CoNLL-2011

Fifteenth Conference on Computational Natural Language Learning

Proceedings of the Conference

23-24 June, 2011
Portland, Oregon, USA

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53704 USA

CoNLL 2011 Best Paper sponsor:

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 978-1-932432-92-3

Preface

The 2011 Conference on Computational Natural Language Learning is the fifteenth in the series of annual meetings organized by SIGNLL, the ACL special interest group on natural language learning. CONLL-2011 will be held in Portland, Oregon, USA, June 23-24 2011, in conjunction with ACL-HLT.

For our special focus this year in the main session of CoNLL, we invited papers relating to massive, linked text data. We received 82 submissions on these and other relevant topics, of which 4 were eventually withdrawn. Of the remaining 78 papers, 13 were selected to appear in the conference program as oral presentations, and 14 were chosen as posters. All accepted papers appear here in the proceedings. Each accepted paper was allowed eight content pages plus any number of pages containing only bibliographic references.

As in previous years, CoNLL-2011 has a shared task, *Modeling unrestricted coreference in OntoNotes*. The Shared Task papers are collected in a companion volume of CoNLL-2011.

We begin by thanking all of the authors who submitted their work to CoNLL-2011, as well as the program committee for helping us select from among the many strong submissions. We are also grateful to our invited speakers, Bruce Hayes and Yee Whye Teh, who graciously agreed to give talks at CoNLL. Special thanks to the SIGNLL board members, Lluís Màrquez and Joakim Nivre, for their valuable advice and assistance in putting together this year's program, and to the SIGNLL information officer, Erik Tjong Kim Sang, for publicity and maintaining the CoNLL-2011 web page. We also appreciate the additional help we received from the ACL program chairs, workshop chairs, and publication chairs.

Finally, many thanks to Google for sponsoring the best paper award at CoNLL-2011.

We hope you enjoy the conference!

Sharon Goldwater and Christopher Manning

CoNLL 2011 Conference Chairs

Program Chairs:

Sharon Goldwater (University of Edinburgh, United Kingdom)
Christopher Manning (Stanford University, United States)

Program Committee:

Steven Abney (University of Michigan, United States)
Eneko Agirre (University of the Basque Country, Spain)
Afra Alishahi (Saarland University, Germany)
Lourdes Araujo (Universidad Nacional de Educación a Distancia, Spain)
Jason Baldridge (University of Texas at Austin, United States)
Steven Bethard (Katholieke Universiteit Leuven, Belgium)
Steven Bird (University of Melbourne, Australia)
Phil Blunsom (University of Oxford, United Kingdom)
Thorsten Brants (Google Inc., United States)
Chris Brew (Ohio State University, United States)
David Burkett (University of California at Berkeley, United States)
Yunbo Cao (Microsoft Research Asia, China)
Xavier Carreras (Technical University of Catalonia, Spain)
Nathanael Chambers (Stanford University, United States)
Ming-Wei Chang (University of Illinois at Urbana-Champaign, United States)
Colin Cherry (National Research Council, Canada)
Massimiliano Ciaramita (Google Research, Switzerland)
Alexander Clark (Royal Holloway, University of London, United Kingdom)
Stephen Clark (University of Cambridge, United Kingdom)
Shay Cohen (Carnegie Mellon University, United States)
Trevor Cohn (University of Sheffield, United Kingdom)
James Curran (University of Sydney, Australia)
Walter Daelemans (University of Antwerp, Belgium)
Mark Dras (Macquarie University, Australia)
Amit Dubey (University of Edinburgh, United Kingdom)
Chris Dyer (Carnegie Mellon University, United States)
Jacob Eisenstein (Carnegie Mellon University, United States)
Micha Elsner (University of Edinburgh, United Kingdom)
Jenny Finkel (Columbia University, United States)
Radu Florian (IBM Watson Research Center, United States)
Robert Frank (Yale University, United States)
Stella Frank (University of Edinburgh, United Kingdom)
Michel Galley (Microsoft Research, United States)
Kevin Gimpel (Carnegie Mellon University, United States)
Yoav Goldberg (Ben Gurion University of the Negev, Israel)
Cyril Goutte (National Research Council, Canada)

Spence Green (Stanford University, United States)
Gholamreza Haffari (BC Cancer Research Center, Canada)
Keith Hall (Google Research, Switzerland)
James Henderson (University of Geneva, Switzerland)
Julia Hockenmaier (University of Illinois at Urbana-Champaign, United States)
Fei Huang (IBM Research, United States)
Rebecca Hwa (University of Pittsburgh, United States)
Richard Johansson (University of Trento, Italy)
Mark Johnson (Macquarie University, Australia)
Rohit Kate (University of Wisconsin at Milwaukee, United States)
Philipp Koehn (University of Edinburgh, United Kingdom)
Mamoru Komachi (Nara Institute of Science and Technology, Japan)
Terry Koo (Google Inc., United States)
Shankar Kumar (Google Inc., United States)
Tom Kwiatkowski (University of Edinburgh, United Kingdom)
Mirella Lapata (University of Edinburgh, United Kingdom)
Shalom Lappin (Kings College London, United Kingdom)
Lillian Lee (Cornell Universiy, United States)
Percy Liang (University of California at Berkeley, United States)
Adam Lopez (Johns Hopkins University, United States)
Rob Malouf (San Diego State University, United States)
André Martins (Carnegie Mellon University, United States)
Yuji Matsumoto (Nara Institute of Science and Technology, Japan)
Takuya Matsuzaki (University of Tokyo, Japan)
David McClosky (Stanford University, United States)
Ryan McDonald (Google Inc., United States)
Paola Merlo (University of Geneva, Switzerland)
Haitao Mi (Institute of Computing Technology, Chinese Academy of Sciences, China)
Yusuke Miyao (University of Tokyo, Japan)
Alessandro Moschitti (University of Trento, Italy)
Lluís Màrquez (Technical University of Catalonia, Spain)
Tahira Naseem (Massachusetts Institute of Technology, United States)
Mark-Jan Nederhof (University of St. Andrews, United Kingdom)
Hwee Tou Ng (National University of Singapore, Singapore)
Vincent Ng (University of Texas at Dallas, United States)
Joakim Nivre (Uppsala University, Sweden)
Miles Osborne (University of Edinburgh, United Kingdom)
Christopher Parisien (University of Toronto, Canada)
Amy Perfors (University of Adelaide, Australia)
Slav Petrov (Google Research, United States)
Hoifung Poon (University of Washington, United States)
Vasin Punyakanok (BBN Technologies, United States)
Chris Quirk (Microsoft Research, United States)
Ari Rappoport (The Hebrew University, Israel)
Lev Ratinov (University of Illinois at Urbana-Champaign, United States)
Roi Reichart (Massachusetts Institute of Technology, United States)

Joseph Reisinger (University of Texas at Austin, United States)
Sebastian Riedel (University of Massachusetts, United States)
Dan Roth (University of Illinois at Urbana-Champaign, United States)
William Sakas (Hunter College, United States)
Anoop Sarkar (Simon Fraser University, Canada)
William Schuler (The Ohio State University, United States)
Libin Shen (Akamai, United States)
Khalil Sima'an (University of Amsterdam, Netherlands)
Noah Smith (Carnegie Mellon University, United States)
Benjamin Snyder (University of Wisconsin-Madison, United States Richard Socher (Stanford University, United States)
Valentin Spitkovsky (Stanford University, United States)
Mark Steedman (University of Edinburgh, United Kingdom)
Mihai Surdeanu (Stanford University, United States)
Jun Suzuki (NTT Communication Science Laboratories, Japan)
Hiroya Takamura (Tokyo Institute of Technology, Japan)
Ivan Titov (Saarland University, Germany)
Kristina Toutanova (Microsoft Research, United States)
Antal van den Bosch (Tilburg University, Netherlands)
Theresa Wilson (Johns Hopkins University, United States)
Peng Xu (Google Inc., United States)
Charles Yang (University of Pennsylvania, United States)
Chen Yu (Indiana University, United States)
Daniel Zeman (Charles University in Prague, Czech Republic)
Luke Zettlemoyer (University of Washington at Seattle, United States)

Invited Speakers:

Bruce Hayes (University of California, Los Angeles, United States)
Yee Whye Teh (Gatsby Unit, University College London, United Kingdom)

Table of Contents

Conference Program

Thursday, June 23, 2011

9:00–9:05 Opening Remarks

Session 1

9:05–9:30 *Modeling Syntactic Context Improves Morphological Segmentation*
Yoong Keok Lee, Aria Haghighi and Regina Barzilay

9:30–9:55 *The Effect of Automatic Tokenization, Vocalization, Stemming, and POS Tagging on Arabic Dependency Parsing*
Emad Mohamed

9:55–10:20 *Punctuation: Making a Point in Unsupervised Dependency Parsing*
Valentin I. Spitkovsky, Hiyan Alshawi and Daniel Jurafsky

10:20–10:50 Coffee Break

Session 2

10:50–11:15 *Modeling Infant Word Segmentation*
Constantine Lignos

11:15–11:40 *Word Segmentation as General Chunking*
Daniel Hewlett and Paul Cohen

11:40–12:40 *(Invited talk) Computational Linguistics for Studying Language in People: Principles, Applications and Research Problems*
Bruce Hayes

12:40–14:00 Lunch Break

Thursday, June 23, 2011 (continued)

Session 3

14:00–14:25 *Search-based Structured Prediction applied to Biomedical Event Extraction*
Andreas Vlachos and Mark Craven

14:25–14:50 *Using Sequence Kernels to Identify Opinion Entities in Urdu*
Smruthi Mukund, Debanjan Ghosh and Rohini Srihari

14:50–15:15 *Subword and Spatiotemporal Models for Identifying Actionable Information in Haitian Kreyol*
Robert Munro

15:15–15:40 *Gender Attribution: Tracing Stylometric Evidence Beyond Topic and Genre*
Ruchita Sarawgi, Kailash Gajulapalli and Yejin Choi

15:40–16:10 Coffee Break

16:10–17:45 Main Session Posters

Improving the Impact of Subjectivity Word Sense Disambiguation on Contextual Opinion Analysis
Cem Akkaya, Janyce Wiebe, Alexander Conrad and Rada Mihalcea

Effects of Meaning-Preserving Corrections on Language Learning
Dana Angluin and Leonor Becerra-Bonache

Assessing Benefit from Feature Feedback in Active Learning for Text Classification
Shilpa Arora and Eric Nyberg

ULISSE: an Unsupervised Algorithm for Detecting Reliable Dependency Parses
Felice Dell'Orletta, Giulia Venturi and Simonetta Montemagni

Language Models as Representations for Weakly Supervised NLP Tasks
Fei Huang, Alexander Yates, Arun Ahuja and Doug Downey

Automatic Keyphrase Extraction by Bridging Vocabulary Gap
Zhiyuan Liu, Xinxiong Chen, Yabin Zheng and Maosong Sun

Thursday, June 23, 2011 (continued)

Using Second-order Vectors in a Knowledge-based Method for Acronym Disambiguation
Bridget T. McInnes, Ted Pedersen, Ying Liu, Serguei V. Pakhomov and Genevieve B. Melton

Using the Mutual k-Nearest Neighbor Graphs for Semi-supervised Classification on Natural Language Data
Kohei Ozaki, Masashi Shimbo, Mamoru Komachi and Yuji Matsumoto

Automatically Building Training Examples for Entity Extraction
Marco Pennacchiotti and Patrick Pantel

Probabilistic Word Alignment under the L_0-norm
Thomas Schoenemann

Authorship Attribution with Latent Dirichlet Allocation
Yanir Seroussi, Ingrid Zukerman and Fabian Bohnert

Evaluating a Semantic Network Automatically Constructed from Lexical Co-occurrence on a Word Sense Disambiguation Task
Sean Szumlanski and Fernando Gomez

Filling the Gap: Semi-Supervised Learning for Opinion Detection Across Domains
Ning Yu and Sandra Kübler

A Normalized-Cut Alignment Model for Mapping Hierarchical Semantic Structures onto Spoken Documents
Xiaodan Zhu

Friday, June 24, 2011

Shared Task on Modeling Unrestricted Coreference in OntoNotes

9:00–10:30 Shared Task Overview and Oral Presentations

10:30–11:00 Coffee Break

11:00–12:30 Shared Task Posters

12:30–14:00 Lunch Break

14:00–15:00 *(Invited talk) Bayesian Tools for Natural Language Learning*
Yee Whye Teh

15:00–15:30 SIGNLL Business Meeting

15:30–16:00 Coffee Break

Session 4

16:00–16:25 *Composing Simple Image Descriptions using Web-scale N-grams*
Siming Li, Girish Kulkarni, Tamara L. Berg, Alexander C. Berg and Yejin Choi

16:25–16:50 *Adapting Text instead of the Model: An Open Domain Approach*
Gourab Kundu and Dan Roth

16:50–17:15 *Learning with Lookahead: Can History-Based Models Rival Globally Optimized Models?*
Yoshimasa Tsuruoka, Yusuke Miyao and Jun'ichi Kazama

17:15–17:40 *Learning Discriminative Projections for Text Similarity Measures*
Wen-tau Yih, Kristina Toutanova, John C. Platt and Christopher Meek

17:40–17:45 Best Paper Award and Closing