

Lecture Notes in Artificial Intelligence 6339

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

José M. Sempere Pedro García (Eds.)

Grammatical Inference: Theoretical Results and Applications

10th International Colloquium, ICGI 2010
Valencia, Spain, September 13-16, 2010
Proceedings

 Springer

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada
Jörg Siekmann, University of Saarland, Saarbrücken, Germany
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

José M. Sempere
Universidad Politécnica de Valencia
Departamento de Sistemas Informáticos y Computación
Camino de Vera s/n, 46022 Valencia, Spain
E-mail: jsempere@dsic.upv.es

Pedro García
Universidad Politécnica de Valencia
Departamento de Sistemas Informáticos y Computación
Camino de Vera s/n, 46022 Valencia, Spain
E-mail: pgarcia@dsic.upv.es

Library of Congress Control Number: 2010933123

CR Subject Classification (1998): I.2, F.1, I.4, I.5, J.3, H.3

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743

ISBN-10 3-642-15487-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-15487-4 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2010
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

The first edition of the International Colloquium on Grammatical Inference (ICGI) was held in Essex (United Kingdom) in 1993. After the success of this meeting there have been eight more editions that have been hosted by different academic institutions across the world: Alicante (Spain, 1994), Montpellier (France, 1996), Ames, Iowa (USA, 1998), Lisbon (Portugal, 2000), Amsterdam (The Netherlands, 2002), Athens (Greece, 2004), Tokyo (Japan, 2006) and Saint-Malo (France, 2008). ICGI 2010 was held in Valencia (Spain) during September 13–16. It was organized by the Research Group on Formal Language Theory, Computability and Complexity from the Technical University of Valencia. This was the tenth edition of ICGI, which is a nice number for celebrations. Ten editions is a sign of good health for any conference. In the case of Grammatical Inference, it means that the topics, problems and applications of this research area are alive and serve as a good framework to study related aspects of artificial intelligence, natural language processing, formal language theory, computability and complexity, bioinformatics, pattern recognition, etc.

There were two reviews and local discussions among the members of the Program Committee (PC) in order to evaluate every work proposed to the conference. This volume contains the texts of 32 papers presented at ICGI 2010. They are divided into two groups of works. There are 18 regular papers (out of 25) and 14 short papers (11 out of 15, and three regular papers proposed as short ones). The topics of the papers range from theoretical results about the learning of different formal language classes (regular, context-free, context-sensitive, etc.) to application papers on bioinformatics, language modelling, software engineering, etc. In addition, there are two invited lectures delivered by distinguished scientists on the following topics:

- Simon Lucas (University of Essex, UK): *Grammatical Inference and Games*
- David B. Searls (University of Pennsylvania, USA): *Molecules, Languages, and Automata*

In this edition, for the first time, there was a *Best Student Paper Award* to motivate young researchers in this area to continue their research work. The award was given to Franco Luque for his paper “Bounding the Maximal Parsing Performance of Non-Terminally Separated Grammars.”

The first day of the conference hosted four tutorial talks given by prominent scientists of the area on different aspects of grammatical inference. We are grateful to the tutorial lecturers for the brilliant talks: Tim Oates, with Sourav Mukherjee, Colin de la Higuera, Francois Coste and Damián López, with Pedro García.

We would like to thank the many people who contributed to the success of ICGI 2010. First of all, we are grateful to the members of the Steering Committee

that supported our proposal to organize the conference. It was very exciting to organize ICGI 2010 given that some members of the Local Organizing Committee were involved in the organization of ICGI 1994.

We are very grateful to the members of the PC for their time and effort in carrying out the reviewing process. The help and the experience that they provided were invaluable, and the suggestions that they proposed to improve different aspects of the conference were brilliant. Thanks are given to external reviewers that helped the PC members during the review process: Kengo Sato, Manuel Vázquez de Parga and Damián López. The joint effort of these people ensured the quality of the works presented in this volume.

The success of the conference was possible due to the work of the Local Organizing Committee. We especially thank the effort and work made by Damián López, who was involved in many aspects of the conference. In addition, we received the support of the Centre for Innovation, Research and Technology Transfer (CTT) and the Continuous Training Centre (CFP) of the Technical University of Valencia. We are grateful to the people of such institutions for helping us to carry out different aspects of the organization of the conference.

Last, but not least, we are grateful to the sponsors of the conference: The PASCAL2 Network of Excellence, the Spanish Ministry of Science and Innovation, BANCAJA, and the Technical University of Valencia together with the Department on Information Systems and Computation and the School of Engineering in Computer Science.

We hope to celebrate the next ten editions of ICGI. We are sure that it will have a brilliant and exciting future in this research area that tries to identify and solve many interesting problems before the limit.

June 2010

José M. Sempere
Pedro García

Conference Organization

Program Chair

José M. Sempere Universidad Politécnica de Valencia, Spain

Program Committee

Pieter Adriaans	Universiteit van Amsterdam, The Netherlands
Dana Angluin	Yale University, USA
Jean-Marc Champarnaud	Université de Rouen, France
Alexander Clark	Royal Holloway University of London, UK
Francois Coste	INRIA, France
Colin de la Higuera	Université de Nantes - LINA, France
Francois Denis	Université de Provence, France
Henning Fernau	Universität Trier, Germany
Pedro García	Universidad Politécnica de Valencia, Spain
Makoto Kanazawa	National Institute of Informatics, Japan
Satoshi Kobayashi	University of Electro-Communications, Japan
Laurent Miclet	ENSSAT-Lannion, France
Tim Oates	University of Maryland Baltimore County, USA
Arlindo Oliveira	Lisbon Technical University, Portugal
Jose Oncina	Universidad de Alicante, Spain
Georgios Paliouras	Institute of Informatics Telecommunications, Greece
Yasubumi Sakakibara	Keio University, Japan
Etsuji Tomita	University of Electro-Communications, Japan
Menno van Zaanen	Tilburg University, The Netherlands
Ryo Yoshinaka	Japan Science and Technology Agency, Japan
Sheng Yu	The University of Western Ontario, Canada
Thomas Zeugmann	Hokkaido University, Japan

Local Organization

All members are from the Universidad Politécnica de Valencia, Spain.

Marcelino Campos
Antonio Cano
Damián López
Alfonso Muñoz-Pomer
Piedachu Peris
Manuel Vázquez de Parga

Sponsoring Institutions

The PASCAL2 Network of Excellence

Ministerio de Ciencia e Innovación (Spain)

Universidad Politécnica de Valencia (UPV)

Department of Information Systems and Computation (DSIC, UPV)

School of Engineering in Computer Science (ETSINF, UPV)

BANCAJA

Table of Contents

Invited Talks

Grammatical Inference and Games: Extended Abstract	1
<i>Simon M. Lucas</i>	
Molecules, Languages and Automata	5
<i>David B. Searls</i>	

Regular Papers

Inferring Regular Trace Languages from Positive and Negative Samples	11
<i>Antonio Cano Gómez</i>	
Distributional Learning of Some Context-Free Languages with a Minimally Adequate Teacher	24
<i>Alexander Clark</i>	
Learning Context Free Grammars with the Syntactic Concept Lattice	38
<i>Alexander Clark</i>	
Learning Automata Teams	52
<i>Pedro García, Manuel Vázquez de Parga, Damián López, and José Ruiz</i>	
Exact DFA Identification Using SAT Solvers	66
<i>Marijn J.H. Heule and Sicco Verwer</i>	
Learning Deterministic Finite Automata from Interleaved Strings	80
<i>Joshua Jones and Tim Oates</i>	
Learning Regular Expressions from Representative Examples and Membership Queries	94
<i>Efim Kinber</i>	
Splitting of Learnable Classes	109
<i>Hongyang Li and Frank Stephan</i>	
PAC-Learning Unambiguous k,l -NTS $^{\leq}$ Languages	122
<i>Franco M. Luque and Gabriel Infante-Lopez</i>	
Bounding the Maximal Parsing Performance of Non-Terminally Separated Grammars	135
<i>Franco M. Luque and Gabriel Infante-Lopez</i>	

CGE: A Sequential Learning Algorithm for Mealy Automata	148
<i>Karl Meinke</i>	
Using Grammar Induction to Model Adaptive Behavior of Networks of Collaborative Agents	163
<i>Wico Mulder and Pieter Adriaans</i>	
Transducer Inference by Assembling Specific Languages	178
<i>Piedachu Peris and Damián López</i>	
Sequences Classification by Least General Generalisations	189
<i>Frédéric Tantini, Alain Terlutte, and Fabien Torre</i>	
A Likelihood-Ratio Test for Identifying Probabilistic Deterministic Real-Time Automata from Positive Data	203
<i>Sicco Verwer, Mathijs de Weerdt, and Cees Witteveen</i>	
A Local Search Algorithm for Grammatical Inference	217
<i>Wojciech Wieczorek</i>	
Polynomial-Time Identification of Multiple Context-Free Languages from Positive Data and Membership Queries	230
<i>Ryo Yoshinaka</i>	
Grammatical Inference as Class Discrimination	245
<i>Menno van Zaanen and Tanja Gaustad</i>	
Short Papers	
MDL in the Limit	258
<i>Pieter Adriaans and Wico Mulder</i>	
Grammatical Inference Algorithms in MATLAB	262
<i>Hasan Ibne Akram, Colin de la Higuera, Huang Xiao, and Claudia Eckert</i>	
A Non-deterministic Grammar Inference Algorithm Applied to the Cleavage Site Prediction Problem in Bioinformatics	267
<i>Gloria Inés Alvarez, Jorge Hernán Victoria, Enrique Bravo, and Pedro García</i>	
Learning PDFAs with Asynchronous Transitions	271
<i>Borja Balle, Jorge Castro, and Ricard Gavaldà</i>	
Grammar Inference Technology Applications in Software Engineering	276
<i>Barrett R. Bryant, Marjan Mernik, Dejan Hrnčič, Faizan Javed, Qichao Liu, and Alan Sprague</i>	

Hölder Norms and a Hierarchy Theorem for Parameterized Classes of CCG	280
<i>Christophe Costa Florêncio and Henning Fernau</i>	
Learning of Church-Rosser Tree Rewriting Systems	284
<i>M. Jayasrirani, D.G. Thomas, Atulya K. Nagar, and T. Robinson</i>	
Generalizing over Several Learning Settings	288
<i>Anna Kasprzik</i>	
Rademacher Complexity and Grammar Induction Algorithms: What It May (Not) Tell Us	293
<i>Sophia Katrenko and Menno van Zaanen</i>	
Extracting Shallow Paraphrasing Schemata from Modern Greek Text Using Statistical Significance Testing and Supervised Learning	297
<i>Katia Lida Kermanidis</i>	
Learning Subclasses of Parallel Communicating Grammar Systems	301
<i>Sindhu J. Kumar, P.J. Abisha, and D.G. Thomas</i>	
Enhanced Suffix Arrays as Language Models: Virtual k -Testable Languages	305
<i>Herman Stehouwer and Menno van Zaanen</i>	
Learning Fuzzy Context-Free Grammar—A Preliminary Report	309
<i>Olgierd Unold</i>	
Polynomial Time Identification of Strict Prefix Deterministic Finite State Transducers	313
<i>Mitsuo Wakatsuki and Etsuji Tomita</i>	
Author Index	317