Preface

Jane Chandlee   JCHANDLEE@HAVERFORD.EDU
Tri-Co Department of Linguistics, Haverford University

Rémi Eyraud   REMI.EYRAUD@UNIV-ST-ETIENNE.FR
Univ Lyon, UJM-Saint-Etienne, CNRS, Laboratoire Hubert Curien UMR 5516, Saint-Etienne

Jeffrey Heinz   JEFFREY.HEINZ@STONYBROOK.EDU
Institute for Advanced Computational Science and Department of Linguistics, Stony Brook University

Adam Jardine   ADAM.JARDINE@RUTGERS.EDU
Department of Linguistics, Rutgers University

Menno van Zaanen   MENNO.VANZAANEN@NWU.AC.ZA
South African Centre for Digital Language Resources

Editors: Jane Chandlee, Rémi Eyraud, Jeffrey Heinz, Adam Jardine, and Menno van Zaanen

1. Introduction

The Fifteenth International Colloquium on Grammatical Inference (ICGI-2020/21) was held virtually online from August 23 to August 27, 2021. Originally scheduled in August 2020 in New York City, NY, USA, the event was postponed due to the outbreak of the COVID-19 pandemic earlier in 2020, which is continuing to this day.

Grammatical inference is the study of how grammars can be learned, either from raw example data or using more sophisticated frameworks. From a purely scientific standpoint, it brings together work in mathematics and theoretical computer science as they pertain to computational and algorithmic learning theories, recursive data structures, such as sequences and trees, and different kinds of formal devices that process them. From a practical standpoint, applications have either been demonstrated or proposed in the literature in areas like natural language processing and acquisition, computational biology, systems verification, robotic planning and control, information retrieval, text processing, data compression and adaptive intelligent agents, and many othersq.

This edition of ICGI specifically called for submissions that explore the relationship between neural networks and formal grammars in the context of machine learning and grammatical inference. Indeed, recent successes of deep learning offer a great platform for grammatical inference. For instance, by providing new insights about the expressivity of neural networks, allowing comparisons between their classes, extracting formal models from already trained deep machines, and adapting theoretically founded learning algorithms to neural models, our field's contributions are at the core of one of the most challenging current issues in deep learning: the understanding of its underlying processes, a research theme usually named interpretability or explainability. In these proceedings, several works are presented on the subject: they aim at extracting weighted automata or context-free grammars, studying the ability of various neural network learning algorithms to correctly

identify context-free structures, or proposing new recurrent neural network architecture to handle long structured sequences.

These proceedings also contain a wide range of results that push further the state of the art in more classical grammar induction frameworks. Some focus on learning subregular classes from raw sequences, a crucial issue in computational linguistics. Others demonstrate new results in the exact learning paradigm, a framework where the learner interacts with an oracle. The hard question of learning (probabilistic) context-free grammars is tackled by several articles, while a couple of works propose novel genetic algorithms for grammar learning. Applications of grammatical inference are not forgotten, with works on privacy preserving data sets or natural language processing. Finally, while the majority of articles strengthen their approach with theoretical results, a couple of theoretically founded learning studies provide important results for less usual classes of grammars, both of them grounded in the field of logic.

2. Program Committee

We would like to thank the members of the program committee for their careful evaluation of the submissions, the impressive quality of the discussions, and more generally for their unwavering support.

- Leonor Becerra Bonache (Université Jean Monnet, France)
- Johanna Björklund (Umeå University, Sweden)
- Jorge Castro (Universitat Politècnica de Catalunya, Spain)
- Alexander Clark (King's College London, UK)
- François Coste (Rennes University, France)
- Colin De La Higuera (University of Nantes, France)
- Witold Dyrka (Politechnika Wroclawska, Poland)
- Henning Fernau (University of Trier, Germany)
- Annie Foret (Rennes University, France)
- Bob Franck (Yale, USA)
- Matthias Gallé (Naver Labs Europe, France)
- Falk Howar (Clausthal University of Technology, Germany)
- Jean-Christophe Janodet (University of Evry, France)
- Tobias Kappé (Cornell University, USA)
- Aurélien Lemay (Université de Lille, France)
- Karl Meinke (KTH Royal Institute of Technology, Sweden)
- Joshua Moerman (Open Universiteit, the Netherlands)
- Matteo Sammartino (Royal Holloway University of London, University College London, UK)
- Ute Schmid (University of Bamberg, Germany)

- Jose M. Sempere (Universitat Politècnica de València, Spain)
- Chihiro Shibata (Tokyo University of Technology, Japan)
- Alexandra Silva (University College London, UK)
- Bernhard Steffen (TU Dortmund, Germany)
- Herbert Tanner (University of Delaware, USA)
- Etsuji Tomita (The University of Electro-Communications, Japan)
- Olgierd Unold (Wroclaw University of Science and Technology, Poland)
- Frits Vaandrager (Radboud University, the Netherlands)
- Sicco Verwer (Delft University of Technology, the Netherlands)
- Wojciech Wieczorek (University of Bielsko-Biala, Poland)
- Gail Weiss (Technion – Israel Institute of Technology, Israel)
- Ryo Yoshinaka (Tohoku University, Japan)

3. Conference Chairs

- Jane Chandlee, Haverford College, USA
- Rémi Eyraud, Saint-Etienne University, France
- Jeffrey Heinz, Stony Brook University, USA
- Adam Jardine, Rutgers University, USA
- Menno van Zaanen, South African Centre for Digital Language Resources, South Africa