ACL 2016

The 54th Annual Meeting of the Association for Computational Linguistics

Proceedings of the SIGFSM Workshop on Statistical NLP and Weighted Automata

August 12, 2016
Berlin, Germany

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-945626-13-5 / 1-945626-13-5

Preface

The past 20 years have seen a fundamental paradigm shift in the field of automated natural language processing: though long dominated by rule-based techniques, the vast majority of contemporary approaches are now based on statistical models. This trend can be observed not only in traditional tasks such as machine translation or morphological analysis, but also in new research areas such as topic modelling. The reasons for such a paradigm shift can be attributed above all to the steadily growing pool of large, high-quality manually annotated training material as well as the availability of suitable statistical methods and easily accessible implementations thereof.

The purpose of the Workshop on Statistical Natural Language Processing and Weighted Automata (StatFSM) was to bring together researchers interested in statistical natural language processing, automata theory and application. We are pleased to say that the program of the workshop reflected in this proceedings volume met these expectations.

These proceedings contain the contributions presented at the StatFSM workshop held on August 12, 2016 in conjunction with ACL 2016 in Berlin, Germany. The workshop was a meeting of the ACL Special Interest Group on Finite-State Methods (SIGFSM). In total, 12 papers (seven long and five short papers) were submitted to a doubly blind refereeing process. Each paper was reviewed by three members of the program committee, which consisted of highly esteemed researchers from the field of automata theory and applications. In the end, 9 papers were selected for presentation at the workshop, and those contributions are included in this volume. They discuss topics ranging from weighted finite-state transducers to weighted tree automata and to devices on even more complicated structures. In addition, the contributions strike the right balance between purely theoretic foundational investigations as well as considerations and solutions to problems entirely motivated from practice.

The workshop would not have been possible without the help of a lot of support. First, we would like to thank the program committee members for providing their expertise and valuable feedback during the review process. We also want to express our gratitude to JASON EISNER for his inspiring keynote lecture. The constant efforts of the ACL administration and the local organizers made this workshop possible. Last but not least we want to thank the authors and the participants who above all others have made this workshop a success.

Bryan Jurish
Andreas Maletti
Uwe Springmann
Kay-Michael Würzner

Organizers:

Bryan Jurish, Berlin-Brandenburg Academy of Sciences and Humanities, Germany
Andreas Maletti, University of Stuttgart, Germany
Uwe Springmann, Ludwig-Maximilians-Universität München, Germany
Kay-Michael Würzner, Berlin-Brandenburg Academy of Sciences and Humanities, Germany

Program Committee:

Borja Balle, Lancaster University, UK
Francisco Casacuberta, Instituto Tecnológico de Informática, Spain
Simon Clematide, University of Zurich, Switzerland
Gregory Crane, University of Leipzig, Germany
Frank Drewes, Umeå University, Sweden
Jason Eisner, Johns Hopkins University, Baltimore, MD, USA
Colin de la Higuera, Nantes University, France
Mans Hulden, University of Colorado, Boulder, CO, USA
Krister Lindén, University of Helsinki, Finnland
Kevin Knight, University of Southern California, CA, USA
Marcus Eichenberger-Liwicki, University of Kaiserslautern, Germany
Stoyan Mihov, Bulgarian Academy of Sciences, Sofia, Bulgaria
Mark-Jan Nederhof, University of St Andrews, UK
Michael Riley, Google Inc., USA
Martin Reynaert, Tilburg University, The Netherlands
Brian Roark, Google Inc., USA
Richard Sproat, Google Inc., USA
Heiko Vogler, Dresden University of Technology, Germany
Bruce Watson, Stellenbosch University, South Africa

Invited Speaker:

Jason Eisner, Johns Hopkins University, Baltimore, MD, USA

Table of Contents

Conference Program

Friday, August 12, 2016

09:00–10:30 Welcome and keynote

09:00–09:30 *Opening*
Organizers

09:30–10:30 *Probabilistic models of related strings*
Jason Eisner

10:30–11:00 Morning break

11:00–12:30 Weighted tree automata

11:00–11:30 *Equivalences between ranked and unranked weighted tree automata via binarization*
Toni Dietze

11:30–12:00 *Adaptive importance sampling from finite state automata*
Christoph Teichmann, Kasimir Wansing and Alexander Koller

12:00–12:30 *Transition-based dependency parsing as latent-variable constituent parsing*
Mark-Jan Nederhof

12:30–14:00 Lunch break

Friday, August 12, 2016 (continued)

14:00–15:30 Weighted finite-state transducers

14:00–14:30 *Distributed representation and estimation of WFST-based n-gram models*
Cyril Allauzen, Michael Riley and Brian Roark

14:30–15:00 *Learning transducer models for morphological analysis from example inflections*
Markus Forsberg and Mans Hulden

15:00–15:30 *Data-driven spelling correction using weighted finite-state methods*
Miikka Silfverberg, Pekka Kauppinen and Krister Lindén

15:30–16:00 Afternoon break

16:00–17:30 Various weighted automata

16:00–16:30 *EM-training for weighted aligned hypergraph bimorphisms*
Frank Drewes, Kilian Gebhardt and Heiko Vogler

16:30–17:00 *On the correspondence between compositional matrix-space models of language and weighted automata*
Shima Asaadi and Sebastian Rudolph

17:00–17:30 *Pynini: A Python library for weighted finite-state grammar compilation*
Kyle Gorman