Proceedings of the International MultiConference of Engineers and Computer Scientists 2011 Vol I,
IMECS 2011, March 16 - 18, 2011, Hong Kong

Advances in Learning Formal Languages

Gend Lal Prajapati

***Abstract*—we present an overview in the advances related to the learning of formal languages i.e. development in the grammatical inference research. The problem of learning correct grammars for the unknown languages is known as grammatical inference. It is considered a main subject of inductive inference, and grammars are important representations to be investigated in machine learning from both theoretical and practical points of view. Application area of grammatical inference is increasing day by day, and it is still required to find a task where grammatical inference models have done much better than other machine learning or pattern recognition programs. However, it is known that making research in this area is a computationally hard problem. This paper mainly explores the area, its applications, various learning paradigms, the case of context-free grammars, challenges, recent trends etc., and cites the important literature on these.**

***Index Terms*— machine learning, grammatical inference, learning model, formal language, context-free grammars**

I. Introduction

DURING the last forty years there has been a growing interest in investigating formal learning models which can learn formally specified classes of grammars from particular kinds of inputs. The study of formal language learning models is at the very heart of computer science. In recent years a number of significant contributions in the field of learning formal grammars have been made, unfortunately inductive learning of context-free grammars is found as computationally hard problem. These results have kindled considerable interest in the study of learning models and the area of *machine learning* of context-free grammars has blossomed into a field of intense interest.

As a broad subfield of *artificial intelligence*, machine learning is concerned with the design and development of learning models and techniques that allow computers to *learn*. At a general level, there are two types of learning: *inductive*, and *deductive*. Inductive machine learning methods extract rules and patterns out of massive data sets, whereas deductive reasoning applies general *principles* to reach specific conclusions.

Machine learning has a wide spectrum of applications including natural language processing, syntactic pattern recognition, search engines, medical diagnosis, bioinformatics, detecting credit card fraud, stock market analysis, classifying DNA sequences, speech and handwriting recognition, object recognition in computer vision, game playing and robot locomotion. Moreover, within the machine learning research field *grammatical inference* has been investigated, more or less independently.

In the area of artificial intelligence, one classical problem is to obtain the rules or patterns that model the structure of a set. When the objects are represented by elements of a formal language, the patterns determine the rules of a grammar which are able to generate the language. The problem of searching these rules belongs to the area of grammatical inference (or *grammar induction*). Thus the general problem of learning formal grammars is an inductive inference problem where target domain is a formal language and the representation class is a family of grammars. The learning objective is to determine a correct grammar for the (unknown) target language, given finite set of examples of the language.

Increasing number of practical applications based on grammatical inference ideas or techniques enable the investigation of new and better grammatical inference models. Hence, it would be useful to know the advancement and important literature in this area.

II. Major Learning Models

There are many somewhat arbitrary choices of learning model. The research activities on grammatical inference have been stimulated by the three major established formal models for learning from examples or inductive inference proposed within computational learning theory framework: Gold's model of *identification in the limit* [38], Angluin's *query learning model* [10], and the *probably approximately correct learning model* by Valiant, addressed *PAC learning model*, in short [101]. The main concentration of these models is on the computational efficiency of the learning algorithm. Each model provides a learning protocol and a criterion for the success of learning.

A. The Model of Identification in the Limit

An early study in the field of Inductive inference can be seen in [93]. A first convincing model for the case of grammatical inference is identification in the limit introduced by Gold in 1967 [38]. The setting of this model is that of on-line, incremental learning. It views learning as an infinite process and provides a learning model where an infinite sequence of examples of the unknown grammar is presented to the inference algorithm (learner) and the limiting behavior is used as the criterion of its success. After each new example the learner must return some hypothesis (guess). Identification is achieved when the learner returns a correct guess and does not change its mind afterwards, then learner is said to identify the unknown grammar in the limit for the target language.

Manuscript received August 05, 2010; revised September 27, 2010.

G. L. Prajapati, Professor in Computer Science & Engineering, is with the Swami Vivekanand College of Engineering, Indore-452020 INDIA, on leave from the Institute of Engineering & Technology, Devi Ahilya University, Indore-452001 INDIA (phone: 91-731-3296144; fax: 91-7324-405200; e-mail: glprajapati1@gmail.com).

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

B. *The Query Learning Model*

Angluin considers a learning situation in which a teacher is available to answer specific kind of queries on the unknown grammar of the target language and presents an elegant formulation of such a teacher and learner paradigm [10]. In the query learning model, there is a fixed set of oracles, known as Minimal Adequate Teacher (MAT) that can answer specific kinds of queries made by the inference algorithm on the unknown grammar. For example, the following types of queries are typical:

1) *Membership.* The input is a string and the output is "yes" if the string belongs to the target language and "no" otherwise.
2) *Equivalence*. The input is a grammar and the output is "yes" if the hypothesis is equivalent to the target, and "no" otherwise. If the answer is "no", a string called counter-example is returned.

In this set-up, an inference algorithm runs with oracles for queries for the unknown grammar, and eventually halts and outputs a correct grammar in a certain finite time.

This is no longer a limiting criterion of learning. A membership query returns one bit of information, but it often plays a vital role in efficient exact identification. For example, the class of deterministic finite automata (DFA) can be identified in polynomial time using equivalence queries and membership queries while it cannot efficiently be identified from equivalence queries only [9], [11].

C. *The PAC Learning Model*

Exact learning has always been considered a hard to achieve goal. Valiant gives the distribution independent probabilistic approach of learning from random examples. This formulation is popularly known as PAC learning model [101], which has also been studied in the context of grammar learning ([65], with also a discussion about these issues and special interest to DFA learning).

III. Learning Finite Automata

Learning regular grammars or DFA has been strongly considered in the area of grammar induction. Efficient algorithms have been presented dealing with learning such grammars in a variety of learning paradigms such as when both examples and counter-examples are provided or when the learning algorithm is allowed to question some teacher.

The possible reasons explaining that most attention has been focused on this class of grammars are that this problem may seem simple enough as it is less general as compared with the next levels of the Chomsky hierarchy, and because of the availability of learning methodologies. As an example, Angluin has investigated the active learning paradigm (*i.e.*, learning with queries asked to an oracle) [7] where efficient DFA learning is possible. The positive results are also seen in the learning from polynomial time and data. But the positive results do not seem to hold for learning context-free grammars in any setting.

The study of the learnability of DFA is a good mean for studying a number of interesting approaches in inductive inference and grammatical inference [68]. In this section, we will see some important results and useful techniques related to DFA learning. A complete information that include early work, learning from good examples, and a survey related with DFA learning can be found in [99], [105], and [68].

A *deterministic finite (state) automaton* (a DFA) is a 5-tuple $A = (Q, \Sigma, q_0, \delta, F)$, where Q is a finite set of *states,* Σ is an alphabet of input symbols, δ is the *state transition function* δ: $Q \times \Sigma \rightarrow Q$, $q_0 \in Q$ is the *initial state,* and $F \subseteq Q$ is a set of *final states.* The language accepted by a DFA A is denoted by $L(A)$.

A. *Learning from Representative Samples*

For learning an unknown (hidden) DFA $A = (Q, \Sigma, q_0, \delta, F)$ from examples, a useful information about A is a *representative sample* S of A, that is, a finite subset of $L(A)$ that exercises every live transition in A. It has been shown that the class of DFA can be learned in polynomial time from a representative sample and using membership queries [7].

B. *Learning with Teachers*

A learning protocol which is based on MAT is considered in [9]. This teacher can answer membership queries and equivalence queries about the unknown DFA A made by an inference algorithm. Angluin has shown that equivalence queries compensate for the lack of representative samples, and presented an efficient inference algorithm for learning DFA using equivalence and membership queries [9]. The positive result is that the class of DFA can be learned in polynomial time using equivalence queries and membership queries.

Yokomori presents an efficient learning of *nondeterministic finite automata* (NFA) in polynomial time from MAT [107]. This provides an alternative algorithm for learning regular languages in polynomial time from MAT.

C. *Learning from Positive Data*

Learning formal languages from *positive presentation* is one interesting and important subject in the field of grammatical inference, and also on the Gold's model of identification in the limit. A positive presentation of a unknown DFA A is any infinite sequence of examples such that the sequence contains all and only the strings in the language $L(A)$. Gold demonstrates that there is a fundamental difference in what could be learned from positive versus complete presentations, and shows a negative result on identification in the limit from positive presentation of superfinite class of languages, *e.g.*, context-free languages [38]. Since the class of regular languages is also superfinite, it is useful to restrict DFA somehow to subclasses in order to obtain learnability results from positive presentation. Angluin takes this path [8].

In order to correct identification in the limit from positive data, we must avoid the problem of "overgeneralization", which means guessing a language that is a strict superset of the unknown language. Angluin presents a series of subclasses of DFA, called k-reversible automata for k = 0, 1, 2, . . ., and shows that the existence

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

of *characteristic samples* is sufficient (to avoid the problem of overgeneralization) for identification from positive presentation for k-reversible automata and there exist such characteristic samples for the class of k-reversible automata [8].

A characteristic sample of a k-reversible automaton A is a finite sample $S \subset L(A)$ such that $L(A)$ is the "smallest" k-reversible language that contains S. If we can determine a characteristic sample for the unknown language among the input examples, then we are confirmed that a guess of the target language will not be an overgeneralization. Therefore, it seems to be true that any characteristic sample is a representative sample for k-reversible automata.

The main conclusion of Angluin's research in the paper [8] is: the class of k-reversible automata, for $k = 0, 1, 2, \ldots$, can be learned in the limit from positive presentation.

Some more results on identification from positive presentation which may not be directly related with DFA can be seen in [6], [52], [92], and [67].

D. Hardness Results

Even we have several positive results for learning DFA, there are also several computationally hardness results related to learnability of DFA. Gold proves that the problem of finding a DFA with a minimum number of states consistent with a given finite sample of positive and negative examples is NP-hard [39]. Moreover, even a very simple case of grammatical inference, learning DFA from positive and negative examples, is intractable. Further, a stronger result on the minimum consistent DFA is illustrated in [69].

Angluin has made significant contributions on learning regular languages or DFA. She also shows negative results for efficient learning of various classes of grammars from equivalence queries only [11], and develops the useful technique of *approximate fingerprints* to obtain negative results for learning from equivalence queries only. Using this technique, she shows that there is no polynomially bounded algorithm using only equivalence queries that learns the class of DFA, NFA, context-free grammars, or the disjunctive or conjunctive normal form boolean formulas.

E. Learning from Erroneous Examples

It may be the case when information provided to the learner is incomplete, *e.g.*, some of the membership queries may be answered by "I don't know." It is natural to assume that the examples may contain some noise, or incomplete information on membership queries. For the standard noise-free model proper exact identification is required. There are fewer tasks to do in the Valiant's probabilistic PAC learning model to deal with the presence of noise. Sakakibara illustrates a model, where each membership query is erroneously answered independently at random, and one can defeat the noise by querying points many times and taking a majority vote until the confidence in the correct answer is high enough [81]. In the related field, Ron and Rubinfeld also consider a model of *persistent* noise in membership queries [78].

IV. The Case of Context-Free Grammars

Context-free grammars are more expressive as compared with regular grammars. In the previous sections, we have seen that there has been huge research into the problem of learning DFA from examples. Here we see whether there are analogous results for learning context-free grammars.

Angluin has shown that there is no polynomial time algorithm using only equivalence queries that exactly identifies context-free grammars [11]. Furthermore, Angluin and Kharitonov have shown that the problem of learning context-free grammars from membership and equivalence queries is computationally as hard as the cryptographic problems, or boolean formulas problems for which there is currently no known polynomial time algorithm [12].

Learning the entire class of the context-free grammars seems to be intractable whichever learning model we choose. The important question as to whether the class can be learned with a polynomial number of queries is still an open question, but widely believed to be also intractable [13]. One hurdle is that of determinism, and the other that of linearity, motivating early studies for the class of linear languages [18]. But in the paradigm of learning from polynomial time and data this class is still not learnable in the limit [41].

Despite above negative results, we will see in the following sections several positive results for learning context-free grammars with additional information or learning subclasses of context-free grammars efficiently.

A. Learning from Structural Information

We discuss a learning problem for context-free grammars where, besides given examples, some additional information is available for the learning algorithm. A useful (and may be reasonable) information would be information on the grammatical structure of the unknown context-free grammar, *i.e.*, example presentations in the form of strings with grammatical structure. Levy and Joshi have already suggested the possibility of efficient grammatical inferences in terms of strings with grammatical structure [57]. Sakakibara has shown in [80] by extending Angluin's inference algorithm [9] for DFA to tree automata that the class of context-free grammars can be learned in polynomial time using structural membership queries and structural equivalence queries.

Theorem 1. ([80]). *The class of context-free grammars can be learned in polynomial time with the help of structural equivalence queries and structural membership queries.*

A structural membership query is a membership query for a structured string to ask whether it is generated by the unknown context-free grammar G, and a structural equivalence query returns "yes" if a queried context-free grammar G' is structurally equivalent to the unknown context-free grammar G and returns "no" with a counter-example otherwise. The counter-example is a structured string in the symmetric difference of the set of unlabelled derivation trees (structured strings) of G and the set of unlabelled derivation trees (structured strings) of G'.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Since the class of context-free grammars is superfinite, it cannot be identified in the limit because of negative result in [38] from positive presentation. Sakakibara has shown a class of context-free grammars, called reversible context-free grammars, which can be identified in the limit from positive presentations of structured strings [82].

Theorem 2. ([82])**.** *The class of reversible context-free grammars can be identified in the limit from positive presentation of structured strings (structural examples) provided that the structured strings are generated with respect to a reversible context-free grammar for the hidden (unknown) context-free language.*

Since the learning algorithm for reversible context-free grammars is an extension of Angluin's learning algorithm which learns zero-reversible automata ([8]), the algorithm learns in time polynomial in the size of the input examples. It is to be noticed that the above result does not imply that the whole class of context-free grammars can be learned from positive presentation of structured strings.

The work in the related paradigm is to learning context-free grammars from positive presentation of structured strings is Crespi-Reghizzi's, where he has described a constructive method for learning a subclass of context-free grammars, which is a different class from reversible context-free grammars [25]. His class of context-free grammars defines only a subclass of context-free languages, called *noncounting context-free languages.* A subclass of reversible context-free grammars, called *type invertible grammars,* is also investigated in [60].

B. Typical Artificial Intelligence Approaches

The harder question of learning context-free grammars has given rise to different approaches over the past few years. Typical artificial intelligence techniques have been used to search for a small consistent context-free grammar. For instance, Sakakibara and Kondo have used genetic algorithms (GAs) to search for a context-free grammar in Chomsky normal form, consistent with the given examples [85]. Experiments suggest that the knowledge of part of the structure (some parenthesis) may help and reduce the number of generations needed to identify the correct grammar and thus contribute to improving the efficiency of the learning algorithm [86], [88].

C. Reduction to Finite Automata Learning Problems

It is quite natural to solve one problem with the help of some other problem whose solution is known. In grammatical inference, a well-known technique often used to establish learnability results is a *reduction* technique that reduces a learning problem to some other learning problem. Takada has shown that the learning problem for even linear grammars can be solved by reducing it to the one for learning DFA [95].

An even linear grammar is a context-free grammar that has productions only of the form $A \rightarrow uBv$ or $A \rightarrow w$ such that u and v have the same length, where A and B are nonterminals and u, v and w are strings over Σ.

Theorem 3. ([95])**.** *The problem of learning the class of even linear grammars is reduced to the problem of learning the class of DFA.*

The class of even linear languages properly contains the class of regular languages and is a proper subclass of context-free languages. Takada has further developed an infinite hierarchy of families of languages whose learning problems are reduced to the learning problem of DFA [96].

D. Learning Subclasses of Context-Free Grammars

Since the identification of context-free grammars seems to be hard without any additional information, the grammatical inference scientists have always put their interest in designing efficient algorithms for identifying subclasses of context-free grammars from examples.

Ishizaka has investigated a subclass of context-free grammars, called *simple deterministic grammars*, and produced a polynomial time algorithm based on the theory of model inference given by Shapiro, for exactly identifying it using membership queries and extended equivalence queries in terms of general context-free grammars [44]. It should be noticed that these grammars are not linear.

A smaller class of simple deterministic grammars has been considered by Yokomori with the goal of finding a polynomial time algorithm to identify it in the limit from positive presentation [106]. He has shown the positive results regarding identification in the limit from positive presentations for the class of very simple grammars.

V. Tree Automata

Tree automata are the direct extension of DFA and NFA for tree languages instead of string. They may be top-down (starting from the root), or bottom-up (starting from the leaves), and deterministic, or nondeterministic. Survey paper [34] provides sufficient details for learning tree automata, and Kosala *et al*. present induction of ranked tree automata for information extraction from Web documents [51].

There is a very deep link between learning context-free grammars from structural information and learning regular tree grammars [80]. Sakakibara has investigated the algorithm *RT* for learning reversible tree automata efficiently from positive samples, and using *RT*, he also demonstrated an algorithm *RC* for learning reversible context-free grammars from positive structural samples [82]. Also, an efficient scheme for incremental learning of context-free grammars in this paradigm is presented by Prajapati *et al*. in [71].

The recent studies here are mainly for variety of tree languages. The algorithm *regular positive and negative inference*, proposed by Oncina and García, is capable of generalizing and identifying DFA. An extension of this algorithm to deal with trees is provided in [35], and a generalization of the well-known algorithm for regular languages learning from stochastic samples to the learning of tree languages is given in [23]. A mathematical basis for the case of learning from positive structural data only, is found in [50]. Moreover, for identifying tree languages from positive data only, Fernau presents a generic inference algorithm with polynomial update time [31].

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

VI. Learning Stochastic Grammars

Now we deal with the *stochastic grammars* emphasizing more on stochastic context-free grammars (SCFGs). Another challenging research problem in grammatical inference is stochastic modeling and training of stochastic grammars. Stochastic grammars have become a tool of wide use in speech recognition, natural language processing, bioinformatics and computational biology. In a stochastic grammar each production is assigned a probability, and therefore each string which it derives is also given a probability indirectly. Stochastic (probabilistic) automata are the probabilistic counterpart of finite automata that are known as *Hidden Markov Models* (HMMs) and are widely used in many applications including speech recognition [15], [55], [73], [46], biological sequence modeling [28], [17], [70], information extraction [91], optical character recognition [56], and learning GA parameters [74]. The potentials of using HMM in extracting free-structure corpus have also been presented in the literature [100]. It finds that HMM is a superior tool for extracting free-structured unique target term. Any grammar in the Chomsky hierarchy can be used in stochastic form. The class of SCFGs extends the class of context-free grammars, and therefore it goes one step beyond HMMs in the Chomsky hierarchy because HMMs extend regular grammars.

The problem of learning stochastic grammars from examples reduces to two sub problems; the first is to learn structure (topology) of the grammar and second is to determine probabilistic parameters in the grammar. There are two well-known efficient algorithms for automatic estimation of probabilities and distributions are available in the literature known as *forward-backward algorithm* for HMMs [72] and *inside-outside algorithm* for SCFGs [16], [54]. Recently [49] gives GA- based approach to the problem of inferring SCFGs from finite language samples. It shows the experiments for learning a range of formal languages and compares the results with those found using the inside-outside algorithm [16]. The results confirm that the learned grammars are compact and fit the corpus data efficiently.

A. Stochastic Context-Free Grammars

A stochastic context-free grammar (SCFG) G consists of a set of nonterminals N, a terminal alphabet Σ, a set P of productions with associated probabilities, and start symbol S. A probability distribution exists over the set of productions which have the same nonterminal on the left hand sides.

SCFGs have exploited in several applications including speech recognition [104] and computational biology [83]. Again, the question of learning probabilistic context-free grammars is hard. One elementary problem is to parsing with such a grammar [94]. Another basic problem that requires study is to check whether a given grammar is consistent with the input examples. This problem is considered in [19] with restrictions, and for guaranteeing the consistency for all SCFGs without restrictions, whenever the probability distributions are learned from the algorithms which are based on growth transformations such as the inside-outside algorithm is illustrated in [90]. Moreover, some other properties related to consistency are also given in [90]. The learning problem has the following possible levels.

If the grammar rules are available, we can try to estimate the probabilities that fit best. An efficient algorithm in this paradigm is demonstrated in [16], and Lari and Young present the estimation process using the well-known inside-outside algorithm in the paper [54]. Some other estimation techniques can also be seen in [83].

In other level of the learning problem, we can first learn the rules and then the probabilities. On having additional knowledge about the data, such as structural information, we can turn to learning rules using tree automata model (this line is followed in [80], and [82]), on the other hand the approach such as described in [75] may be followed. It has already been proved in the literature that the direct approach (without any heuristics) of learning directly the context-free grammars is hard and seems to be achieved only by some robust search techniques, like GAs [47].

VII. Other Language Representations

Type-3, type-2, type-1 and type-0 (from most restrictive to most general) are the four classes of grammars that Chomsky, in 1959, cataloged in a hierarchy of grammars according to the structure of their productions. Usually, formal grammars in Chomsky hierarchy are used for modeling languages. But other ways have also been applied for language representations. In the following, we discuss some of these.

A. About Pattern Languages

Angluin has originated the study of pattern languages [5]. The sole reason to use pattern is to reflect characteristics of the target language. They can be formed as sequences of letters, variables and wild-cards. Strings belong to the language if they obey the pattern. An example of pattern is a regular expression, which defines a regular language. Some typical applications of pattern languages are in text processing, automated data entry tasks, and bioinformatics. Different people have explored variants of pattern languages. A good picture of the situation, efficient robust (against noise) techniques and links to the field can be seen in [40]. The field is also covered in the prestigious forums ALT and COLT. There has been a lot of very specific research in the field. However, representative pattern problems have been found as hard. Some lines of research that have been followed to deal with the problem of inductive learning of pattern languages are described nicely in the literature [63], [52], [29], [45], and [79].

B. Categorial Grammars

The study of categorial grammars is mainly dates from 1935. In this grammar only small number of rules is used, and remaining syntactic behaviors are derived from the lexical entries of specific strings. A formal learning theory for the field within Gold's view of identification in the limit on positive examples has been explored in [48], and syntactic formalism based on Lambek categorial grammars

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

(a different class of categorial grammars) and semantic representation with logical formulas can be seen in paper [97]. The learnability results are different for different classes of grammars, *e.g.*, the negative results have been presented for different classes of Lambek grammars, whereas positive results are seen for *k*-valued classical categorial grammars ([33]). Another example of hardness can also be seen in [32].

VIII. Applications

There are number of problems where solutions based on grammatical inference ideas or techniques have been tested. However, it is still required to find a task where grammatical inference models have done much better than other machine learning or pattern recognition programs. In this section, we discuss some major applications of grammatical inference. The papers in pattern recognition [34], [62], Sakakibara's articles [84], [87] and Higuera's study in [42] are some places where survey work on the area along with its applications has taken place.

A. Robotics and Control Systems

Dean *et al.* consider an application of grammatical inference in map learning [27] where robot may learn with some possibility of error. Related work can also be seen in [77]. Moreover, Rieger constructs a prefix tree from robot traces [76]. This tree is used to derive and estimate the parameters of a deterministic, as well as a probabilistic automaton model for further navigation. Luzeaux has also used grammar induction techniques in the field of control theory [59]. Here grammatical inference technique has been applied to model the numerical data with regular grammars in order to acquire a qualitative model.

B. Structural Pattern Recognition

Syntactic and structural methods use models and techniques of formal language theory. Structural representations are useful to define likelihood relationship from a candidate pattern to a model of a class to design efficient algorithms for recognition tasks. Grammatical inference is a main example of syntactic and structural pattern analysis in machine learning. The articles [61], [62] and [34] present details of some of the applications of grammatical inference in the area. The work by Lucas *et al.*, where a most accurate method for character recognition, which process less than 1 character per second is seen in [58], and by Ney, for a general study in [66].

C. Computational Linguistics

Grammatical inference for natural languages is hard. However, it has always been in the interest of several scientists including Adriaans [2] to relate grammatical inference with natural language. The theoretical work of Adriaans is the basis of the Entity Modeling Intelligent Learning Engine prototype (*i.e.*, EMILE 4.1 implementation). This toolbox is useful to analyze the grammatical structure of free text.

D. Speech

In the field of speech recognition, obtaining the better language model is an important research goal in order to lower the error rate of the speech recognizer. Some typical language models are described in [46], [36], and [98]. A study for the construction of language model using grammatical inference techniques for language simplification task is illustrated in [4]. [103] achieves satisfactory results on the airline travel information system task [104] by learning smaller size context-free grammars.

E. Applications in Computational Biology

In molecular biology, we need to deal with very large size sequences of DNA, RNA and protein, and therefore the problems in this area are generally computationally hard. In fact, there is still ample room to apply grammar induction techniques in this discipline, and to introduce approaches like HMMs [53] and other approaches [43].

Determining similarity among a family of sequences, producing a multiple sequence alignment, classifying the members in different families according to their evolutionary distances and discovering the right ancestor of a candidate member will continue to be some of the most important and fundamental computational tasks. Wang *et al.* describe techniques for DNA sequence classification tasks in [102]. Sakakibara *et al.* learn stochastic context-free grammars from tRNA sequences [83]. The inferred grammar allows to discover and to model part of the secondary structure. The same line of work can also be seen in [1]. The work in [89], which presents a combination of context-free grammars and bi-gram models, obtains good results. Here, context-free grammars are used to represent relations of the structured part of RNA sequences and others that are not structured.

F. Other Applications

1) *Inductive Logic Programming.* As a subfield of machine learning, inductive logic programming [64] has also relation with grammatical inference. Boström's system merlin compiles the data by the background information in order to learn a DFA [21], or a HMM one [22].
2) *Data Mining*. Extraction of relevant information efficiently from large amounts of data is still a big issue in this field. Some of the places where grammar induction techniques have been tested for knowledge discovery include:
   a) Borges and Levene describe learning user behaviors from their navigation patterns based on the theory of probabilistic grammars [20].
   b) Chidlovskii *et al.* present a method for automatically generating wrappers for meta-search engines using incremental grammar induction algorithm for effective search on the World Wide Web [24].
3) *Music*. Other than the conventional application areas, *e.g.*, pattern recognition and language modeling, music processing is also one of the interesting application areas of grammatical inference. As an example, stochastic automata have been used to model musical styles (Renaissance, Baroque, Ragtime, Bach, etc.) [26]; the learned automata can then be used in automatic composition (to synthesize new melodies) or in automatic musical style recognition (to classify test melodies).

4) *Time Series Prediction.* Time series prediction is the use of a methodology to predict future behavior based on past experience. Giles *et al.*'s have used grammatical inference with recurrent neural networks for conversion into a symbolic representation [37]. Their method predicts the direction of change for the next day with an error rate of 47.1%.
5) *Document Management.* The job of writing, storing, and retrieving documents in electronic form has become popular now days. Context-free grammars are found as common means for describing the structure of a document. There are several possible applications of grammatical inference to deal with documents as data. Some typical applications can be seen in [3], and [108].

IX. Conclusion

We have found very few articles for learning context-sensitive grammars and also for phrase structure grammars; reasons certainly include higher computational complexity of these classes. Levy and Joshi have suggested an open problem to extend the methods of skeletal structural descriptions to arbitrary phrase structure grammars [57]. Arikawa *et al.* describe elementary formal systems for identification of context-sensitive languages [14]. Moreover, an approach for identification for subclasses of context-sensitive languages can also be found in [30].

While producing the material in this paper, our fundamental goal was to provide detailed introduction about the active and challenging area of grammatical inference (mainly, for beginners or someone interested to get research related activities in this field and to find most appropriate ideas or techniques for their own research work).

References

[1] Abe, N., and Mamitsuka, H., "Predicting Protein Secondary Structure Using Stochastic Tree Grammars," *Machine Learning Journal*, vol. 29, pp. 275–301, 1997.

[2] Adriaans, P., "Language Learning from a Categorical Perspective," *Ph.D. Thesis*, Universiteit van Amsterdam, 1992.

[3] Ahonen, H., Mannila, H., and Nikunen, E., "Forming Grammars for Structured Documents: An Application of Grammatical Inference," in: Carrasco, R.C., and Oncina, J. (Eds.), *Grammatical Inference: Algorithms and Applications, Proceedings of ICGI '94*, LNAI, vol. 862, Springer, Berlin, Heidelberg, pp. 153–167, 1994.

[4] Amengual, J.-C., Sanchis, A., Vidal, E., and Benedí, J.-M., "Language Simplification Through Error-Correcting and Grammatical Inference Techniques," *Machine Learning Journal*, vol. 44 (1), pp. 143–159, 2001.

[5] Angluin, D., "Finding Patterns Common to a Set of Strings," in: *Conference Record of the 11th Annual ACM Symposium on Theory of Computing*, ACM Press, New York, NY, USA, pp. 130–141, 1979.

[6] Angluin, D., "Inductive Inference of Formal Languages from Positive Data," *Information and Control*, vol. 45, pp. 117–135, 1980.

[7] Angluin, D., "A Note on the Number of Queries Needed to Identify Regular Languages," *Information and Control*, vol. 51, pp. 76–87, 1981.

[8] Angluin, D., "Inference of Reversible Languages," *Journal of the Association for Computing Machinery*, vol. 29, pp. 741-765, 1982.

[9] Angluin, D., "Learning Regular Sets from Queries and Counter-examples," *Information and Computation*, vol. 75, pp. 87-106, 1987.

[10] Angluin, D., "Queries and Concept Learning," *Machine Learning*, vol. 2, pp. 319-342, 1988.

[11] Angluin, D., "Negative Results for Equivalence Queries," *Machine Learning*, vol. 5, pp. 121-150, 1990.

[12] Angluin, D., and Kharitonov, M., "When Won't Membership Queries Help?" in: *Proceedings 23rd Annual ACM Symposium on Theory of Computing*, ACM Press, New York, pp. 444-454, 1991.

[13] Angluin, D., "Queries Revisited," in: Abe, et al. (Eds.), *Proceedings of ALT 2001*, Lecture Notes in Computer Science, vol. 2225, Springer, Berlin, Heidelberg, pp. 12–31, 2001.

[14] Arikawa, S., Shinohara, T., and Yamamoto, A., "Elementary Formal Systems as a Unifying Framework for Language Learning," in: *Proceedings 2nd Workshop on Computational Learning Theory*, Morgan Kaufmann, pp. 312-327, 1989.

[15] Bahl, L., Brown, P., Souza, P. de, and Mercer, R., "Estimating Hidden Markov Models Parameters so as to Maximize Speech Recognition Accuracy," *IEEE Transactions on Speech Audio Process*, vol. 1 (1), pp. 77–83, 1993.

[16] Baker, J.K., "Trainable Grammars for Speech Recognition," *Speech Communication Papers for the 97th Meeting of the Acoustical Society of America*, pp. 547-550, 1979.

[17] Baldi, P., and Brunak, S., *Bioinformatics: A Machine Learning Approach*, second ed., MIT Press, 2001.

[18] Biermann, A., "A Grammatical Inference Program for Linear Languages," in: *Fourth Hawaii International Conference on System Sciences*, pp. 121–123, 1971.

[19] Booth, T.L., and Thompson, R.A., "Applying Probability Measures to Abstract Languages," *IEEE Transactions on Computers*, vol. C-22 (5), pp. 442–450, 1973.

[20] Borges, J., and Levene, M., "Data Mining of User Navigation Patterns," in: Masand, B., and Spiliopoulou, M. (Eds.), Web Usage Mining and User Profiling, *Lecture Notes in Computer Science*, vol. 1836, Springer, Berlin, pp. 92-111, 2000.

[21] Boström, H., "Theory-Guided Induction of Logic Programs by Inference of Regular Languages," in: *13th International Conference on Machine Learning*, Morgan Kaufmann, Los Altos, CA, 1996.

[22] Boström, H., "Predicate Invention and Learning from Positive Examples Only," in: Nédellec, C., and Rouveirol, C. (Eds.), *10th European Conference on Machine Learning*, Lecture Notes in Artificial Intelligence, vol. 1398, Springer, Berlin, pp. 226–237, 1998.

[23] Carrasco, R.C., Oncina, J., and Calera-Rubio, J., "Stochastic Inference of Regular Tree Languages," *Machine Learning Journal*, vol. 44 (1), pp. 185–197, 2001.

[24] Chidlovskii, B., Ragetli, J., and Rijke, M. de, "Wrapper Generation via Grammar Induction," in: *Machine Learning*: ECML 2000, *11th European Conference on Machine Learning*, vol. 1810, Springer, Berlin, pp. 96-108, 2000.

[25] Crespi-Reghizzi, S., "An Effective Model for Grammar Inference," in: Gilchrist, B. (Ed.), *Information Processing*, vol. 71, Elsevier, Amsterdam, pp. 524-529, 1972.

[26] Cruz, P., and Vidal, E., "Learning Regular Grammars to Model Musical Style: Comparing Different Coding Schemes," in: Honavar, V., and Slutski, G. (Eds.), *Grammatical Inference, Proceedings of ICGI '98*, Lecture Notes in Artificial Intelligence, vol. 1433, Springer, Berlin, Heidelberg, pp. 211–222, 1998.

[27] Dean, T., Basye, K., Kaelbling, L., Kokkevis, E., Maron, O., Angluin, D., and Engelson, S., "Inferring Finite Automata with Stochastic Output Functions and an Application to Map Learning," in: Swartout, W. (Ed.), *Proceedings of the 10th National Conference on Artificial Intelligence,* MIT Press. San Jose, CA, pp. 208-214, 1992.

[28] Durbin, R., Eddy, S., Krogh, A., and Mitchison, G., *Biological Sequence Analysis*, Cambridge University Press, Cambridge, 1998.

[29] Erlebach, T., Rossmanith, P., Stadtherr, H., Steger, A., and Zeugmann, T., "Learning One-Variable Pattern Languages Very Efficiently on Average, in Parallel, and by Asking Queries," in: Li, M., and Maruoka, A. (Eds.), *Proceedings of ALT '97*, Lecture Notes in Computer Science, vol. 1316, Springer, Berlin, Heidelberg, pp. 260–276, 1997.

[30] Fahmy, A.F., and Biermann, A.W., "Synthesis of Real Time Acceptors," *Journal of Symbolic Computation*, vol. 15, pp. 807-842, 1993.

[31] Fernau, H., "Learning Tree Languages from Text," in: Kivinen, J., and Sloan, R.H. (Eds.), *Proceedings of COLT 2002*, LNAI, vol. 2375, Springer, Berlin, Heidelberg, pp. 153–168, 2002.

[32] Florêncio, C.C, "Consistent identification in the Limit of Rigid Grammars from Strings is NP-hard," in: Adriaans, P., et al. (Eds.), *Grammatical Inference: Algorithms and Applications Proceedings of ICGI '00*, Lecture Notes in Artificial Intelligence, vol. 2484, Springer, Berlin, Heidelberg, pp. 49–62, 2002.

[33] Foret, A., and Nir, Y.L., "On Limit Points for Some Variants of Rigid Lambek Grammars," in: Adriaans, P. et al. (Eds.), *Grammatical Inference: Algorithms and Applications, Proceedings of ICGI '00,* Lecture Notes in Artificial Intelligence, vol. 2484, Springer, Berlin, Heidelberg, pp. 106–119. 2002.

[34] Fu, K.S., and Booth, T.L., "Grammatical Inference: Introduction and Survey. Part I and II," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 5, pp. 59 -72 and 409-423, 1975.

[35] García, P., and Oncina, J., "Inference of Recognizable Tree Sets," *Technical Report DSIC-II/47/93*, Departamento de Lenguajes y Sistemas Informáticos, Universidad Politécnica de Valencia, Spain, 1993.
[36] García, P., Segarra, E., Vidal, E., and Galiano, I., "On the Use of the Morphic Generator Grammatical Inference (mggi) Methodology in Automatic Speech Recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 4, pp. 667-685, 1994.
[37] Giles, C.L., Lawrence, S., and Tsoi, A. "Noisy Time Series Prediction Using Recurrent Neural Networks and Grammatical Inference," *Machine Learning Journal*, vol. 44 (1), pp. 161–183, 2001.
[38] Gold, E. M., "Language Identification in the Limit," *Information and Control*, vol. 10**,** pp. 447-474, 1967.
[39] Gold, E.M., "Complexity of Automaton Identification from Given Data," *Information and Control*, vol. 37, pp. 302–320, 1978.
[40] Goldman, S.A., and Kwek, S.S., "On Learning Unions of Pattern Languages and Tree Patterns," in: Watanabe, O., and Yokomori, T. (Eds.), *Proceedings of the 10th International Conference on Algorithmic Learning Theory (ALT 1999)*, Lecture Notes in Computer Science, vol. 1720, Springer, Berlin, pp. 347–363, 1999.
[41] Higuera, C. de la, "Characteristic Sets for Polynomial Grammatical Inference," *Machine Learning Journal*, vol. 27, pp. 125–138, 1997.
[42] Higuera, C. de la, "A Bibliographical Study of Grammatical Inference," *Pattern Recognition*, vol. 38, pp. 1332-1348, 2005.
[43] Hunter, L., *Artificial Intelligence and Molecular Biology*, AAAI Press/MIT Press, New York, 1993.
[44] Ishizaka, H., "Polynomial Time Learnability of Simple Deterministic Languages," *Machine Learning*, vol. 5, pp. 151-164, 1990.
[45] Jantke, K.P., and Lange, S., "Case-Based Representation and Learning of Pattern Languages," *Theoretical Computer Science*, vol. 137 (1), 25–51, 1995.
[46] Jelinek, F., *Statistical Methods for Speech Recognition*, MIT Press, Cambridge, MA, 1998.
[47] Kammeyer, T., and Belew, R.K., "Stochastic Context-Free Grammar Induction with a Genetic Algorithm Using Local Search," in: Belew, R.K., and Vose, M. (Eds.), *Foundations of Genetic Algorithms IV*, Morgan Kaufmann, University of San Diego, CA, USA, 1996.
[48] Kanazawa, M., *Learnable Classes of Categorial Grammars*, CSLI Publications, Stanford, CA, 1998.
[49] Keller, B., and Lutz, R., "Evolutionary Induction of Stochastic Context-Free Grammars," *Pattern Recognition*, vol. 38, pp. 1393-1406, 2005.
[50] Knuutila, T., and Steinby, M., "Inference of Tree languages from a Finite Sample: an Algebraic Approach," *Theoretical Computer Science*, vol. 129, pp. 337–367, 1994.
[51] Kosala, R., Blockeel, H., Bruynooghe, M., and Bussche, J.V. den, "Information Extraction from Structured Documents Using k-Testable Tree Automaton Inference," *Data & Knowledge Engineering*, vol. 58, pp. 129–158, 2006.
[52] Koshiba, T., "Typed Pattern Languages and Their Learnability," in: *Proceedings 2nd European Conference on Computational Learning Theory* (EuroCOLT'95), Lecture Notes in Artificial Intelligence, vol. 904, Springer, Berlin, pp. 367-379, 1995.
[53] Krogh, A., Brown, M., Mian, I. S., Sjölander, K., and Haussler, D., "Hidden Markov Models in Computational Biology: Applications to Protein Modeling," *Journal of Molecular Biology*, vol. 235, pp. 1501-1531, 1994.
[54] Lari, K., and Young, S.J., "The Estimation of Stochastic Context-Free Grammars Using the Inside-Outside Algorithm," *Computer Speech and Language*, vol. 4, pp. 35-56, 1990.
[55] Lee, K.F., "Large Vocabulary Speaker Independent Continuous Speech Recognition: the SPHINX System," *Ph.D. Dissertation*, Computer Science Department, Carnegie Mellon University, 1988.
[56] Levin, E., and Pieraccini, R., "Planar Hidden Markov Modeling: From Speech to Optical Character Recognition," in: Giles, C.L., Hanton, S.J. and Cowan, J.D. (Eds.), *Advances in Neural Information Processing Systems*, vol. 5, Morgan Kauffman, Los Altos, CA, pp. 731–738, 1993.
[57] Levy, L.S., and Joshi, A.K., "Skeletal Structural Descriptions," *Information and Control*, vol. 39, pp. 192-211, 1978.
[58] Lucas, S., Vidal, E., Amari, A., Hanlon, S., and Amengual, J.C., "A Comparison of Syntactic and Statistical Techniques for Offline OCR," in: Carrasco, R.C., and Oncina, J. (Eds.), *Grammatical Inference and Applications, Proceedings of ICGI '94*, Lecture Notes in Artificial Intelligence, vol. 862, Springer, Berlin, Heidelberg, pp. 168–179, 1994.
[59] Luzeaux, D., "Machine Learning Applied to the Control of Complex Systems," in: *Proceedings of the 8th International Conf. on Artificial Intelligence and Expert Systems Applications*, Paris, France, 1996.
[60] Makinen, E., "On the Structural Grammatical Inference Problem for Some Classes of Context-Free Grammars," *Information Processing Letters*, vol. 42, pp. 193-199, 1992.
[61] Miclet, L., *Structural Methods in Pattern Recognition*, Chapman & Hall, New York, 1986.
[62] Miclet, L., "Grammatical Inference, in: Syntactic and Structural Pattern Recognition, Theory and Applications," *World Scientific*, Singapore, pp. 237-290, 1990.
[63] Mitchell, A., Scheffer, T., Sharma, A., and Stephan, F., "The VC-Dimension of Subclasses of Pattern Languages," in: Watanabe, O., and Yokomori, T. (Eds.), *Proceedings of the 10th International Conference on Algorithmic Learning Theory* (ALT 1999), LNCS, vol. 1720, Springer, Berlin, pp. 93–105, 1999.
[64] Muggleton, S., *Inductive Logic Programming*, in: The MIT Encyclopedia of the Cognitive Sciences (MITECS), MIT Press, Cambridge, MA, 1999.
[65] Natarajan, B.L., *Machine Learning: A Theoretical Approach*, Morgan Kauffman Pub., San Mateo, CA, 1991.
[66] Ney, H., "Stochastic Grammars and Pattern Recognition," in: Laface, P., and Mori, R.D. (Eds.), *Proceedings of the NATO Advanced Study Institute*, Springer, Berlin, pp. 313–344, 1992.
[67] Oncina, J., Garcia, P., and Vidal, E., "Learning Subsequential Transducers for Pattern Recognition Interpretation Tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 448-458, 1993.
[68] Pitt, L., "Inductive Inference, DFAs, and Computational Complexity," in: *Proceedings AII-89 Workshop on Analogical and Inductive Inference*, Lecture Notes in Computer Science, vol. 397, Springer, Berlin, pp. 18-44, 1989.
[69] Pitt, L., and Warmuth, M.K., "The Minimum Consistent DFA Problem Cannot be Approximated Within any Polynomial," in: *Proceedings 21st Annual ACM Symposium on Theory of Computing*, ACM Press, New York, 1989.
[70] Plötz., T., and Fink, G.A., "Pattern Recognition Methods for Advanced Stochastic Protein Sequence Analysis Using HMMs," *Pattern Recognition*, vol. 39, pp. 2267-2280, 2006.
[71] Prajapati G.L., Chaudhari N.S., and Chandwani M., "Efficient Incremental Model for Learning Context-Free Grammars from Positive Structural Examples," 5th Hellenic Conference on Artificial Intelligence, SETN-08, Syros, Greece, Proceedings in *Artificial Intelligence: Theories, Models and Applications*, Lecture Notes in Artificial Intelligence (LNAI), Springer Verlag, Berlin, ISSN: 0302-9743, ISBN: 978-3-540-87880-3, vol. 5138, pp. 250-262, 2008.
[72] Rabiner, L.R., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings IEEE*, vol. 77(2), pp. 257-286, 1989.
[73] Rabiner, L., and Juang, B.-H., *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
[74] Rees, J., and Koehler, G.J., "Learning Genetic Algorithm Parameters Using Hidden Markov Models," *European Journal of Operational Research*, vol. 175, pp. 806-820, 2006.
[75] Rico-Juan, J.R., Calera-Rubio, J., and Carrasco, R.C., "Stochastic *k*-Testable Tree Languages and Applications," in: Adriaans, P., et al. (Eds.), *Grammatical Inference: Algorithms and Applications, Proceedings of ICGI '00*, Lecture Notes in Artificial Intelligence, vol. 2484, Springer, Berlin, Heidelberg, pp. 199–212, 2002.
[76] Rieger, A., "Inferring Probabilistic Automata from Sensor Data for Robot Navigation," in: Kaiser, M. (Ed.), *Proceedings of the MLnet Familiarization Workshop and Third European Workshop on Learning Robots*, pp. 65–74, 1995.
[77] Rivest, R.L., and Schapire, R.E., "Inference of Finite Automata Using Homing Sequences," *Information and Computation*, vol. 103, pp. 299–347, 1993.
[78] Ron, D., and Rubinfeld, R., "Learning Fallible Deterministic Finite Automata," *Machine Learning*, vol. 18, pp. 149-185, 1995.
[79] Rossmanith, P. and Zeugmann, T., "Stochastic Finite Learning of the Pattern Languages," *Machine Learning Journal*, vol. 44 (1), pp. 67–91, 2001.
[80] Sakakibara, Y., "Learning Context-Free Grammars from Structural Data in Polynomial Time," *Theoretical Computer Science*, vol. 76, pp. 223-242, 1990.
[81] Sakakibara, Y., "On learning from Queries and Counter- Examples in the Presence of Noise," *Information Processing Letters*, vol. 37, pp. 279-284, 1991.
[82] Sakakibara, Y., "Efficient Learning of Context-Free Grammars from Positive Structural Examples," *Information and Computation*, vol. 97, pp. 23-60, 1992.
[83] Sakakibara, Y., Brown, M., Hughey, R., Mian, I. S., Sjolander, K., Underwood, R. C., and Haussler, D., "Stochastic Context-Free Grammars for tRNA Modeling," *Nucleic Acid Research*, vol. 22, pp. 5112-5120, 1994.
[84] Sakakibara, Y., "Recent Advances of Grammatical Inference," *Theoretical Computer Science*, vol. 185, pp. 15-45, 1997.

[85] Sakakibara, Y., and Kondo, M., "GA-Based Learning of Context-Free Grammars Using Tabular Representations," In: *Proceedings of 16th International Conf. Machine Learning*, pp. 354–360, 1999.

[86] Sakakibara, Y., and Muramatsu, H., "Learning Context-Free Grammars from Partially Structured Examples," in: Oliveira, A. de (Ed.), *Grammatical Inference: Algorithms and Applications, Proceedings of ICGI '00*, Lectures Notes in Artificial Intelligence, vol. 1891, Springer, Berlin, Heidelberg, pp. 229–240, 2000.

[87] Sakakibara, Y., "Grammatical Inference in Bioinformatics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1051-1062, 2005.

[88] Sakakibara, Y., "Learning Context-Free Grammars Using Tabular Representations," *Pattern Recognition*, vol. 38, pp. 1372-1383, 2005.

[89] Salvador, I., and Benedí, J.-M., "RNA Modeling by Combining Stochastic Context-Free Grammars and *n*-Gram Models," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 16 (3), pp. 309–316, 2002.

[90] Sánchez, J.A., and Benedí, J.M., "Consistency of Stochastic Context-Free Grammars from Probabilistic Estimation Based on Growth Transformation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19 (9), pp. 1052–1055, 1997.

[91] Seymore, K., McCallum, A., and Rosenfeld, R., "Learning Hidden Markov Model Structure for Information Extraction," in: *AAAI'99 Workshop on Machine Learning for Information Extraction*, Orlando, Florida, pp. 37–42, 1999.

[92] Shinohara, T., "Rich Classes Inferable from Positive Data: Length-Bounded Elementary Formal Systems," *Information and Computation*, vol. 108, pp. 175-186, 1994.

[93] Solomonoff, R., "A Formal Theory of Inductive Inference," *Information and Control*, vol. 7, pp. 224–254, 1964.

[94] Stolcke, A., "An Efficient Probabilistic Context-Free Parsing Algorithm that Computes Prefix Probabilities," *Computational Linguistics*, vol. 21 (2), pp. 165–201, 1995.

[95] Takada, Y., "Grammatical Inference for Even Linear Languages Based on Control Sets," *Information Processing Letters*, vol. 28, pp. 193-199, 1988.

[96] Takada, Y., "A Hierarchy of Language Families Learnable by Regular Language Learners," *Information and Computation*, vol. 123, pp. 138-145, 1995.

[97] Tellier, I., "Meaning Helps Learning Syntax," in: Honavar, V., and Slutski, G. (Eds.), *Grammatical Inference, Proceedings of ICGI '98*, Lecture Notes in Artificial Intelligence, vol. 1433, Springer, Berlin, Heidelberg, pp. 25–36, 1998.

[98] Thollard, F., Dupont, P., and Higuera, C. de la, "Probabilistic DFA Inference Using Kullback-Leibler Divergence and Minimality," in: *Proceedings of the 17th International Conference on Machine Learning*, Morgan Kaufmann, San Francisco, CA, pp. 975–982, 2000.

[99] Trakhtenbrot, B., and Barzdin, Y., *Finite Automata: Behavior and Synthesis*, North-Holland, Amsterdam, 1973.

[100] Tso, B., and Chang, P.Y., "Mining Free-Structured Information Based on Hidden Markov Models," *Expert Systems with Applications*, vol. 32, pp. 97–102, 2007.

[101] Valiant, L.G., "A Theory of the Learnable," *Communications of the ACM* vol. 27, pp. 1134-1142, 1984.

[102] Wang, J.T.-L., Rozen, S., Shapiro, B.A., Shasha, D., Wang, Z., and Yin, M., "New Techniques for DNA Sequence Classification," *Journal of Computational Biology*, vol. 6 (2), pp. 209–218, 1999.

[103] Wang, Y., and Acero, A., "Grammar Learning for Spoken Language Understanding," in: *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding*, 2001.

[104] Wang, Y., and Acero, A., "Evaluation of Spoken Language Grammar Learning in the ATIS Domain," in: *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 2002.

[105] Wiehagen, R., "From Inductive Inference to Algorithmic Learning Theory," *New Generation Computing*, vol. 12, pp. 321-335, 1994.

[106] Yokomori, T., "Polynomial-Time Learning of Very Simple Grammars from Positive Data," in: *Proceedings 4th Workshop on* COLT, Morgan Kaufmann, Los Altos, CA, pp. 213-227, 1991.

[107] Yokomori, T., "Learning Nondeterministic Finite Automata from Queries and Counterexamples," in: Furukawa, Michie, and Muggleton (Eds.), *Machine Intelligence*, vol. 13, Oxford University Press, Oxford, pp. 169-189, 1994.

[108] Young-Lai, M., and Tompa, F.W., "Stochastic Grammatical Inference of Text Database Structure," *Machine Learning Journal*, vol. 40 (2), pp. 111–137, 2000.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)