A Proposal for Contextual Grammatical Inference

Leonor Becerra-Bonache(✉), María Galván, and François Jacquenet

Laboratoire Hubert Curien, Jean Monnet University, 18 rue Benoit Lauras, 42000 Saint-Etienne, France
{leonor.becerra,maria.galvan,francois.jacquenet}@univ-st-etienne.fr

Abstract. Grammatical Inference deals with the learning of formal languages from data. Research in this field has mainly reduced the problem of language learning to syntax learning. Taking into account that the theoretical results obtained in Grammatical Inference show that learning formal languages only from syntax is generally hard, in this paper we propose to also take into account contextual information during the language learning process. First, we review works in the area of Artificial Intelligence that use the concept of context, and then, we present the theoretical, algorithmic and practical aspects of our proposal.

Keywords: Grammatical inference · Context · Natural language

1 Introduction

Grammatical Inference (GI) is a specialized subfield of Machine Learning that deals with the learning of formal languages from a set of data. Roughly speaking, in a GI problem, we have a learner (or learning algorithm) that learns a language from the information that a teacher provides to it. Hence, a GI problem has some similarities with the problem of children's language acquisition: a child, like the learning algorithm, learns a language from the data that he/she receives from the linguistic community that is around him/her. Moreover, depending on the input that the child receives, he/she will learn one language or another, that is why a child growing up in a linguistic community that speaks Spanish acquires Spanish, but if the language spoken by the community is Chinese, the child will learn Chinese.

The initial theoretical foundations of GI were given by E.M. Gold at the end of 60's [20]. His goal was to formalize the acquisition of natural language. Since then, a big amount of research has been done to establish a theory of GI, to find efficient methods for inferring grammars, and to apply those methods to practical problems (e.g., syntactic pattern recognition, adaptive intelligent agents, computational biology, natural language processing, etc.). An excellent survey of the field can be found in [15].

Most of the research that has been developed within the field of GI has focused on *syntax learning* (i.e., on learning the rules of a grammar), and tends

I. Rojas et al. (Eds.): IWANN 2015, Part I, LNCS 9094, pp. 18–28, 2015.
DOI: 10.1007/978-3-319-19258-1_2

to omit any other kind of information. The theoretical results obtained in GI show that learning formal languages only from syntax is generally hard [15,20], usually leading to negative results. What about taking into account another kind of information during the language learning process?

Linguistic and cognitive studies suggest that *contextual information* seems to play an important role in the early stages of children's linguistic development [14]. For instance, let us take a conversation extracted from the CHILDES database [24] (i.e., Child Language Data Exchange System, which collects transcripts of child-parent dialogs):

Abe: milk. milk.
Father: you want milk?
Abe: uh-huh.
Father: ok. Just a second and I'll get you some.

The child in this conversation is two and a half years old, and she is concretely in the linguistic stage called two-word stage, in which children go from the production of one word to the combination of two elements. Specially in this stage, the correspondence between sentences and the context in which they are made seems to be a very important source of information for both: the child, trying to learn the language, and the adult, trying to make sense of the imperfect sentences produced by the child. For example, in the previous conversation, the child produces only two words as *milk milk* to express a whole sentence like *I want milk*. Thanks to the *context* in which the sentence is produced (which is shared by both the adult and the child), the adult can understand the meaning of the child's sentence, although it is not syntactically correct.

Therefore, contextual information seems to play an important role in language acquisition. The presence of this kind of information not only seems to facilitate (i.e. to speed-up) the learning process to the child, but also allows the communication between adult and child [14].

Taking into account that, in natural situations, contextual information is also available to the child, and that GI studies show that learning from only syntax is hard, the following questions arise: Why not to take contextual information into account during the learning process to improve language learning? Can contextual information simplify (speed-up) the learning problem?

The remainder of the paper is organized as follows. An overview of different works developed in the field of Artificial Intelligence that take into account the context is presented in Section 2. This section is divided into two subsections: Section 2.1 is dedicated to exploring different existing theories about context, and Section 2.2 reviews works on context modeling. Then, Section 3 presents the theoretical, algorithmic and practical aspects of our proposal. Finally, concluding remarks are presented in Section 4.

2 Context in Artificial Intelligence

2.1 Theories of Context

The notion of formalized contexts was introduced by McCarthy [26] in his Turing Award Lecture in the late 80's as a way of focussing the problem of generality in Artificial Intelligence. Thenceforth, it has been extensively discussed in Artificial Intelligence and other fields. Three main formalizations have been developed: *Propositional Logic of Context* (PLC), *Local Models Semantics/MultiContext Systems* (LMS/MCS) and *Situation Theory*. Next we briefly review these approaches and related papers.

Propositional Logic of Context. McCarthy [27] worked to formalize context and to develop a theory of introducing context as formal objects. He introduced a new modality $ist(c, p)$ (pronounced as "is true"), meaning that the proposition p is true in the context c. Guha's [23] PhD dissertation under McCarthy's supervision was the first in doing a depth study of context. Guha extended McCarthy's notion of context and motivated the CYC[1] ontology together with D. Lenat [22]. Knowledge statements in CYC were divided into microtheories, which become a common sense knowledge base. Each microtheory contains a set of axioms or rules and a vocabulary which provides the syntax and semantics with predicates and functions.

From these works, Buvač and others developed a *Propositional Logic of Context* (PLC) [10,11] and a *Quantificational Logic of Context* [9,25]. They described the syntax and semantics of a general propositional language of context. For this, they introduced $ist(c, p)$, meaning that the sentence p holds in the context c, and each context has its own vocabulary, i.e. a set of propositional atoms which are defined or meaningful in that context.

Local Models Semantics/MultiContext Systems. Working under a different approach, Giunchiglia [18] reformulated the problem of context in terms of locality of reasoning. He devised a formalization of context based on the problem of locality. It consists in a problem of modeling reasoning using a subset of information that reasoners know about world. Namely, to solve a problem set in a concrete situation, people do not use all their knowledge; by contrast, it is solved by a *local theory* considering who really know all the essential events of the problem. This proposal differs from reasoning, since people can move between contexts, changing from one to another when necessary. Definitively, this approach gave more importance to *formalize contextual reasoning* than *formalizing contexts as first-class objects*.

In *MultiContext Systems* (MCS), devised by Giunchiglia and Serafini [19], they introduced the idea of bridging rules. These rules relate knowledge among other contexts. MCS later became associated with the *Local Model Semantics* proposed by Ghidini and Giunchiglia [17]. Ghidini and Giunchiglia argued that

[1] CYC: http://www.cyc.com/

there are two main principles underlying the use of context: the *principle of locality* (reasoning uses a part of the environment that is available, which is called context) and the *principle of compatibility* (different kinds of reasoning performed in an environment are compatible).

Situation Theory. *Situation Theory* is a theory of meaning and communication in some situations which are recognized as primary events (as opposed to derivatives). Barwise and Perry [6] argued that for an expression to be meaningful, it should transmit information. They assumed that people use language in limited parts of the world to discuss other limited parts of the world (situations). So, they developed a theory of situations as a relation between these situations. The theory provides a system of abstract objects that makes it possible to describe the meaning of both expressions and mental states in terms of the information they carry about the external world [16].

The main ideas of *Situation Theory* are infons and situations. Infons are the basic units of information and they are of the form: $<< P, a_1, ..., a_n, i >>$ where P is an n-place relation, the value i represents polarity and a_1, ..., a_n are appropriate objects for P.

Following Barwise [7], Akman and Surav treated context as "an amalgamation of a grounding situation and the rules which govern the relations within the context" [5,31].

2.2 Examples of Context Modeling

Several authors have tried to model the context. Here we present an overview of some of these works.

Nickles and Rettinger [28] introduced a *Relational Reinforcement Learning* (RRL) approach. It uses symbolic interaction between artificial agents and humans for learning denotational concept semantics. This framework does not impose any specific kind of formal context. Results were mostly positive and showed the applicability and significance of the learning framework for realistic semantic learning tasks. They could show that the agent: (a) can learn different meanings of a concept, (b) scales to different levels of complexity, including a very ambitious 10 blocks world, (c) uses communication to ease the task, (d) deals with non-stationarity and more complex scenarios.

Goldwasser and Roth [21] were interested in providing a way for a human teacher to interact with an automated learner using natural instructions, thus allowing the teacher to communicate the relevant domain expertise to the learner without necessarily knowing anything about the internal representations used in the learning process. They evaluated their approach by applying it to the rules of the solitaire card game. They showed that their learning approach can eventually use natural language instructions to learn the target concept and play the game legally. Furthermore, they showed that the learned semantic interpreter also generalizes to previously unseen instructions. This presents an analogy to human learning, where a learner tests her understanding in an actionable setting. Such a setting can be simulated as a world environment in which the linguistic representation can be directly executed by a computer system. For example, in semantic parsing, the learning goal is to produce and successfully execute a meaning representation. Executable system actions include access to databases of simulated card games.

Chen [13] focused his research on how to build systems that use both text and the perceptual context in which it is used in order to learn a language. He first presented a system that can describe events in RoboCup 2D simulation games by learning only from sample language commentaries paired with traces of simulated activities without any language-specific prior knowledge. By applying an EM-like algorithm (Expectation-Maximization algorithm), the system was able to simultaneously learn a grounded language model as well as align the ambiguous training data. The goal of this research was to learn to map words and phrases into logical components that can be composed together to form complete meanings. Chen studied the problem in simulated environments that retained many of the important properties of a dynamic world with multiple agents and actions while avoiding many of the complexities of robotics and computer vision. His commentator system learned to semantically interpret and generate language in the RoboCup soccer domain by observing an on-going commentary of the game paired with the evolving simulator state.

The process that determines the emotional tone in a series of words is known as *Sentiment Analysis*. It is used to try to understand the attitudes, opinions and emotions expressed by a writer in an online publication. Context is an instrument to one of the existing methods for carrying out this process, as it can be seen in the paper written by Vanzo et al. [33]. They analysed the feelings in the social network (concretely Twitter). They considered a tweet within its context, i.e. the stream of related posts. In this case they studied three types of contextual information for a target tweet: an explicit conversation, the user attitude and the overall set of recent tweets about a topic. They considered the sentiment prediction as a sequential classification task over a context that associates tweets and for this, they proposed the SVM^{hmm} algorithm. Finally, they concluded that contextual information is relevant because it eliminates the ambiguity of the sentiment polarity.

3 Our Proposal

As we have pointed out previously, research in the field of GI does not take into account some relevant aspects of natural language learning like, for example, the availability of contextual information to the child. In this paper we propose to exploit syntactic and contextual information for language learning. In order to do that, we are going to focus on three different aspects: theoretical, algorithmic and practical aspects.

3.1 Theoretical Aspects

Three important formal models of language learning have been widely investigated in the field of GI.

- *Identification in the limit*, proposed by E.M. Gold [20]. In this model, examples of an unknown language are presented to the learner (or learning algorithm), and the learner has to produce a hypothesis of this language. Its hypothesis is updated after receiving each example; if the new examples received are not consistent with the current hypothesis, it changes its hypothesis. According to Gold's model, the learner identifies the target language *in the limit* if after a finite number of examples, the learner makes a correct hypothesis and does not change it from there on. There are two traditional settings within this model: a) learning from text, where only examples of the target language are given to the learner (i.e., only positive data); b) learning from informant, where examples that belong and do not belong to the target language are provided to the learner (i.e., positive and negative information).
- *Query learning model*, proposed by D. Angluin [1]. In this model, also known as active learning model, the learner is allowed to *interact* with the teacher, by making questions about the strings of the language. There are different kinds of queries, but the standard combination to be used are: a) *membership queries*: the learner asks if a concrete string belongs to the target language and the teacher answers "yes" or "no"; b) *equivalence queries*: the learner asks if its hypothesis is correct and the teacher answers "yes" if it is correct or otherwise gives a counterexample. According to Angluin's model, the learner has successfully learnt the target language if it returns the correct hypothesis after asking a finite number of queries.
- PAC learning model, proposed by L.G. Valiant [32]. The Probably Approximately Correct (PAC) model is a probabilistic model of learning from random examples. According to this model, the learner receives examples sampled under an unknown distribution. The learner is required to learn under any distribution, but exact learning is not required (since one may be unlucky during the sampling process). A successful learning algorithm is one that with high probability finds a grammar whose error is small.

As we can see, each of these models is based on different learning settings (what kind of data is used in the learning process and how these data are provided to the learner) and different criteria for a successful inference (under what conditions we say that a learner has been successful in the language learning task). Although they have aspects that make them useful to study the problem of natural language acquisition to a certain extent, other aspects make them unsuitable for this task. For example:

- in Gold's model, the definition of identification in the limit postulates greatly idealized conditions, as compared to the real-world conditions under which children learn language. Moreover, it makes some assumptions that are controversial from a linguistic point of view. For instance, there is not limit on how long it can take the learner to guess the correct language (but children are able to learn language in an efficient way), the learner hypothesizes complete grammars instantaneously (this is not the case in children's language acquisition), and the learner passively receives strings of the language (but children also interact with their environment).

- in Angluin's model, the queries are quite unnatural for real learning environments (a child will never ask if his/her grammar is the correct one). Moreover, the learner has to learn exactly the target language (but everybody has imperfections in their linguistic competence) and the teacher is assumed to be perfect (i.e., he knows everything and always gives the correct answers), which is an ideal teacher that does not occur in a real situation.
- in Valiant's model, the requirement that the examples have the same distribution throughout the process is too strong for practical situations.

Moreover, a common feature of all these formal models is that they do not take into account some relevant aspects of natural language learning, like for example, the availability of contextual information to the child. These models take into account only the syntax; so, their goal is to learn the rules of a grammar by taking just into account if, for example, a string belongs or not to the language, but not the context in which this string has been produced. Therefore, we claim that the existing models in GI are not well adapted to introduce contextual information.

Due to the absence of formal models in GI that consider this natural aspect during the learning process, one of our goals is to develop a new theoretical framework that takes syntactic and contextual information into account for language learning.

To reach this goal, we will need to answer some crucial questions. First, how should we represent contextual information? As we have seen in the previous section, there exists different possibilities. We will opt for representing contextual information in the form of a logical language. We will explore various fields of logical languages ranging from some old formalisms such as (constraint) first order logic to more recent ones such as description logic, widely used nowadays in the area of semantic web, leading to the design and use of ontologies.

Second, how should we prove that taking contextual information into account during the learning process can not only make learning easier, but also faster? We postulate that the use of contextual information will reduce the number of data necessary to theoretically learn the target language. In other words, we think that we should be able to drive tighter generalization bounds to this additional and meaningful information. Moreover, these theoretical results should also allow us to well understand why we obtained better results with contextual information, and answer the following questions: How contextual information speeds up the learning process? What guarantees the convergence of the algorithm?, etc.

This theoretical study will thus give us a framework to develop learning algorithms that use syntactic and contextual information during the learning process.

3.2 Algorithmic Aspects

Ideally, in order to correctly simulate some aspects of natural language acquisition, the information provided to the algorithm should be the same as the one available to children. As we have seen, contextual information not only is one component of language learning, but also seems to play an important role in the first stages of children's language acquisition. Therefore, it is also of great interest to study this component and the way new learning algorithms could take it into account during the learning process.

Unfortunately, most learning algorithms in GI have been focused on syntax learning, omitting any consideration to semantics. Thus, the problem of language learning has been mainly reduced to syntax learning. For example, many algorithms in GI are based on state merging mechanisms applied on (probabilistic) automata. More precisely, these methods allow us to generalize the knowledge by merging step by step some states (which describe prefixes) while not challenging some statistical properties (e.g. in ALERGIA [12]) or the correct classification of the training data (e.g. in RPNI [29]). Therefore, this kind of algorithms are definitely not adapted to our context which has also to consider contextual information.

Our algorithmic objective is to develop new algorithms, theoretically founded, that incorporate contextual information for language learning, that is something novel in the field of GI. Our conjecture is that contextual information speeds-up the learning process and reduces the amount of examples needed to learn a language.

3.3 Practical Aspects

One of the limitations for researchers working on context is the lack of training data with contextual annotation. A common resource of training data remains conspicuously absent. Moreover, the lack of standardization of the type of data available to researchers makes it very difficult to compare different models with one another.

Our aim is to develop a new benchmark for the international community. More concretely, we will develop datasets combining natural language descriptions with semantically annotated visual events. This resource will be valuable for researchers who study language learning, particular for those who study syntax and semantics together. The datasets will be freely available for use, and the researchers will not only be able to use these data, but also produce new data and share them. Hence, these datasets aim to be collaborative tools, allowing other researchers to add new data.

This database will also allow us to evaluate our algorithms and the ones proposed by other researchers from the international community.

We can distinguish two types of useful datasets:

1) A first goal is to develop a benchmark based on the *Miniature Language Acquisition task* [30]. This task consists of learning a subset of a natural

language from sentence-picture pairs that involve geometric shapes with different properties (color, size and position). Although this task is not as complex as those faced by children, it involves enough complexity to be compared to many real-world tasks.

2) A second goal is to develop a benchmark based on real data. A first perspective is to use a software platform to develop a real syntactic-semantic dataset (in collaboration with linguists). An idea would be to present several pictures to a child, to say a sentence and ask to the child which one better reflects the content of the picture. With that, we would validate our models (we could verify, for example, if our system is able to understand what we are talking about, like the child does).

It is worth noting that our proposal relies on some preliminary works that we have done. In [2–4], it was proposed a simple computational model of language learning. This model takes into account some aspects of natural language acquisition, such as: the interaction between the adult and the child, the context in which the sentences are produced and the meaning-preserving corrections made by the adults to well understand the imperfect sentences of the child. Thus, this model has a learner and a teacher that interact in a sequence of situations by producing sentences that intend to denote one object in each situation. The learner uses cross-situational correspondences to learn to comprehend and produce denoting sentences in a given situation (there is no explicit semantic annotation of the sentences). The goal of the learner is to be able to produce every sentence denoting one object in any given situation. It was empirically showed that the access to semantic information facilitates language learning, and the presence of corrections by the teacher has an effect on language learning by the learner. Following this line of research, a work based on pair-Hidden Markov Models was proposed in [8]. It was showed that, by taking into account semantics, it is possible to accelerate the language learning process. Therefore, these preliminary works constitute the building-block of our proposal.

4 Conclusion

One of the goals of Artificial Intelligence is to enable computers to interact with the real world. To achieve this goal, we need to construct machines that are able to understand, among other things, natural language sentences addressed to them and also to produce meaningful sentences in a given context. This paper presents *ongoing work* in this line of research.

The field of GI provides a good theoretical framework for investigating the process of natural language learning. Taking into account that studies in GI are based exclusively on learning syntax, we have proposed in this paper to exploit syntactic and contextual information for language learning. Hence, our proposal establishes a methodological break up with conventional approaches in GI.

Our work can contribute to the definition and implementation of models that may simplify and improve human-computer interfaces. The results we expect to obtain from our work are not only theoretical or restricted to the field of GI, they are also related to cognitive science, and in general, to the topic of how humans actually process natural language. So, our proposal is an interdisciplinary work and the expected results may be useful in different research areas.

References

1. Angluin, D.: Learning Regular Sets from Queries and Counterexamples. Information and Computation **75**, 87–106 (1987)
2. Angluin, D., Becerra-Bonache, L.: Learning meaning before syntax. In: Clark, A., Coste, F., Miclet, L. (eds.) ICGI 2008. LNCS (LNAI), vol. 5278, pp. 1–14. Springer, Heidelberg (2008)
3. Angluin, D., Becerra-Bonache, L.: A model of Semantics and Corrections in Language Learning. Technical Report YALE/DCS/TR-1425, Computer Science Department, Yale University (2010)
4. Angluin, D., Becerra-Bonache, L.: Effects of meaning-preserving corrections on language learning. In: International Conference on Computational Natural Language Learning (CONLL), pp. 97–105 (2011)
5. Akman, V., Surav, M.: The use of Situation Theory in Context Modeling. Computational Intelligence **13**(3), 427–438 (1997)
6. Barwise, J., Perry, J.: Situations and Attitudes. MIT Press, Cambridge (1983)
7. Barwise, J.: Conditionals and conditional information. In: Traugott, E.C., Ferguson, C.A., Reilly, J.S. (eds.) On Conditionals, pp. 21–54. Cambridge University Press, Cambridge (1986)
8. Becerra-Bonache, L., Fromont, E., Habrard, A., Perrot, M., Sebban, M.: Speeding up Syntactic Learning Using Contextual Information. International Colloquium on Grammatical Inference (ICGI) **21**, 49–53 (2012)
9. Buvač, S.: Quantificational logic of context. In: National Conference on Artificial Intelligence (AAAI), Portland, Oregon, 4–8 August 1996, vol. 1, pp. 600–606 (1996)
10. Buvač, S., Mason, I.: Propositional logic of context. In: National Conference on Artificial Intelligence (AAAI), pp. 412–419 (1993)
11. Buvač, S., Buvač, V., Mason, I.: Metamathematics of Contexts. Fundamenta Informaticae **23**(3), 263–301 (1995)
12. Carrasco, R.C., Oncina, J.: Learning stochastic regular grammars by means of a state merging method. In: International Colloquium on Grammatical Inference (ICGI), pp. 139–152 (1994)
13. Chen, D.L.: Learning Language from Perceptual Context. Department of Computer Sciences, University of Texas at Austin, PhD. proposal (2009)
14. Chouinard, M.M., Clark, E.V.: Adult reformulations of child errors as negative evidence. Journal of Child Language **30**, 637–669 (2003)
15. de la Higuera, C.: Grammatical Inference: Learning Automata and Grammars. Cambridge University Press, Cambridge (2010)
16. Devlin, K.: Logic and information. Cambridge University Press, New York (1991)
17. Ghidini, C., Giunchiglia, F.: Local models semantics, or contextual reasoning = locality + compatibility. Artificial Intelligence **127**(2), 221–259 (2001)
18. Giunchiglia, F.: Contextual reasoning. Epistemologia, XVI, pp. 345–364 (1993)
19. Giunchiglia, F., Serafini, L.: Multilanguage hierarchical logics, or: how we can do without modal logics. Artificial Intelligence **65**(1), 29–70 (1994)

20. Gold, E.M.: Language identification in the limit. Information and Control **10**, 447–474 (1967)
21. Goldwasser, D., Roth, D.: Learning from natural instructions. Machine Learning **94**(2), 205–232 (2014)
22. Guha, R.V., Lenat, D.B.: Cyc: A midterm report. AI Magazine **11**(3), 32–59 (1990)
23. Guha, R.V.: Contexts: A formalization and some applications. Stanford PhD Thesis (1991)
24. MacWhinney, B.: The CHILDES Project: Tools for analyzing talk, 3rd edn. Lawrence Erlbaum Associates, Mahwah (2000)
25. Makarios, S.: A Model Theory for a Quantified Generalized Logic of Contexts. Technical Report KSL-06-08, Knowledge Systems, AI Laboratory (2006)
26. McCarthy, J.: Generality in Artificial. Communication of the ACM **30**(12), 1029–1035 (1987)
27. McCarthy, J.: Notes on Formalizing Context. In: Proceedings of the Thirteenth International Joint Conference in Artificial Intelligence (IJCAI-1993), Chambery, France, pp. 555–560 (1993)
28. Nickles, M., Rettinger, A.: Interactive Relational Reinforcement Learning of Concept Semantics. Machine Learning **94**(2), 169–204 (2014)
29. Oncina, J., García, P.: Identifying regular languages in polynomial time. In: Bunke, H. (ed.) Advances in Structural and Syntactic Pattern Recognition, pp. 99–108. World Scientific Publishing, Singapore (1992)
30. Stolcke, A., Feldman, J.A., Lakoff, G., Weber, S.: Miniature Language Acquisition: A Touchstone for Cognitive Science. Cognitive Science, pp. 686–693 (1994)
31. Surav, M., Akman, V.: Modeling Context with Situations. In: IJCAI-95 Workshop on Modeling Context in Knowledge Representation and Reasoning, Research Report 95/11, LAFORIA, pp. 145–156 (1995)
32. Valiant, L.G.: A Theory of the Learnable. Communication of the ACM **27**(11), 1134–1142 (1984)
33. Vanzo, A., Croce, D., Basili, R.: A context-based model for sentiment analysis in Twitter. In: International Conference on Computational Linguistics (COLING), pp. 2345–2354 (2014)