset $\mathcal{S}'' = \{Y_1, Y_2, \ldots, Y_r\}$ of $((\mathcal{L} - \mathcal{S}) \cup \mathcal{S}')$, of size r such that $X_i \subset Y_i$ for $1 \leq i \leq r$.*

Proof. Suppose there exist $\mathcal{S}$ of size n, $r \geq n - m + 1$, $\mathcal{S}' = \{X_1, X_2, \ldots, X_r\} \subseteq \mathcal{S}$ and $\mathcal{S}'' = \{Y_1, Y_2, \ldots, Y_r\} \subseteq (\mathcal{L} - \mathcal{S}) \cup \mathcal{S}'$ such that $X_i \subset Y_i$, where X_i's are pairwise different and Y_i's are pairwise different. Then, we show that $\mathcal{L}$ cannot be (m, n)-**TxtFin**-learnt. Suppose any learner $\mathbf{M}$ is given. Suppose we give texts $T_1, T_2, \ldots, T_n$ for $X_1, X_2, \ldots, X_r, Z_{r+1}, \ldots, Z_n$ to $\mathbf{M}$, where $Z_{r+1}, \ldots, Z_n$ are the pairwise different members of $\mathcal{S} - \mathcal{S}'$. Suppose the learner outputs conjectures $p_1, p_2, \ldots, p_n$ on the above texts after seeing $T_1[s], T_2[s], \ldots, T_n[s]$.

Without loss of generality assume that $L_j \not\subseteq L_i$, for $1 \leq i < j \leq r$. Let $E_1, E_2, \ldots, E_r$ be as given by Lemma 24.

Now, for $1 \leq i \leq r$, let T_i' be a text for E_i such that T_i' extends $T_i[s]$. Then, $\mathbf{M}$ fails to learn the texts $T_1', T_2', \ldots, T_r'$ when given $(T_1', T_2', \ldots, T_r', T_{r+1}, \ldots, T_n)$ as input. Thus, $\mathcal{L}$ cannot be (m, n)-**TxtFin**-learnt.

On the other hand, suppose for every $\mathcal{S} \subseteq \mathcal{L}$ of size n, for every $r \geq n - m + 1$ and for every subset $\mathcal{S}' = \{X_1, X_2, \ldots, X_r\}$ of $\mathcal{S}$ of size r, there does not exist a subset $\mathcal{S}'' = \{Y_1, Y_2, \ldots, Y_r\}$ of $((\mathcal{L} - \mathcal{S}) \cup \mathcal{S}')$, of size r such that $X_i \subset Y_i$ for $1 \leq i \leq r$. Then, we claim that $\mathcal{L}$ is (m, n)-**TxtFin**-learnable. For this consider the following learner which (m, n)-**TxtFin**-learns $\mathcal{L}$.

Let $S = \{\min(L - L') : L, L' \in \mathcal{L}, L - L' \neq \emptyset\}$. For $L \in \mathcal{L}$, let $S_L = S \cap L$.

On input texts $T_1, T_2, \ldots, T_n$, the learner keeps track of $S_i = S \cap \text{content}(T_i)$, based on the input seen so far (let S_i^s denote the value of S_i after having seen input $T_1[s], T_2[s], \ldots, T_n[s]$)). If and when the learner finds, for some s, that there are n pairwise distinct languages $L_1, L_2, \ldots, L_n$ in $\mathcal{L}$ such that $S_i^s = S_{L_i}$, the learner outputs conjectures for L_i on T_i. Now, for $\mathbf{M}$ to make an error in learning T_i, T_i should be a text for some superset of L_i. Suppose $\mathcal{S} = \{L_1, L_2, \ldots, L_n\}$, and $\mathcal{S}'$ is the subset of $\mathcal{S}$ on which the above learner made errors. Then, if $\mathcal{S}' = \{X_1, X_2, \ldots, X_r\}$ is of size $r > n - m$, then there must exist some subset $\mathcal{S}'' = \{Y_1, Y_2, \ldots, Y_r\}$ of $((\mathcal{L} - \mathcal{S}) \cup \mathcal{S}')$, of size r, with $X_i \subset Y_i$ for $1 \leq i \leq r$. However, by the hypothesis, such an $\mathcal{S}'$ does not exist. Thus, the learner (m, n)-**TxtFin**-learns $\mathcal{L}$. ∎

Here note that the hypothesis in the statement of Theorem 33, implies that each X_i is properly in some Y_j which belongs to $\mathcal{L} - \mathcal{S}$. This can be seen by considering any maximal subset chain among X_i's: if $X_{i_1} \subset X_{i_2} \subset \ldots \subset X_{i_r}$ is a maximal subset chain, then all these languages are contained in Y_{i_r} which is in $\mathcal{L} - \mathcal{S}$. In other words, we can consider $X_1, X_2, \ldots, X_r$ to be divided into maximal subset chains (with no common language between different chains), where each chain is properly contained in some member $Y \in \mathcal{L} - \mathcal{S}$ (where the Y's are pairwise different for different maximal chains).

The following corollaries give relationships between (m, n)-**TxtFin**-learners for different parameters m and n when the learnable classes are small.

Corollary 34. *Suppose $\mathcal{L}$ is a finite automatic class that has at least $n + 1$ languages. Then, for $0 < m \leq n$, $\mathcal{L}$ is (m, n)-**TxtFin** learnable implies that $\mathcal{L}$ is $(m + 1, n + 1)$-**TxtFin**-learnable.*

Proof. If $\mathcal{L}$ is not $(m+1, n+1)$-**TxtFin**-learnable, then by Theorem 33 there exist $\mathcal{S} \subseteq \mathcal{L}$ of size $\geq n + 1$, $\mathcal{S}' = \{X_1, X_2, \ldots\} \subseteq \mathcal{S}$ of size at least $n - m + 1$ and $\mathcal{S}'' = \{Y_1, Y_2, \ldots\} \subseteq (\mathcal{L} - \mathcal{S}) \cup \mathcal{S}'$ of the same size as $\mathcal{S}'$, where $X_i \subset Y_i$, for each i. Then, if $\mathcal{S} \neq \mathcal{S}'$, by removing one member from $\mathcal{S}$ not belonging to $\mathcal{S}'$, and using $\mathcal{S}', \mathcal{S}''$ as witnesses in Theorem 33, we get that $\mathcal{L}$ is not (m, n)-**TxtFin**-learnable. In the case $\mathcal{S} = \mathcal{S}'$, we could just remove some minimal member (subset wise) from $\mathcal{S}, \mathcal{S}'$ to get the result. ∎

Corollary 35. *Suppose $0 < m \leq n$ and $0 < m' \leq n'$, where $n - m > n' - m'$. Suppose $r \geq \max(n, n' + 1)$ (where r can be infinity). Then there exists an automatic class $\mathcal{L}$ having r languages such that $\mathcal{L}$ can be (m, n)-**superTxtFin**-learnt, but not (m', n')-**TxtFin**-learnt.*

Proof. Let $\mathcal{L}$ consist of the languages

$L_\epsilon = \{a^i : i \in N\}$,

$L_{a^i} = \{a^j : 1 \leq j \leq i\}$, for $1 \leq i \leq n - m$, and

$L_{b^i} = \{b^i\}$, for $n - m < i < r$.

Then, clearly $\mathcal{L}$ is (m, n)-**superTxtFin**-learnable (as only the languages L_i, $1 \leq i \leq n - m$ do not have a characteristic sample with respect to $\mathcal{L}$). However, $\mathcal{L}$ is not (m', n')-**TxtFin**-learnable as for $\mathcal{S} = \{L_{a^1}, L_{a^2}, \ldots, L_{a^{\min(n-m,n')}}\}$, one can choose $X_i = L_{a^i}$, for $1 \leq i \leq \min(n - m, n')$, $Y_i = X_{i+1}$, for $1 \leq i < \min(n-m, n')$, and $Y_{\min(n-m,n')} = L_\epsilon$. Then, by Theorem 33, $\mathcal{L}$ is not (m', n')-**TxtFin**-learnable. ∎

As we already noted, for $r = n'$, the classes of size r are easily seen to be (m', n')-**superTxtFin**-learnable, and for classes of size smaller than n, (m, n)-**superTxtFin**-learnability is also trivial. Obviously, the same applies to (m, n)-**TxtFin**-learnability. Thus, the above corollary handles all interesting cases when $n - m > n' - m'$. Now, the only remaining case where the separation problem "(m, n)-$\mathbf{TxtFin} - (m'n')$-$\mathbf{TxtFin} = \emptyset$?" is not solved by above results is when $n - m \leq n' - m'$, and $n > n'$. We consider this case now.

Corollary 36. *Suppose $0 < m \leq n$, $0 < m' \leq n'$. Suppose, further, that $n > n'$, $n - m \leq n' - m'$ and $r \geq n$. Let F be as defined in Proposition 19.*

*(a) If $n' - m' + 1 - F(n' - m' + 1, r - n', n - n') \leq n - m$, then there exists an automatic class of size r which is (m, n)-**superTxtFin**-learnable but not (m', n')-**TxtFin**-learnable.*

*(b) If $n'-m'+1-F(n'-m'+1, r-n', n-n') > n-m$, then every automatic (m,n)-**TxtFin**-learnable class $\mathcal{L}$ of size r is (m',n')-**TxtFin**-learnable.*

Proof. (a) Suppose $n'-m'+1-F(n'-m'+1, r-n', n-n') \leq n-m$. Note that this implies r is finite. Let Z_i, $n' < i \leq r$ be a partition of $\{m', m'+1, \ldots, n'\}$ such that Z_i's are of size either $\lfloor \frac{n'-m'+1}{r-n'} \rfloor$ or $\lceil \frac{n'-m'+1}{r-n'} \rceil$.
For $1 \leq i < m'$, let $L_i = \{a^i\}$.
For $m' \leq i \leq n'$, let $L_i = \{a^j : j \leq i$ and i, j belong to the same partition in the above partitioning of $\{m', m'+1, \ldots, n'\}\}$.
For $n' < i \leq r$, let $L_i = \{b^i\} \cup \bigcup_{j \in Z_i} L_j$, Note that $L_i = \{b^i\} \cup \{a^j : j \in Z_i\}$, for $n' < i \leq r$.

Intuitively, think of L_i, $i \in \{m', m'+1, \ldots, n'\}$ as balls and L_i, $n' < i \leq r$ as boxes. The languages with indices in Z_i form a subset chain with L_i, $n' < i \leq r$, containing all the languages L_j, $j \in Z_i$.

Let $\mathcal{L} = \{L_i : 1 \leq i \leq r\}$.

Now, $\mathcal{L}$ is not (m',n')-**TxtFin**-learnable. To see this, choose $\mathcal{S}$ to be $\{L_1, L_2, \ldots, L_{n'}\}$, $\mathcal{S}'$ to be $\{L_{m'}, L_{m'+1}, \ldots, L_{n'}\}$, and X_k, Y_k, for $m' \leq k \leq n'$ as follows:

$X_k = L_k$, for $m' \leq k \leq n'$;

$Y_k = L_i$, if k is the maximum element in the partition Z_i;

$Y_k = L_j$, if $k \in Z_i$, and j is the least element in Z_i which is larger than k.

Then, by Theorem 33 it follows that $\mathcal{L} \notin (m',n')$-**TxtFin**.

To see that $\mathcal{L}$ is (m,n)-**superTxtFin**-learnable we proceed as follows. Suppose $\mathcal{S}$ is a set of n languages whose texts are given to the learner as input. The worst case for the learner would be when there are as many languages as possible, $X_1, X_2, \ldots, X_p$, such that for some $Y_1, Y_2, \ldots, Y_p$ in $\mathcal{L} - \mathcal{S} \cup \{X_1, X_2, \ldots, X_p\}$, $X_q \subset Y_q$. We could think of these X_i/Y_i as forming disjoint subset chains, where the last member of each subset chain is from $\mathcal{L} - \mathcal{S}$ and the other members are from $X_1, X_2, \ldots, X_p$. Thus, we can think of this as balls being placed in boxes based on the analogy above.

In the discussed worst case, for a subset $\mathcal{S}$ of $\mathcal{L}$ of size n, languages in $\mathcal{S}$ are all the languages $L_i, 1 \leq i \leq n'$ plus $n-n'$ of the languages among $L_i, n' < i \leq r$.

Now, using the above balls/boxes analogy, with $n-n'$ boxes in $\mathcal{S}$, we can apply Proposition 19(b). Thus, as the total number of balls is $n'-m'+1$, there can be at most $n'-m'+1-F(n'-m'+1, r-n', n-n')$ languages in $\mathcal{S}$ which are contained in some language in $\mathcal{L} - \mathcal{S}$. Thus, by Theorem 16, $\mathcal{L}$ is (m,n)-**superTxtFin**-learnable.

(b) Suppose $n'-m'+1-F(n'-m'+1, r-n', n-n') > n-m$. Suppose $\mathcal{L}$ of size r is not (m',n')-**TxtFin**-learnable. Then, by Theorem 33, there exists a subset $\mathcal{S}$ of $\mathcal{L}$ of size n' such that at least $n'-m'+1$ languages in $\mathcal{S}$ can be arranged in pairwise disjoint maximal subset chains such that each chain has a distinct superset in $\mathcal{L} - \mathcal{S}$. We can think of the members of these subset chains as balls and the corresponding supersets in $\mathcal{L} - \mathcal{S}$ as boxes. But then, by Proposition 19(a) one can select $n-n'$ languages $A_1, A_2, \ldots, A_{n-n'}$ in $\mathcal{L} - \mathcal{S}$ such that there are pairwise distinct languages in $\mathcal{L} - \mathcal{S} - \{A_1, A_2, \ldots, A_{n-n'}\}$ which are supersets of pairwise distinct subset chains containing in total at least

$n' - m' + 1 - F(n' - m' + 1, r - n', n - n') > n - m$ many languages in $\mathcal{S}$ (where the different chains have pairwise different supersets). This would imply by, Theorem 33, that $\mathcal{L}$ is not (m, n)-**TxtFin**-learnable. ∎

6. (m, n)-TxtEx-learning

In this section we consider (m, n)-learning of automatic classes in the limit.

Note that any finite automatic class of languages can be easily **TxtEx**-learnt and thus (m, n)-**TxtEx**-learnt. Thus, for the following we will assume that the classes under consideration are infinite. For the rest of this section, we assume that the hypothesis spaces are always automatic.

We will give a characterization of (m, n)-**TxtEx**-learning in terms of existence of tell-tale sets. We also show that the number of languages learnable in parallel can be increased or decreased when $n - m$, the number of languages which may be erroneously identified, remains the same. It will also follow from our characterization that, for (m, n)-learnability in the limit, superlearnability and learnability have the same power.

Definition 37. (Based on [BB75, Ful90]) $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ (each of same length) is called a *stabilizing sequence* for $\mathbf{M}$ on $(L_1, L_2, \ldots, L_n)$ iff

(a) for each i, content$(\sigma_i) \subseteq L_i$.

(b) for all $\tau_1, \tau_2, \ldots, \tau_n$ of same length such that $\sigma_i \subseteq \tau_i$ and content$(\tau_i) \subseteq L_i$, and hypotheses output by $\mathbf{M}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ is same as the corresonding hypotheses output by $\mathbf{M}(\tau_1, \tau_2, \ldots, \tau_n)$.

Suppose $\mathbf{M}(\sigma_1, \sigma_2, \ldots, \sigma_n)\downarrow_{hyp} = (g_1, g_2, \ldots, g_n)$. If, additionally, for at least m different $i \in \{1, 2, \ldots, n\}$, g_i is an index for L_i (in the hypothesis space used by $\mathbf{M}$), then $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ is called a *locking sequence* for $\mathbf{M}$ on $(L_1, L_2, \ldots, L_n)$.

Lemma 38. *(Based on [BB75]) If $\mathbf{M}$ (m, n)-**TxtEx**-identifies $\mathcal{L}$, and $L_1, L_2, \ldots, L_n$ are pairwise distinct members of $\mathcal{L}$, then there exists a (m, n)-**TxtEx**-locking sequence for $\mathbf{M}$ on each $(L_1, L_2, \ldots, L_n)$.*

The following two theorems represent a generalization of the result of [JLS12] that an automatic class $\mathcal{L}$ is **TxtEx**-learnable iff every language in $\mathcal{L}$ has a tell-tale set with respect to $\mathcal{L}$ (see also Angluin [Ang80]).

Theorem 39. *Suppose $0 < m \leq n$. Suppose $\mathcal{L}$ is an automatic class. If $\mathcal{L}$ has at most $n - m$ languages which do not have a tell-tale set with respect to $\mathcal{L}$ then $\mathcal{L}$ is (m, n)-**superTxtEx**-learnable.*

Proof. Suppose $\mathcal{L} = \{L_\alpha : \alpha \in I\}$, where I is a regular index set and the alphabet set is Σ. For $\alpha \in I$, if L_α has a tell-tale set with respect to $\mathcal{L}$, then let b_α be length-lexicographically least element in Σ^* such that $L_\alpha \cap \{x : x \leq_{ll} b_\alpha\}$ is a tell-tale set for L_α; otherwise b_α is a special symbol #. Note that, by Lemma 1, such b_α can be obtained effectively from α for automatic families.

Now the learner $\mathbf{M}$ on input texts $T_1, T_2, \ldots, T_n$ behaves as follows. It searches for an r, a subset X of $\{1, 2, \ldots, n\}$ of cardinality m, and α_i for each $i \in X$ such that

(a) $b_{\alpha_i} \neq \#$ and

(b) for each $i \in X$, content$(T_i) \subseteq L_{\alpha_i}$ and content$(T_i[r])$ contains $L_{\alpha_i} \cap \{x : x \leq_{ll} b_{\alpha_i}\}$.

In this case $\mathbf{M}$ outputs (in the limit) α_i on T_i and indicates it as being learnt. Conjectures on other texts are irrelevant and they are specified as not being learnt.

Note that if the texts $T_1, T_2, \ldots, T_n$ are indeed for n pairwise distinct languages in $\mathcal{L}$, then r, X and α_i, $i \in X$ as above will exist and can be found in the limit. Thus, the above learner will succeed to (m, n)-**superTxtEx**-learn $\mathcal{L}$. ∎

Theorem 40. *Suppose $0 < m \leq n$. Suppose $\mathcal{L}$ is an automatic class. If $\mathcal{L}$ has at least $n - m + 1$ languages which do not have a tell-tale set with respect to $\mathcal{L}$, then $\mathcal{L}$ is not (m, n)-***TxtEx***-learnable.*

Proof. Suppose, by way of contradiction, that $\mathbf{M}$ (m, n)-**TxtEx**-learns the class $\mathcal{L}$. Let $L_1, L_2, \ldots, L_{n-m+1}$ be $n - m + 1$ languages in $\mathcal{L}$ which do not have a tell-tale set with respect to $\mathcal{L}$. Let $L_{n-m+2}, \ldots, L_n$ be pairwise distinct languages in $\mathcal{L} - \{L_1, L_2, \ldots, L_{n-m+1}\}$.

As $\mathbf{M}$ (m, n)-**TxtEx**-learns $\mathcal{L}$, there must be a stabilizing sequence for $\mathbf{M}$ on $L_1, L_2, \ldots, L_n$. Suppose the stabilizing sequence was $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ (each should be of same length). Let $g_1, g_2, \ldots, g_n$ be the grammar output by $\mathbf{M}$ on input $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ Then, for $1 \leq i \leq n - m + 1$, let T_i be texts extending σ_i for pairwise distinct languages L'_i in $\mathcal{L} - \{L_1, L_2, \ldots, L_{n-m+1}, \ldots, L_n\}$ such that g_i is not a grammar for L'_i. Note that there exist such pairwise distinct languages in $\mathcal{L}$ as $L_1, L_2, \ldots, L_{n-m+1}$ do not have a tell-tale set with respect to $\mathcal{L}$. Thus, we have that $\mathbf{M}$ fails on all of T_i, $1 \leq i \leq n - m + 1$. Thus, $\mathbf{M}$ does not (m, n)-**TxtEx**-learn $\mathcal{L}$. ∎

The following corollaries follow from the above Theorems 39 and 40.

Corollary 41. *Suppose $0 < m \leq n$. If $\mathcal{L}$ is an automatic class which is (m, n)-***TxtEx***-learnable then $\mathcal{L}$ is (m, n)-***superTxtEx***-learnable.*

Corollary 42. *Suppose $0 < m \leq n$. Suppose $\mathcal{L}$ is an automatic class. Then, $\mathcal{L}$ is (m, n)-***TxtEx***-learnable iff it is $(m + 1, n + 1)$-***TxtEx***-learnable.*

Corollary 43. *Suppose $0 < n$. (n, n)-***TxtEx*** $= (1, 1)$-***TxtEx*** $=$ ***TxtEx***.

Theorem 44. *Suppose $0 < n$. There exists an automatic class $\mathcal{L}$ which can be $(1, n + 1)$-***superTxtEx***-learnt but not $(1, n)$-***TxtEx***-learnt.*

Proof. Suppose $x, z \in \Sigma^*$. Let $L_{x,\epsilon} = \{conv(x, y) : y \in \Sigma^*\}$. Let $L_{x,z} = \{conv(x, y) : y \leq_{ll} z\}$, for $z \neq \epsilon$.

Let $S \subseteq \Sigma^*$ be a set of cardinality n. Let $\mathcal{L} = \{L_{x,z} : z \in \Sigma^*, x \in S\}$. Then, it is easy to verify that $\mathcal{L}$ can be $(1, n+1)$-**TxtEx**-learnt. To see this, note that for $n+1$ texts $T_1, T_2, \ldots, T_{n+1}$, for different languages from $\mathcal{L}$, at least one text must be for $L_{x,z}$, for some $x \in S$, $z \in \Sigma^* - \{\epsilon\}$. Then the **TxtEx**-learner can algorithmically search for one such x, z and in the limit identify that text (along with specifying that text as having been identified) and output arbitrary fixed conjectures on the other texts in the limit.

Using an argument from Gold [Gol67] we show that $\mathcal{L}$ cannot be $(1, n)$-**TxtEx**-learnt by a learner **M**: Otherwise, there will be some locking sequence $(\sigma_{x_1}, \sigma_{x_2}, \ldots, \sigma_{x_n})$, for **M** on languages $(L_{x_1,\epsilon}, L_{x_2,\epsilon}, \ldots, L_{x_n,\epsilon})$ (here $x_1, x_2, \ldots, x_n$ are the different members of S). Suppose the grammars to which **M** converges on $(\sigma_{x_1}, \ldots, \sigma_{x_n})$ is $g_{x_1}, \ldots, g_{x_n}$. Then, let w be length-lexicographically largest element such that $conv(x, w)$ appears in σ_x, for some x. Let $w' \geq_{ll} w$ be such that none of the grammars g_{x_i} are for $L_{x_i,w'}$. Then the learner fails if the input texts are for $L_{x_i,w'}$, which extend σ_{x_i}. ∎

Corollary 45. *For $0 < m \leq n$, there exists an automatic class $\mathcal{L}$ such that $\mathcal{L}$ can be $(m, n+1)$-**TxtEx**-learnt but not (m, n)-**TxtEx**-learnt.*

Theorem 46. *Suppose $\mathcal{L}$ is an automatic class and $\mathcal{S}$ is a finite subset of $\mathcal{L}$ such that for all $L \in \mathcal{L} - \mathcal{S}$, L has a characteristic sample with respect to $\mathcal{L} - \mathcal{S}$. Then $\mathcal{L}$ is **TxtEx**-learnable.*

Proof. We show that each $L \in \mathcal{S}$ has a tell-tale set with respect to $\mathcal{L}$. This would imply that every language in $\mathcal{L}$ has a tell-tale set with respect to $\mathcal{L}$ and thus, by [JLS12], $\mathcal{L}$ is **TxtEx**-learnable.

To see this, consider any $L \in \mathcal{S}$. Let A be a finite subset of L such that for all $L' \in \mathcal{S} - \{L\}$, $A \subseteq L'$ implies $L \subseteq L'$. Note that there exists such an A as $\mathcal{S}$ is finite. If there is no $L'' \in \mathcal{L} - \mathcal{S}$ such that $L'' \subset L$, then clearly A is a tell-tale set for L with respect to $\mathcal{L}$. So suppose there exists an $L'' \in \mathcal{L} - \mathcal{S}$ such that $L'' \subseteq L$. Let A'' be characteristic subset of L'' with respect to $\mathcal{L} - \mathcal{S}$. Let x be a member of $L - L''$, if any (otherwise, x can be an arbitrary member of L). Then $A \cup A'' \cup \{x\}$ is a tell-tale set for L with respect to $\mathcal{L}$, as no language in $\mathcal{L} - \mathcal{S} - \{L''\}$, by the definition of characteristic sample, contains A''. ∎

Corollary 47. *Suppose $\mathcal{L}$ is an automatic class and $\mathcal{S}$ is a finite subset of $\mathcal{L}$ such that for all $L \in \mathcal{L} - \mathcal{S}$, L has a characteristic sample with respect to $\mathcal{L} - \mathcal{S}$. Suppose $0 < n$. Then $\mathcal{L}$ is (n, n)-**superTxtEx**-learnable.*

Corollary 48. $(1, n)$-**TxtFin** $\subseteq$ **TxtEx**.

7. Automatic (m, n)-Finite Learning

In this section, we consider finite (m, n)-learning by finite automata. For finite automatic classes, (m, n)-**superAutoTxtFin** and (m, n)-**AutoTxtFin** are same as (m,n)-**superTxtFin** and (m,n)-**TxtFin**, respectively. This follows from the proofs of Theorem 16 and 33, as the positive side in those results can be easily implemented using automatic learners.

Theorem 49. *Suppose $\mathcal{L}$ is an automatic finite class. Then,*
*(a) $\mathcal{L}$ is (m,n)-**TxtFin**-learnable implies $\mathcal{L}$ is (m,n)-**AutoTxtFin**-learnable.*
*(b) $\mathcal{L}$ is (m,n)-**superTxtFin**-learnable implies $\mathcal{L}$ is (m,n)-**superAutoTxtFin**-learnable.*

Note that, for automatic learnability, the characterization results (in terms of characteristic samples) of Sections 4 and 5 do not hold. This is illustrated by the following result, based on techniques from [JLS12].

Theorem 50. *[JLS12] Let $\Sigma = \{a, b\}$. $\mathcal{L} = \{L : (\exists n)(\exists x \in \Sigma^n)[L = \Sigma^n - \{x\}]\}$. Then $\mathcal{L}$ is not **AutoTxtEx**-learnable.*

Note that every language in the class $\mathcal{L}$ above has a characteristic sample (the language itself) with respect to $\mathcal{L}$. The proof for the above theorem can be generalized to show that $\mathcal{L}$ is not $(1,k)$-**AutoTxtEx**-learnable. To see this, suppose, by way of contradiction, that $\mathbf{M}$ $(1,k)$-**AutoTxtEx**-learns $\mathcal{L}$. Let n be large enough. Partition Σ^n into $k+1$ groups $S_1, S_2, \ldots, S_{k+1}$ of roughly equal size (their sizes being either $\lfloor \frac{2^n}{k} \rfloor$ or $\lceil \frac{2^n}{k} \rceil$). Now, we claim that there exist σ_i^j, τ_i^j, for $1 \leq i \leq k$, $1 \leq j \leq k$ such that

(a) $\sigma_i^j = \tau_i^j = \#^n$ for $i \neq j$;

(b) $|\sigma_i^i| = |\tau_i^i| = n$ and $\mathrm{content}(\sigma_i^i) \neq \mathrm{content}(\tau_i^i)$ and each contain n elements from S_i;

(c) Let $\gamma_i^j = \sigma_i^1 \sigma_i^2 \ldots \sigma_i^j$, for $1 \leq i, j \leq n$. Then, $\mathbf{M}(\gamma_1^i, \gamma_2^i, \ldots, \gamma_{i-1}^i, \gamma_i^{i-1}\sigma_i^i, \gamma_{i+1}^i, \ldots, \gamma_k^i) = \mathbf{M}(\gamma_1^i, \gamma_2^i, \ldots, \gamma_{i-1}^i, \gamma_i^{i-1}\tau_i^i, \gamma_{i+1}^i, \ldots, \gamma_k^i)$, for $1 \leq i \leq k$.

Note that there exist such σ_i^i, τ_i^i as needed in parts (b), (c). The claim is true, as the memory of the automatic learner after having seen such σ_i^i, τ_i^i can be of at most $O(n)$ bits (as an automatic learner, after seeing $O(n)$ data items of length $\leq n$, can have memory of the length at most $O(n)$) though the number of possibilities for such σ_i^i/τ_i^i is approximately $\binom{2^n/k}{n}$. Now, let x_i, y_i be such that $x_i \in \mathrm{content}(\sigma_i^i) - \mathrm{content}(\tau_i^i)$ and $y_i \in \mathrm{content}(\tau_i^i) - \mathrm{content}(\sigma_i^i)$. Let T_i be a text for $\Sigma^n - \{x_i, y_i\}$. Then $\mathbf{M}$ can be made to fail by considering $\mathbf{M}$'s behaviour on the k texts: $T_i' = \gamma_i^k T_i$, where its limiting behaviour does not change if one replaces, for some i's, σ_i^i by τ_i^i in T_i'. Thus, we can make $\mathbf{M}$ fail on each of the k-inputs T_i' by appropriately replacing or not replacing σ_i^i by τ_i^i.

In the sequel, without loss of generality, assume that all languages have at most one grammar in the hypothesis space (which is automatic). So below equality of languages is equivalent to grammars being the same.

Our next goal is to show that, for large enough automatic classes, automatic finite $(m+1, n+1)$-learnability implies automatic finite (m,n)-learnability. We begin with a technical lemma.

Lemma 51. *Given $k, n > 0$, there exists a number $r_{k,n}$ such that the following holds.*

Suppose we are given conjectures $(p_1^i, p_2^i, \ldots, p_n^i)$, for $1 \le i \le r_{k,n}$.

Then, for some $S \subseteq \{1, 2, \ldots, r_{k,n}\}$ of size at least k, and for some $p_1, p_2, \ldots, p_n$, we have, for all $j \in \{1, 2, \ldots, n\}$, either

(a) for all $i \in S$, $p_j^i = p_j$, or

(b) for all distinct $i, i' \in S$, $p_j^i \ne p_j^{i'}$,

where p_j's can be computed automatically from the different p_r^i's. (Note that p_j for j satisfying (b) above can be arbitrary).

Proof. We will define $r_{k,n}$ by induction on n.

Case 1: $n = 1$.

For this, we can take $r_{k,1} = (k-1)^2 + 1$.

If $\{p_1^i : 1 \le i \le r_{k,1}\}$ has at least k elements, then (b) can be satisfied. Otherwise, by pigeonhole principle, there exists a p_1 such that for k different i's, $p_1^i = p_1$.

Case 2: $n > 1$.

For this, we can take $r_{k,n} = (r_{k,n-1} - 1)^2 + 1$.

If $\{p_1^i : 1 \le i \le r_{k,n}\}$ has at least $r_{k,n-1}$ elements, then we can satisfy the lemma by induction. Otherwise, by pigeonhole principle, there exists a p_1 such that for $r_{k,n-1}$ different i's in $\{1, 2, \ldots, r_{k,n}\}$, $p_1^i = p_1$. Then, again, we can satisfy the lemma by induction. ∎

Theorem 52. *Let $0 < m \le n$. Suppose $\mathcal{L}$ is an infinite automatic class which is $(m+1, n+1)$-**AutoTxtFin**-learnable. Then $\mathcal{L}$ is (m, n)-**AutoTxtFin**-learnable.*

Proof. Suppose $\mathbf{M}$ $(m+1, n+1)$-**AutoTxtFin**-learns $\mathcal{L}$. Consider a learner $\mathbf{M}'$ (which will (m, n)-**AutoTxtFin**-learn $\mathcal{L}$) defined as follows. Suppose input texts for $\mathbf{M}'$ are $T_1, T_2, \ldots, T_n$, for pairwise different languages in $\mathcal{L}$. Let $k = n(n+2)$. Let $X_1, X_2, \ldots, X_{r_{k,n}+n}$, be $r_{k,n} + n$ pairwise different languages from $\mathcal{L}$. Let T_{n+i} be a text for X_i. (Note: We need some automaticity here. So we assume $T_{n+i}(t) = x_t$ if $x_t \in X_i$; otherwise $T_{n+i}(t) = \#$, where $x_0, x_1, \ldots$ is a length-lexicographic ordering of elements of Σ^*. Furthermore, we assume that the learner, after having seen $T_{n+i}[t+1]$, remembers x_t, in addition to any other items, in its memory so that it can compute x_{t+1}.)

Run $\mathbf{M}$ on $(T_1, T_2, \ldots, T_n, T_{n+i})$, for $1 \le i \le r_{k,n} + n$. Note that, for at most n of these values of i, hypothesis of $\mathbf{M}$ may converge to $(?, ?, \ldots, ?)$ (since one of the texts $T_1, T_2, \ldots, T_n$ may be a text for an X_i). So consider the first $r_{k,n}$ of the i's in $\{1, 2, \ldots, r_{k,n} + n\}$ on which the learner $\mathbf{M}$ outputs a hypothesis different from $(?, ?, \ldots, ?)$. Without loss of generality, assume that these i's are $1, 2, \ldots, r_{k,n}$. Let the corresponding grammars/indices being output on the texts T_j, $1 \le j \le n$, be p_j^i. Now using Lemma 51, we can automatically obtain grammar/index p_j, $1 \le j \le n$ satisfying (a) or (b) in the statement of Lemma 51, for some $S \subseteq \{i : 1 \le i \le r_{k,n}\}$ of size k. The (final) output of $\mathbf{M}'$ on $T_1, T_2, \ldots, T_n$, will then be $(p_1, p_2, \ldots, p_n)$ (before $\mathbf{M}'$ determines these p_i's, its output will be $(?, ?, \ldots, ?)$). Thus, there exists an $i \in S$ such that:

(1) for at least m many $j \in \{1, 2, \ldots, n\}$, p^i_j is a grammar/index for content(T_j),

(2) for none of the j satisfying (b) in the statement of Lemma 51, p^i_j is a grammar/index for content(T_j).

The above holds, as there are at most n many different i's in S which do not satisfy (1), and there are at most n many i's in S, for which some j satisfying (b) in the statement of Lemma 51 also satisfies that p^i_j is a grammar/index for T_j.

Fix an i satisfying (1) and (2) above. It follows that there are at least m different $j \in \{1, 2, \ldots, n\}$, for which the p^i_j's are grammar/index for content(T_j). As these j's all satisfy (a) in the statement of Lemma 51, we have $p^i_j = p_j$. Hence, $\mathbf{M}'$ (m, n)-**AutoTxtFin**-learns $\mathcal{L}$. ∎

Remark 53. *Suppose $0 < m \leq n$, $\mathcal{L}$ is finite and contains at least $2n + 2 - m$ languages, and $\mathcal{L}$ is $(m + 1, n + 1)$-**AutoTxtFin**-learnable. Then, $\mathcal{L}$ is (m, n)-**AutoTxtFin**-learnable. This holds, as by Theorem 28 and Theorem 29, $\mathcal{L}$ is (m, n)-**TxtFin**-learnable and thus by Theorem 49, $\mathcal{L}$ is (m, n)-**AutoTxtFin**-learnable.*

We have not been able to prove that (m, n)-**AutoTxtFin**-learnability implies $(m + 1, n + 1)$-**AutoTxtFin**-learnability (for infinite automatic classes). Yet, we can show that (m, n)-**AutoTxtFin**-learnability does not imply $(m, n - 1)$-**TxtFin**-learnability.

Proposition 54. *Suppose $r \geq 1$. Let $L_{a^i} = \{a^i\}$, Let $L_{b^i} = \{b^i\}$, and $L_{c^i} = \{b^i, c^i\}$.*

Let $\mathcal{L} = \{L_{a^i} : i \geq 1\} \cup \{L_{b^i} : 1 \leq i \leq r\} \cup \{L_{c^i} : 1 \leq i \leq r\}$.

*Then, for $m \geq 1$, $\mathcal{L}$ is $(m, m + r)$-**superAutoTxtFin**-learnable, but not $(m, m + r - 1)$-**TxtFin**-learnable.*

Proof. An automatic learner waits until it sees that at least m of the input texts contain either a string in a^+ or c^+. Then, it can easily identify these input texts (assuming they are for languages in $\mathcal{L}$).

$\mathcal{L} \notin (m, m + r - 1)$-**TxtFin** follows from Theorem 28. ∎

Corollary 55. *For all m, n such that $0 < m \leq n - 1$, there exists an automatic family which can be (m, n)-**AutoTxtFin**-learnt but not $(m, n - 1)$-**AutoTxtFin**-learnt.*

As the following two theorems show, for **superAutoTxtFin**-learning of large enough classes, (m, n)-learnability implies $(m + 1, n + 1)$-learnability, and vice versa.

Theorem 56. *Suppose $0 < m \leq n$. Suppose $\mathcal{L}$ is an automatic $(m + 1, n + 1)$-**superAutoTxtFin**-learnable class containing at least $2n - m + 1$ languages. Then, $\mathcal{L}$ is (m, n)-**superAutoTxtFin**-learnable.*

Proof. If $\mathcal{L}$ is finite, then this follows from Corollary 12 and Theorem 49. If $\mathcal{L}$ is infinite, then let $L_1, L_2, \ldots, L_r$ be r distinct languages in $\mathcal{L}$, where $r = n^2+n+1$. Let T_{n+i} be an automatic text for L_i[2].

Let $\mathbf{M}$ be an automatic learner which $(m+1, n+1)$-**superAutoTxtFin**-learns $\mathcal{L}$. Now consider a learner $\mathbf{M}'$ which on input texts $T_1, T_2, \ldots, T_n$ simulates $\mathbf{M}$ on input $T_1, T_2, \ldots, T_n, T_k$, for $n+1 \leq k \leq n+r$, memorizing the conjectures made, if any, for each of these k until it finds a subset X of $\{1, 2, \ldots, n\}$ of size m along with p_j for each $j \in X$ such that

(a) for each $j \in X$, for at least $n+1$ different values of $k \in \{n+1, n+2, \ldots, n+r\}$, $\mathbf{M}$'s conjecture on input $T_1, T_2, \ldots, T_n, T_k$ is p_j on T_j along with T_j being specified as having been learnt.

If and when such X (and corresponding p_j for $j \in X$) is found, $\mathbf{M}'$ outputs conjectures p_j on T_j for $j \in X$, along with specifying it as being learnt. Conjectures of $\mathbf{M}'$ on the remaining texts are irrelevant (with them being specified as not having been learnt). Note that an automatic learner can memorize all the conjectures as above, and can test for existence of appropriate X and p_j as above.

Note that if such X and corresponding p_j are found, then clearly p_j must be a correct grammar for content(T_j), as at most n of the $k \in \{n+1, n+2, \ldots, n+r\}$ can be spoiled due to T_k being a text for content(T_i), for some $i \in \{1, 2, \ldots, n\}$. Thus, it suffices to argue that there will exist such X and corresponding p_j for $j \in X$. For this, we argue as follows.

(b) There can be at most n values of $k \in \{n+1, n+2, \ldots, n+r\}$) such that one of T_i, $1 \leq i \leq n$ is a text for L_k (thus spoiling the corresponding output conjectures, if any, of $\mathbf{M}$ on input $T_1, T_2, \ldots, T_n, T_k$); Let Z be the set of values of such $k \in \{n+1, n+2, \ldots, n+r\}$.

(c) Let X' be the set of numbers $j \in \{1, 2, \ldots, n\}$ such that T_j is specified as being learnt by $\mathbf{M}$ for at most n values of $k \in \{n+1, n+2, \ldots, n+r\} - Z$.

(d) Let $Z' = \{k \in \{n+1, n+2, \ldots, n+r\} - Z :$ for some $j \in X'$, T_j was specified as being learnt by $\mathbf{M}$ on input $T_1, T_2, \ldots, T_n, T_k\}$. Note that cardinality of Z' is at most n^2.

(e) Then, for any $k \in \{n+1, n+2, \ldots, n+r\} - Z - Z'$, the texts T_j which are specified as being learnt by $\mathbf{M}$ on $T_1, T_2, \ldots, T_n, T_k$ must have been specified as learnt by $\mathbf{M}$ on input $T_1, T_2, \ldots, T_n, T_k$ for at least $n+1$ different $k \in \{n+1, n+2, \ldots, n+r\}$ each time with the same conjecture (say q_j) of $\mathbf{M}$ on T_j.

As $r > n + n^2$, it follows that there exists X and corresponding $p_j = q_j$ as needed in the construction above (since there exists only one grammar for each language in the indexing used for hypothesis space). ∎

[2] For example, we can have $T_{n+i}(s) = w_s$, if $w_s \in L_i$ and # otherwise, where w_s is the s-th string in length-lexicographic order. To obtain such a text, the learner remembers in its memory w_s after having seen s inputs from the texts. This memory is updated to w_{s+1} after the next element is seen.

Theorem 57. *Suppose* $0 < m \leq n$.
Suppose $\mathcal{L}$ *is an automatic class which is* (m,n)-**superAutoTxtFin**-*learnable. Then* $\mathcal{L}$ *is* $(m+1,n+1)$-**superAutoTxtFin**-*learnable.*

Proof. Suppose $\mathbf{M}$ (m,n)-**superAutoTxtFin**-learns $\mathcal{L}$. Then consider the following $\mathbf{M}'$, which on input text $T_1, T_2, \ldots, T_{n+1}$ simulates $\mathbf{M}$ on $T_1, T_2, \ldots, T_{i-1}$, $T_{i+1}, \ldots, T_{n+1}$, for each possible value of $i \in \{1, 2, \ldots, n+1\}$. Note that each of these simulations must eventually output conjectures different from $(?, ?, \ldots, ?)$. Then $\mathbf{M}'$ outputs p_j on T_j (and specifies it as having been learnt) if in one of these simulations $\mathbf{M}$ outputs p_j on T_j and specifies it as having been learnt. Conjectures of $\mathbf{M}'$ on the remaining texts are irrelevant and they are not being identified. Suppose $S \subseteq \{1, 2, \ldots, n+1\}$ is the set of j's such that T_j was not specified by $\mathbf{M}$ as having been learnt on input $T_1, T_2, \ldots, T_{i-1}, T_{i+1}, \ldots, T_{n+1}$ for any value $i \in \{1, 2, \ldots, n+1\}$. We claim that S is of cardinality at most $n-m$: otherwise, taking $i \in \{1, 2, \ldots, n+1\} - S$, we have that $\mathbf{M}$ failed to specify at least m texts as being identified on input $T_1, T_2, \ldots, T_{i-1}, T_{i+1}, \ldots, T_{n+1}$.

Note that the learner $\mathbf{M}'$ can be made automatic.

It follows that $\mathbf{M}'$ $(m+1,n+1)$-**superAutoTxtFin**-learns $\mathcal{L}$. ∎

Corollary 58. *Suppose* $0 < m \leq n$. *Suppose* $\mathcal{L}$ *is an automatic class containing at least* $2n - m + 1$ *languages. Then,* $\mathcal{L}$ *is* (m,n)-**superAutoTxtFin**-*learnable iff it is* $(m+1,n+1)$-**superAutoTxtFin**-*learnable.*

The proof of Corollary 31 also shows:

Corollary 59. *Suppose* $0 < m < n$. *There exists an automatic class* $\mathcal{L}$ *that is* (m,n)-**superAutoTxtFin**-*learnable, but not* $(m,n-1)$-**TxtFin**-*learnable.*

8. Automatic (m,n)-TxtEx-learning

We do not have a good characterization of (m,n)-**AutoTxtEx**-learnability. Our main result in this section is that, for this type of learning, $(m+1,n+1)$-learnability implies (m,n)-learnability (same is true for superlearners). We also show that, for superlearners, (m,n)-learnability implies $(m+1,n+1)$-learnability.

But first we show that automatic learners in the limit from one text can sometimes learn more than general finite $(1,n)$-learners.

Theorem 60. *Suppose* $0 < n$. *Then,* **AutoTxtEx** $-$ $(1,n)$-**TxtFin** $\neq \emptyset$.

Proof. Let $\Sigma = \{a\}$ and $L_{a^i} = \{a^j : j \leq i\}$. Let $\mathcal{L} = \{L_{a^i} : i \in N\}$. Then, it is easy to see that $\mathcal{L}$ is an automatic family which can be **AutoTxtEx**-learnt. However, $\mathcal{L} \notin (1,n)$-**TxtFin** by Theorem 25, as no language in $\mathcal{L}$ has a characteristic sample with respect to $\mathcal{L} - \mathcal{S}$ for any finite subset $\mathcal{S}$ of $\mathcal{L}$. ∎

Theorem 61. *Suppose $0 < m \leq n$ and $\mathcal{L}$ is an automatic class.*

*(a) Suppose $\mathcal{L}$ is $(m+1, n+1)$-**AutoTxtEx**-learnable. Then, $\mathcal{L}$ is (m, n)-**AutoTxtEx**-learnable.*

*(b) Suppose $\mathcal{L}$ is $(m+1, n+1)$-**superAutoTxtEx**-learnable. Then, $\mathcal{L}$ is (m, n)-**superAutoTxtEx**-learnable.*

Proof. We only show part (a). Part (b) can be done similarly.

If $\mathcal{L}$ is finite, then clearly $\mathcal{L}$ is in **AutoTxtEx** and thus (m, n)-**AutoTxtEx** and (m, n)-**superAutoTxtEx**-learnable. So assume that $\mathcal{L}$ is infinite. Let $L_1, L_2, \ldots, L_{n+1}$ be $n+1$ pairwise distinct languages in $\mathcal{L}$. Let $S = \{\min(L_i - L_j) : 1 \leq i, j \leq n+1, L_i - L_j \neq \emptyset\}$. Note that $L_i \cap S$ is different for different i, $1 \leq i \leq n+1$.

Let T_{n+i} be a text for L_i such that $T_{n+i}(s) = w_s$, if $w_s \in L_i$; $T_{n+i}(s) = \#$ otherwise. Here $w_0, w_1, \ldots,$ is a length lexicographic ordering of Σ^*.

Suppose $\mathbf{M}$ $(m+1, n+1)$-**TxtEx**-learns $\mathcal{L}$. Now we define $\mathbf{M}'$ as follows.

On any input texts $(T_1, T_2, \ldots, T_n)$ $\mathbf{M}'$ simulates $\mathbf{M}$ on input $(T_1, T_2, \ldots, T_n, T_{n+i})$, for $1 \leq i \leq n+1$, remembering, after having seen $(T_1[s], T_2[s], \ldots, T_n[s])$ the memory of $\mathbf{M}$ on $(T_1[s], T_2[s], \ldots, T_n[s], T_{n+i}[s])$, for each $i \in \{1, 2, \ldots, n+1\}$. Additionally, for $i \in \{1, 2, \ldots, n\}$, $\mathbf{M}'$ memorizes $C_i^s = \text{content}(T_i[s]) \cap S$.

Now suppose the conjectures of $\mathbf{M}$ on input $(T_1[s], T_2[s], \ldots, T_n[s], T_{n+i}[s])$ are $(g_{1,i}^s, g_{2,i}^s, \ldots, g_{n,i}^s, g_{n+i,i}^s)$. Then the conjectures of $\mathbf{M}'$ after seeing input $T_1[s], T_2[s], \ldots, T_n[s]$ are $(g_{1,j}^s, g_{2,j}^s, \ldots, g_{n,j}^s)$ where j is the least member of $\{1, 2, \ldots, n\}$ such that none of C_i^s, $i \in \{1, 2, \ldots, n\}$, equals $L_j \cap S$.

Now suppose $T_1, T_2, \ldots, T_n$ are texts for pairwise distinct languages in $\mathcal{L}$. Then, in the limit j as computed above will be such that L_j is not equal to content(T_i) for all $i \in \{1, 2, \ldots, n\}$. Thus, $(T_1, T_2, \ldots, T_n, T_{n+j})$ will be texts for pairwise distinct languages in $\mathcal{L}$, and thus $\mathbf{M}'$ will correctly identify at least m of the texts $T_1, T_2, \ldots, T_n$ (as $\mathbf{M}$ identifies at least $m+1$ of the $n+1$ texts $T_1, T_2, \ldots, T_n, T_{n+j}$). ∎

Theorem 62. *Suppose $0 < n$.*

*(n, n)-**superAutoTxtEx** = (n, n)-**AutoTxtEx** = **AutoTxtEx**.*

Proof. **AutoTxtEx** $\subseteq$ (n, n)-**superAutoTxtEx** $\subseteq$ (n, n)-**AutoTxtEx** follows by definition. On the other hand, by Theorem 61, (n, n)-**AutoTxtEx** $\subseteq$ $(1, 1)$-**AutoTxtEx** = **AutoTxtEx**. ∎

A proof similar to the one for Theorem 57 can show the following.

Theorem 63. *Suppose $0 < m \leq n$. Suppose $\mathcal{L}$ is automatic and is (m, n)-**superAutoTxtEx**-learnable. Then, $\mathcal{L}$ is $(m+1, n+1)$-**superAutoTxtEx**-learnable.*

It is open at present whether (m, n)-**AutoTxtEx**-learnability implies $(m+1, n+1)$-**AutoTxtEx**-learnability.

9. Conclusion

We defined and explored a model of parallel learning of n languages at a time when at least m languages are required to be learnt correctly. Similarly to (m, n)-computation being a deterministic alternative to probabilistic computation based on randomization, our model suggests a deterministic alternative to traditional probablistic learnability of languages (explored, for example, in [Pit89] and [WFK84]; as L. Pitt showed in [Pit89], learning using traditional probability is strongly related to another type of parallel deterministic learning — learning a language by a team). It turns out that, for the finite (m, n)-learnability, the maximum number $n - m$ of languages in the automatic family that do not have characteristic samples, and, for (m, n)-learnability in the limit, the maximum number of $n - m$ of languages in the automatic family that do not have tell-tale sets are the crucial factors defining learnability (and not the frequency m out of n of correct conjectures — as it follows from our results, increasing frequency not necessarily diminishes learnability of families of languages). Since a family of languages with a larger number of languages without characteristic samples (or tell-tale sets) is more topologically complex, the number $n-m$ can be interpreted as a measure of this complexity, and we have shown that there are learnability hierarchies based on this complexity measure.

Several interesting problems remain open. The main problem is finding characterizations, if any, for (m, n)-**AutoTxtFin**-learnability. It is open at present whether (m, n)-**AutoTxtFin**-learnability implies $(m+1, n+1)$-**AutoTxtFin**-learnability. Another potentially interesting area of research would be finding if and how frequency learnability can help in terms of efficiency of learning.

Acknowledgements

We thank the referees of ALT 2014 for several helpful comments which improved the presentation of the paper.

References

[ADHP05] H. Austinat, V. Diekert, U. Hertrampf, and H. Petersen. Regular frequency computations. *Theoretical Computer Science*, 330:15–20, 2005.

[AGS89] D. Angluin, W. Gasarch, and C. Smith. Training sequences. *Theoretical Computer Science*, 66:255–272, 1989.

[Ang80] D. Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.

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

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

[BG00] A. Blumensath and E. Grädel. Automatic structures. In *15th Annual IEEE Symposium on Logic in Computer Science (LICS)*, pages 51–62. IEEE Computer Society, 2000.

[BGK96] R. Biegel, W. Gasarch, and E. Kinber. Frequency computation and bounded queries. *Theoretical Computer Science*, 163:177–192, 1996.

[BKF11] K. Balodis, I. Kucevalovs, and R. Freivalds. Frequency prediction of functions. In *MEMICS*, pages 76–83, 2011.

[Gol67] E. M. Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.

[JLS12] S. Jain, Q. Luo, and F. Stephan. Learnability of automatic classes. *Journal of Computer and System Sciences*, 78(6):1910–1927, 2012.

[JORS99] S. Jain, D. Osherson, J. Royer, and A. Sharma. *Systems that Learn: An Introduction to Learning Theory.* MIT Press, Cambridge, Mass., second edition, 1999.

[Kin76] E. Kinber. Frequency computations in finite automata. *Kibernetika*, 2:7–15, 1976. In Russian. English translation in Cybernetics, vol. 12, pp 179–187.

[KN95] B. Khoussainov and A. Nerode. Automatic presentations of structures. In *Logical and Computational Complexity, (International Workshop LCC 1994)*, volume 960 of *Lecture Notes in Computer Science*, pages 367–392. Springer, 1995.

[KSVW95] E. Kinber, C. Smith, M. Velauthapillai, and R. Wiehagen. On learning multiple concepts in parallel. *Journal of Computer and System Sciences*, 50:41–52, 1995.

[LZ92] S. Lange and T. Zeugmann. Types of monotonic language learning and their characterization. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 377–390. ACM Press, 1992.

[LZ93] S. Lange and T. Zeugmann. Language learning in dependence on the space of hypotheses. In *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, pages 127–136. ACM Press, 1993.

[Muk92] Y. Mukouchi. Characterization of finite identification. In K. Jantke, editor, *Analogical and Inductive Inference, Proceedings of the Third International Workshop*, pages 260–267, 1992.

[PS98] C. H. S. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Dover, 1998.

[Pit89] L. Pitt. Probabilistic inductive inference. *Journal of the ACM*, 36:383–433, 1989.

[Ros60] G. Rose. An extended notion of computability. In *Abstracts of International Congress for Logic, Methodology and Philosophy of Science*, page 14, 1960.

[Smi82] C. Smith. The power of pluralism for automatic program synthesis. *Journal of the ACM*, 29:1144–1165, 1982.

[Tra64] B. Trakhtenbrot. On the frequency of computation of functions. *Algebra i Logika*, 2:25–32, 1964.

[WFK84] R. Wiehagen, R. Freivalds, and E. Kinber. On the power of probabilistic strategies in inductive inference. *Theoretical Computer Science*, 28:111–133, 1984.