Preface — ICGI 2012

Jeffrey Heinz HEINZ@UDEL.EDU
University of Delaware, USA

Colin de la Higuera CDLH@UNIV-NANTES.FR
University of Nantes, France

Tim Oates OATES@CS.UMBC.EDU
University of Maryland, Baltimore County, USA

Editors: Jeffrey Heinz, Colin de la Higuera and Tim Oates

These proceedings contain the papers accepted to, and presented at, the 11th International Conference on Grammatical Inference (ICGI) which took place from September 5–8, 2012 at the University of Maryland, College Park near Washington, D.C. It is the second time ICGI has been held in the United States of America, and we were very glad to be part of the effort bringing it here.

The conference began with a day of four tutorials. The tutorial program has proven to be a valuable aspect of past ICGIs, introducing attendees to a wide variety of topics related to grammatical inference, and this program was no exception. The tutorials included:

- *Gold-Style Learning Theory: General Highlights Since Gold* by John Case of the University of Delaware. Gold's learning theory is seminal and is a recurring element of talks at every ICGI, so ensuring a common footing among the attendees on the topic was extremely helpful.

- *Distributional learning of context free and mildly context sensitive languages* by Alexander Clark of Royal Holloway, London and Ryo Yoshinaka of Kyoto University. This tutorial covered ground in the Chomsky hierarchy above regular languages, which is where much of the work on grammatical inference is done, showing the power of distributional learning.

- *Learning Probability Distributions Generated by Finite-State Machines* by Ricard Gavaldà and Jorge Castro of UPC, Barcelona. This tutorial nicely complemented the Pautomac competition held at this ICGI on learning probabilistic finite automata and hidden Markov models, preparing attendees for the discussion on the day of the competition workshop.

- *Computational Approaches to Child Language Acquisition* by Shuly Wintner of the University of Haifa. Though work in grammatical inference owes much of its inspiration to human language learning, it is often very formal and abstract. This tutorial returned attendees to one of the roots of the field.

Three experts in fields related to grammatical inference graciously agreed to give invited talks. These talks were:

- *Grammar Induction: Beyond Local Search* by Jason Eisner (Johns Hopkins University). Common methods for inducing probabilistic grammars for natural language face challenges in escaping local optima. Two strategies studied in other fields (operations research and vision) were explored in this talk.

- *Inducing Partially Observable Markov Decision Processes* by Michael Littman (Brown University). Markov Decision Processes are close to probabilistic finite automata, even if they differ by their semantics. In this talk, challenges related to the learnability of these objects were presented.

- *Active Automata Learning: From DFA to Interface Programs and Beyond* by Bernhard Steffen (University of Dortmund). Whereas the well known L^* algorithm was originally used for theoretical reasons, this is no longer the case in software engineering where verification and testing require precise and fast grammatical inference algorithms. This talk explored current work in this area.

The core of the conference, as always, revolves around high quality research done by members of our community. This year we received a total of 28 submissions for long papers, of which twelve were accepted for full presentation ($\sim$43%). Another six of these submissions were accepted as short papers ($\sim$22%). Additionally we received five short paper submissions, three of which were accepted. The talks and papers presented in these proceedings cover a wide range of topics.

The program committee gave the Best Student Paper award to Borja Balle for his contribution "Bootstrapping and Learning PDFA in Data Streams."

Another important component of this year's ICGI was the Probabilistic Automata Learning (Pautomac) Competition. Finite state automata (or machines) are well-known models for characterizing the behavior of systems or processes. They have been used for several decades in computer and software engineering to model, for example, complex behaviors of electronic circuits and software. A nice feature of automaton models is that they are easy to interpret. Unfortunately, in many applications the original design of a system is unknown. That is why learning is important.

Participants were able to test their algorithms on automatically generated training sets from probabilistic finite state automata or hidden Markov Models and get scores on hidden testing data. Despite a storm on the final day of the competition that knocked out power to millions of people on the East coast of the United States, including the servers running the competition, a winner was declared. We are happy to congratulate the winning team of team of Chihiro Shibata and Ryo Yoshinaka whose paper "Marginalizing Out Transition Probabilities for Several Subclasses of PFAs" describes the techniques they used to best the competition.

In summary, ICGI 2012 serves as a forum for the dissemination of the latest and most important research in grammatical inference. It also educates attendees broadly and deeply through the tutorial program and invited talks. We hope that the energy of the conference will continue as the attendees return to their home institutions, ready to tackle new problems that will help drive our field forward.

Organizing a conference like this requires help from a number of people. We would like to thank the program committee for their work on providing excellent reviews of the submitted

paper, the steering committee for their sage advice on the many difficult issues that naturally arise, the Pautomac organizers and scientific committee for making the competition such a success (both logistically and scientifically), and the many other people who helped out along the way. We are also extremely grateful to the different organizations who generously sponsored the event, The European Network of Excellence PASCAL 2 and the University of Maryland Center for Advanced Studies of Language (CASL).

Jeffrey Heinz, Colin de la Higuera, and Tim Oates
ICGI 2012, co-chairs

ICGI Steering Committee

Alexander Clark, François Coste, Henning Fernau, Colin de La Higuera, José Oncina, Yasubumi Sakakibara, José Sempere, Etsuji Tomita, Menno van Zaanen

ICGI 2012 Program Committee

Pieter Adriaans, Dana Angluin, Tom Armstrong, Robert Berwick, John Case, Alexander Clark, François Coste, Colin de la Higuera, Jason Eisner, Rémi Eyraud, Henning Fernau, Pedro García, Jeffrey Heinz, Makoto Kanazawa, Satoshi Kobayashi, Greg Kobele, Laurent Miclet, Tim Oates, José Oncina, Yasubumi Sakakibara, José Sempere, Bernhard Steffen, Etsuji Tomita, Menno van Zaanen, Sicco Verwer, Enrique Vidal, Ryo Yoshinaka, Franǫis Yvon, Thomas Zeugmann

ICGI 2012 External reviewers

Stephanie Antetomaso, Hendrik Blockeel, Peter Habermehl, Emma Hartman, Falk Howar, Malte Isberner, Anna Kasprzik, Hiroshi Sakamoto, James Scicluna, Yasuhiro Tajima, Mitsuo Wakatsuki

Pautomac Organizing Committee

Hasan Ibne Akram, Rémi Eyraud, Jeffrey Heinz, Colin de la Higuera, James Scicluna, Sicco Verwer

Pautomac Scientific Committee

Pieter Adriaans, Dana Angluin, Alexander Clark, Pierre Dupont, Ricard Gavaldà, Colin de la Higuera, Jean-Christophe Janodet, Tim Oates, José Oncina, Menno van Zaanen