Fundamenta Informaticae 146 (2016) i-ii
DOI 10.3233/FI-2016-1390
IOS Press

Grammatical Inference

Preface

Grammatical Inference (GI) is concerned with learning automata, grammars, and other objects that make it possible to generate, derive, represent, or recognize formal languages when given some information about the languages. The algorithms, techniques, and theoretical results which have been obtained in the field are used in a variety of very different applications, including natural language, bioinformatics, psychology, software engineering, and many others.

Grammatical Inference is at a crossing point between a number of fields, influenced by algorithms and results from these. It has applications, but it also has a strong theoretical basis: learnability questions arise, complexity issues are often crucial, bounds concerning the number of examples to be sampled are often strongly linked with similar bounds from the field of randomized algorithms. Furthermore, several of the applications require guarantees on the results returned by GI algorithms.

This importance of theory was what convinced us to propose to publish this special issue in Fundamenta Informaticae. Clearly, many important lines of research followed in the community, closer to the applications, were not going to be targeted. This special issue presents four articles from a total of 10 submissions received, eight of which were extended submissions from the 12th International Conference on Grammatical Inference held at Kyoto University in September, 2014.

The first round of the review process involved 2 to 4 reviewers, different in most cases from those involved during the reviewing process of the ICGI conference. Among the submissions, 4 were co-authored by guest editors of the special issue. These 4 papers were handled separately by the guest editors who were not co-authors of the papers. A second round of reviews took place for a limited number of papers. At the end of this round, 4 papers were selected, closely adhering to the reviewer recommendations.

The special issue includes papers on a variety of topics, showing some of the key theoretical questions under scrutiny in the community.

The first paper titled *Distributional Learning of Some Nonlinear Tree Grammars* is concerned with distributional learning and extending the results obtained in previous work from linear to non-linear structures. The authors, Makoto Kanazawa, Alexander Clark, Gregory M. Kobele and Ryo Yoshinaka follow the long tradition in computational linguistics and the mathematics of language of using *substitutability* to define language.

The second paper is *Grammatical Inference of PCFGs applied to Language Modelling and Unsupervised Parsing* by James Scicluna and Colin de la Higuera. If the learning of regular languages is fairly well understood today and the extension to the stochastic counterparts has now been studied

for more than 20 years, this is not the case for learning probabilistic context-free languages. Yet these are important modelling tools in bio-informatics and in natural language processing. In this work, an algorithm to learn PCFGs is proposed, using constraint solving as one of the algorithmic mechanisms.

In their article *Designing and Learning Substitutable Plane Graph Grammars*, Rémi Eyraud, Jean-Christophe Janodet, Tim Oates, Frédéric Papadopoulos study grammars for graph languages. Again, this is a fairly unexplored domain and one for which, up to now, essentially only statistical techniques involving the optimization of parameters had been investigated. This work concentrates on a special class of graphs (the plane graphs) and provides us with some positive learnability results possibly opening the path to further research in the field.

The fourth and final paper of this special issue concerns transducers. Transducers define bi-languages and are typically used in translation tasks. The state of the art concerning the learnability of transducers was that only a strongly deterministic type of finite state machines were learnable. In the work *A Canonical Semi-Deterministic Transducer*, Achilles Beros and Colin de la Higuera propose a new class of transducers, semi-deterministic, prove the existence of a normal form for these and exhibit an algorithm allowing to learn them from polynomial time and data.

We believe this particular collection of papers is representative of the theoretical part of the field of Grammatical Inference, but more importantly of some of the key challenges to face, and we hope it is of interest to the general readership of Fundamenta Informaticae as well as to specialists in Grammatical Inference.

Lastly, for this special issue, we required the help of more than 30 reviewers which we really would like to thank today. We also want to thank Damian Niwinski for his help, advice and encouragements.

Rémi Eyraud
QARMA team, Marseille, France
remi.eyraud@lif.univ-mrs.fr

Colin de la Higuera
Nantes University, Nantes, France
cdlh@univ-nantes.fr

Makoto Kanazawa
National Institute of Informatics, Tokyo, Japan
kanazawa@nii.ac.jp

Ryo Yoshinaka
Tohoku University, Sendai, Japan
ry@ecei.tohoku.ac.jp

Editors of this special issue