Grammatical inference: strengths and weaknesses

Petter ERICSON [a] and Anna JONSSON [a]
[a] *Department of Computing Science, Umeå University, Sweden*

1. Introduction

Grammatical inference (GI) refers to the inference of a set of grammatical rules that defines a formal language from queries or data about that language. Most prominently, these grammars are either *regular* or *context-free*, and most often defined on either strings, trees, or graphs – discrete, structured data. Though there are several advantages to this form of machine learning, including generally small, transparent and efficient models, there are also several drawbacks that impede their use outside some very specific contexts, including sensitivity to noise and inference processes that are hard to parallelise. This extended abstract explores some of these strengths and weaknesses.

2. GI techniques

There is a large variety of GI techniques, all sharing the objective of inferring some formal device (typically a grammar or automaton) describing a language. A non-exhaustive list of models include Gold's language identification in the limit [1], Angluin's Minimally Adequate Teacher (MAT) model [2], which has evolved into later techniques such as model learning [3] and many others. For learning of (probabilistic) context-free languages, the two major techniques are the Inside-Outside algorithm [4,5] and various Bayesian techniques [6,7] generally based on either the former, or on Markov chain Monte Carlo methods [8]. Further GI models exist for even more powerful grammar formalisms, up to and including certain subclasses of multiple context-free grammars [9].

2.1. MAT learning

In order to discuss the strengths and weaknesses in a more concrete setting, we introduce one of the techniques in more detail, namely Minimal Adequate Teacher (MAT) learning [2]. MAT learning models grammar acquisition with a *learner* querying a *teacher* who knows the language that is to be learned (generally through access to a grammar or automaton implementing the language). The teacher can respond to two types of queries: (1) Is the element x part of the target language? and (2) Does the automaton/grammar A characterise the target language? If a query of type (2) is asked and the answer is 'no', the teacher additionally provides a *counterexample*, while a 'yes' answer terminates the learning procedure. In the standard implementation, where an automaton is learned, each query results in either a new state or a new transition, gradually introducing a more fine-grained recognition of the target language, eventually resulting in a minimal (deterministic) regular automaton.

3. Strengths

Below, we list the strengths of GI and exemplify them in the context of MAT learning.

Traceability We can trace how the model is learning, where the rules come from and how they are formed based on the data. This allows us to correct potential errors in the algorithm, or reconsider the learning strategy or data used. For example in the MAT model, each state and transition is tied directly to a query to the teacher, and we can moreover trace the various language hypotheses that the learner submits for equivalence queries.

Transparency The finished model is not a black box, but a formal device with a clear path of different decisions leading to its final judgement.

Explainability As a result of the transparency, the model contains all of the information needed to present the reasoning behind a given output to a user. This significantly facilitates the building of a user-friendly explanation system.

Brevity With appropriate advance knowledge, it is possible to learn the target function from a minimal set of examples. In particular, there is a variant of MAT learning that in lieu of equivalence queries have access to a *representative sample* of positive and negative examples, while minimising the amount of membership queries.

4. Weaknesses

Here we repeat the structure of the above section, but this time for GI weaknesses.

Sensitivity Exact and structured models such as automata have generally a harder time to allow for outliers and errors. In particular, MAT learning has guarantees on the maximum amount of queries needed to learn a language described by an automaton of a certain size, but even a single mistake by the oracle can upend those guarantees entirely.

Nuance Another impact of the structured and discrete nature of grammars is that the choice of discretisation is largely unavoidable, which can make extrapolation to completely new data hard or impossible.

Serialisation GI techniques (such as MAT learning) that rely on incremental improvements to a central grammar are hard to parallelise, and thus can not make proper use of the large improvements in multi-core and GPU processing of recent years.

5. Discussion

We hope to expand and clarify on these strengths and weaknesses in collaboration with the other workshop participants, as well as to find ways to combine GI with other complementary ML techniques, as attempted e.g. in [10,11].

References

[1] Gold EM. Language identification in the limit. Information and control. 1967;10(5):447-74.

[2] Angluin D. Learning regular sets from queries and counterexamples. Information and Computation. 1987;75(2):87-106.

[3] Vaandrager F. Model learning. Communications of the ACM. 2017;60(2):86-95.

[4] Baker JK. Trainable grammars for speech recognition. The Journal of the Acoustical Society of America. 1979;65(S1):S132-2.

[5] Lari K, Young SJ. The estimation of stochastic context-free grammars using the inside-outside algorithm. Computer speech & language. 1990;4(1):35-56.

[6] Kurihara K, Sato T. An application of the variational Bayesian approach to probabilistic context-free grammars. In: International Joint Conference on Natural Language Processing Workshop Beyond Shallow Analyses; 2004. .

[7] Snyder B, Naseem T, Barzilay R. Unsupervised Multilingual Grammar Induction. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Association for Computational Linguistics; 2009. p. 73-81.

[8] Johnson M, Griffiths TL, Goldwater S. Bayesian inference for PCFGs via Markov chain Monte Carlo. In: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference; 2007. p. 139-46.

[9] Clark A, Yoshinaka R. Distributional learning of parallel multiple context-free grammars. Machine Learning. 2014;96(1):5-31.

[10] Xu Z, Wen C, Qin S, He M. Extracting automata from neural networks using active learning. PeerJ Computer Science. 2021;7:e436.

[11] Kim Y, Dyer C, Rush AM. Compound probabilistic context-free grammars for grammar induction. arXiv preprint arXiv:190610225. 2019.