Introduction to the Special Issue on Grammar Induction

P. Adriaans
ILLLC, University of Amsterdam, The Netherlands
H. Fernau
WSI für Informatik, Universität Tübingen, Germany
C. de la Higuera
EURISE, Université de Saint-Etienne, France
M. van Zaanen
ILK, Tilburg University, The Netherlands

1   Introduction

Grammatical Inference, likewise known as Grammar Induction, has traditionally been investigated by many researchers from very distinct fields, such as formal grammars, computational linguistics, formal linguistics, psychology, learning theory, computer science, etc. This has resulted in many interesting results, that unfortunately do not cross-over to the other fields.

More recently, there has been a trend to overcome the boundaries of the separate areas. This has already led to many discussions between researchers from these fields. We expect that in the near future many new results and systems will arise from the work that has been started by researchers from the different fields.

The underlying idea behind the special issue that now lies before you is two-fold. Firstly, we want to show that the problems in Grammatical Inference have been interesting to many people, that the research questions in themselves are not particularly hard to understand and that therefore, the field is quite accessible. Secondly, we aim to bring more researchers with a different research background together. This is already an on-going process. We merely try to aid, facilitate and accelerate this new co-operation, which is already showing nice "synergetic effects."

In this special issue, we tried to combine articles that will be interesting to many people with different backgrounds, selecting three out of nine submissions. Furthermore, we added two introductory chapters, especially aimed at people with a more formal computational background and people coming from the (broad) field of linguistics. We hope you will enjoy reading the articles in this special issue and that they may help your introduction in Grammatical Inference or perhaps give you new ideas for future research or collaboration.

If you got interested in the area, please visit our website `http://eurise.univ-st-etienne.fr/gi/` or come to the biennial conference ICGI (International Conference on Grammatical Inference) which will take place in Athens this year.

2   Articles

2.1   Grammar Induction: An Invitation For Formal Language Theorists

The first introductory article, by Fernau and de la Higuera, introduces the field of Grammatical Inference to experts from formal language theory or complexity theory. They believe that progress will be made by studying more complex classes of grammars and languages, or by insisting on specific algebraic properties that will allow new classes to be learned. They also believe that formal language theorists and grammatical inferentists share objects and techniques and that therefore the effort to move from one field to the other has a good pay-back. Moreover, some specific problems are provided for which complexity theory techniques may yield a solution.

2.2   Computational Grammar Induction for Linguists

The second introductory article, by Adriaans and van Zaanen, initiates linguists who are interested in problems of Grammatical Inference into the more formal aspects of the field. They try to relate the formal computational grammar research to more practical linguistics. They describe ongoing research on the borders of formal language theory and (computational) linguistics. Additionally, they indicate interesting and promising research questions and directions.

2.3   e-GRIDS: Computationally Efficient Grammatical Inference from Positive Examples

In this article, by Petasis *et al.*, a new, computationally efficient algorithm is discussed. Using positive sample sentences, it applies a measure of simplicity to direct the search process. The technique they propose is based on earlier work by Wolff, or Langley and Stromsten. This article shows how theoretical results can be applied to the practical problem of finding context-free grammars.

2.4   Black Box and White Box Identification of Formal Languages using Testsets

Bezem *et al.* study conditions that are necessary to compute a grammar for a given language. Their article shows that these conditions must be quite strong, not only for context-free grammars, but also for regular grammars. They improve some algorithms and also provide new results.

2.5 Learning k-Testable and k-Piecewise Testable Languages from Positive Data

In the line of classic work in structural and syntactic pattern recognition, the authors propose new classes of grammars, parameterized with an integer constant k, measuring some sort of window over the grammars, and obtain positive learning results for these classes.