CoNLL-2010

Fourteenth Conference on Computational Natural Language Learning

Proceedings of the Conference

15-16 July 2010
Uppsala University
Uppsala, Sweden

Production and Manufacturing by
Taberg Media Group AB
Box 94, 562 02 Taberg
Sweden

CoNLL-2010 Best Paper Sponsors:

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 978-1-932432-83-1 / 1-932432-83-3

Introduction

The 2010 Conference on Computational Natural Language Learning is the fourteenth in the series of annual meetings organized by SIGNLL, the ACL special interest group on natural language learning. CONLL-2010 will be held in Uppsala, Sweden, 15-16 July 2010, in conjunction with ACL.

For our special focus this year in the main session of CoNLL, we invited papers relating to grammar induction, from a machine learning, natural language engineering and cognitive perspective. We received 99 submissions on these and other relevant topics, of which 18 were eventually withdrawn. Of the remaining 81 papers, 12 were selected to appear in the conference programme as oral presentations, and 13 were chosen as posters. All accepted papers appear here in the proceedings. Following the ACL 2010 policy we allowed an extra page in the camera ready paper for authors to incorporate reviewer comments, so each accepted paper was allowed to have nine pages plus any number of pages containing only bibliographic references.

As in previous years, CoNLL-2010 has a shared task, *Learning to detect hedges and their scope in natural language text*. The Shared Task papers are collected into an accompanying volume of CoNLL-2010.

First and foremost, we would like to thank the authors who submitted their work to CoNLL-2010. We are grateful to our invited speakers, Lillian Lee and Zoubin Ghahramani, 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 each step of the way, and Erik Tjong Kim Sang, who acted as the information officer and maintained the CoNLL-2010 web page.

We also appreciate the help we received from the ACL programme chairs, especially Stephen Clark. The help of the ACL 2010 publication chairs, Jing-Shin Chang and Philipp Koehn, technical support by Rich Gerber from softconf.com, as well as input from Priscilla Rasmussen was invaluable in producing these proceedings.

Finally, many thanks to Google for sponsoring the best paper award at CoNLL-2010.

We hope you enjoy the conference!

Mirella Lapata and Anoop Sarkar

CoNLL 2010 Conference Chairs

Program Chairs

Mirella Lapata (University of Edinburgh, United Kingdom)
Anoop Sarkar (Simon Fraser University, Canada)

Program Committee:

Steven Abney (University of Michigan, United States)
Eneko Agirre (University of the Basque Country, Spain)
Afra Alishahi (Saarland University, Germany)
Jason Baldridge (The University of Texas at Austin, United States)
Tim Baldwin (University of Melbourne, Australia)
Regina Barzilay (Massachusetts Institute of Technology, United States)
Phil Blunsom (University of Oxford, United Kingdom)
Thorsten Brants (Google Inc., United States)
Chris Brew (Ohio State University, United States)
Nicola Cancedda (Xerox Research Centre Europe, France)
Yunbo Cao (Microsoft Research Asia, China)
Xavier Carreras (Technical University of Catalonia, Spain)
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)
James Clarke (University of Illinois at Urbana-Champaign, United States)
Walter Daelemans (University of Antwerp, Netherlands)
Vera Demberg (University of Edinburgh, United Kingdom)
Amit Dubey (University of Edinburgh, United Kingdom)
Chris Dyer (Carnegie Mellon University, United States)
Jenny Finkel (Stanford University, United States)
Radu Florian (IBM Watson Research Center, United States)
Robert Frank (Yale University, United States)
Michel Galley (Stanford University, United States)
Yoav Goldberg (Ben Gurion University of the Negev, Israel)
Cyril Goutte (National Research Council, Canada)
Gholamreza Haffari (University of British Columbia, Canada)
Keith Hall (Google Research, Switzerland)
Marti Hearst (University of California at Berkeley, United States)
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 (The University of Texas at Austin, United States)
Frank Keller (University of Edinburgh, United Kingdom)
Philipp Koehn (University of Edinburgh, United Kingdom)
Terry Koo (Massachusetts Institute of Technology, United States)

Shankar Kumar (Google Inc., United States)
Shalom Lappin (Kings College London, United Kingdom)
Adam Lopez (University of Edinburgh, United Kingdom)
Rob Malouf (San Diego State University, United States)
Yuji Matsumoto (Nara Institute of Science and Technology, Japan)
Takuya Matsuzaki (University of Tokyo, Japan)
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)
Raymond Mooney (University of Texas at Austin, United States)
Alessandro Moschitti (University of Trento, Italy)
Gabriele Musillo (FBK-IRST, Italy)
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)
Grace Ngai (Hong Kong Polytechnic University, China)
Joakim Nivre (Uppsala University, Sweden)
Franz Och (Google Inc., United States)
Miles Osborne (University of Edinburgh, United Kingdom)
Christopher Parisien (University of Toronto, Canada)
Slav Petrov (Google Research, United States)
Hoifung Poon (University of Washington, United States)
David Powers (Flinders University of South Australia, Australia)
Vasin Punyakanok (BBN Technologies, United States)
Chris Quirk (Microsoft Research, United States)
Lev Ratinov (University of Illinois at Urbana-Champaign, United States)
Roi Reichart (The Hebrew University, Israel)
Sebastian Riedel (University of Massachusetts, United States)
Ellen Riloff (University of Utah, United States)
Brian Roark (Oregon Health & Science University, United States)
Dan Roth (University of Illinois at Urbana-Champaign, United States)
William Sakas (Hunter College, United States)
William Schuler (The Ohio State University, United States)
Sabine Schulte im Walde (University of Stuttgart, Germany)
Libin Shen (BBN Technologies, United States)
Benjamin Snyder (Massachusetts Institute of Technology, United States)
Richard Sproat (Oregon Health & Science University, United States)
Mark Steedman (University of Edinburgh, United Kingdom)
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)
Peng Xu (Google Inc., United States)
Charles Yang (University of Pennsylvania, United States)
Daniel Zeman (Charles University in Prague, Czech Republic)
Luke Zettlemoyer (University of Washington at Seattle, United States)

Invited Speakers:

Zoubin Ghahramani, University of Cambridge and Carnegie Mellon University
Lillian Lee, Cornell University

Table of Contents

Conference Program

Thursday, July 15, 2010

9:00–9:15 Opening Remarks

Session 1: Parsing (9:15–10:30)

9:15–9:40 *Improvements in Unsupervised Co-Occurrence Based Parsing*
Christian Hänig

9:40–10:05 *Viterbi Training Improves Unsupervised Dependency Parsing*
Valentin I. Spitkovsky, Hiyan Alshawi, Daniel Jurafsky and Christopher D. Manning

10:05–10:30 *Driving Semantic Parsing from the World's Response*
James Clarke, Dan Goldwasser, Ming-Wei Chang and Dan Roth

10:30–11:00 Break

Session 2: Grammar Induction (11:00–12:15)

11:00–11:25 *Efficient, Correct, Unsupervised Learning for Context-Sensitive Languages*
Alexander Clark

11:25–11:50 *Identifying Patterns for Unsupervised Grammar Induction*
Jesús Santamaría and Lourdes Araujo

11:50–12:15 *Learning Better Monolingual Models with Unannotated Bilingual Text*
David Burkett, Slav Petrov, John Blitzer and Dan Klein

12:15–14:15 Lunch

14:15–15:30 *(Invited Talk) Clueless: Explorations in Unsupervised, Knowledge-Lean Extraction of Lexical-Semantic Information*
Lillian Lee

15:30–16:00 Break

Thursday, July 15, 2010 (continued)

CoNLL 2010 Shared Task, Overview and Oral Presentations (16:00–17:30)

16:00–16:20 *The CoNLL 2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text*
Richárd Farkas, Veronika Vincze, György Móra, János Csirik and György Szarvas

16:20–16:30 *A Cascade Method for Detecting Hedges and their Scope in Natural Language Text*
Buzhou Tang, Xiaolong Wang, Xuan Wang, Bo Yuan and Shixi Fan

16:30–16:40 *Detecting Speculative Language using Syntactic Dependencies and Logistic Regression*
Andreas Vlachos and Mark Craven

16:40–16:50 *A Hedgehop over a Max-margin Framework using Hedge Cues*
Maria Georgescul

16:50–17:00 *Detecting Hedge Cues and their Scopes with Average Perceptron*
Feng Ji, Xipeng Qiu and Xuanjing Huang

17:00–17:10 *Memory-based Resolution of In-sentence Scopes of Hedge Cues*
Roser Morante, Vincent Van Asch and Walter Daelemans

17:10–17:20 *Resolving Speculation: MaxEnt Cue Classification and Dependency-Based Scope Rules*
Erik Velldal, Lilja Øvrelid and Stephan Oepen

17:20–17:30 *Combining Manual Rules and Supervised Learning for Hedge Cue and Scope Detection*
Marek Rei and Ted Briscoe

Shared Task Discussion Panel (17:30–18:00)

Friday, July 16, 2010

9:15–10:30 *(Invited Talk) Bayesian Hidden Markov Models and Extensions*
Zoubin Ghahramani

10:30–11:00 Break

Joint Poster Session: Main conference and shared task posters (11:00–12:30)

11:00–12:30 Main conference posters

Improved Unsupervised POS Induction Using Intrinsic Clustering Quality and a Zipfian Constraint
Roi Reichart, Raanan Fattal and Ari Rappoport

Syntactic and Semantic Structure for Opinion Expression Detection
Richard Johansson and Alessandro Moschitti

Type Level Clustering Evaluation: New Measures and a POS Induction Case Study
Roi Reichart, Omri Abend and Ari Rappoport

Recession Segmentation: Simpler Online Word Segmentation Using Limited Resources
Constantine Lignos and Charles Yang

Computing Optimal Alignments for the IBM-3 Translation Model
Thomas Schoenemann

Semi-Supervised Recognition of Sarcasm in Twitter and Amazon
Dmitry Davidov, Oren Tsur and Ari Rappoport

Learning Probabilistic Synchronous CFGs for Phrase-Based Translation
Markos Mylonakis and Khalil Sima'an

A Semi-Supervised Batch-Mode Active Learning Strategy for Improved Statistical Machine Translation
Sankaranarayanan Ananthakrishnan, Rohit Prasad, David Stallard and Prem Natarajan

Improving Word Alignment by Semi-Supervised Ensemble
Shujian Huang, Kangxi Li, Xinyu Dai and Jiajun Chen

Friday, July 16, 2010 (continued)

A Comparative Study of Bayesian Models for Unsupervised Sentiment Detection
Chenghua Lin, Yulan He and Richard Everson

A Hybrid Approach to Emotional Sentence Polarity and Intensity Classification
Jorge Carrillo de Albornoz, Laura Plaza and Pablo Gervás

Cross-Caption Coreference Resolution for Automatic Image Understanding
Micah Hodosh, Peter Young, Cyrus Rashtchian and Julia Hockenmaier

Improved Natural Language Learning via Variance-Regularization Support Vector Machines
Shane Bergsma, Dekang Lin and Dale Schuurmans

11:00–12:30 Shared Task posters: Systems for Shared Task 1 and 2

Hedge Detection using the RelHunter Approach
Eraldo Fernandes, Carlos Crestana and Ruy Milidiú

A High-Precision Approach to Detecting Hedges and Their Scopes
Halil Kilicoglu and Sabine Bergler

Exploiting Rich Features for Detecting Hedges and Their Scope
Xinxin Li, Jianping Shen, Xiang Gao and Xuan Wang

Uncertainty Detection as Approximate Max-Margin Sequence Labelling
Oscar Täckström, Sumithra Velupillai, Martin Hassel, Gunnar Eriksson, Hercules Dalianis and Jussi Karlgren

Hedge Detection and Scope Finding by Sequence Labeling with Procedural Feature Selection
Shaodian Zhang, Hai Zhao, Guodong Zhou and Bao-liang Lu

Learning to Detect Hedges and their Scope using CRF
Qi Zhao, Chengjie Sun, Bingquan Liu and Yong Cheng

Exploiting Multi-Features to Detect Hedges and Their Scope in Biomedical Texts
Huiwei Zhou, Xiaoyan Li, Degen Huang, Zezhong Li and Yuansheng Yang

Friday, July 16, 2010 (continued)

11:00–12:30 Shared Task posters: Systems for Shared Task 1

A Lucene and Maximum Entropy Model Based Hedge Detection System
Lin Chen and Barbara Di Eugenio

HedgeHunter: A System for Hedge Detection and Uncertainty Classification
David Clausen

Uncertainty Learning using SVMs and CRFs
Vinodkumar Prabhakaran

Exploiting CCG Structures with Tree Kernels for Speculation Detection
Liliana Paola Mamani Sanchez, Baoli Li and Carl Vogel

Features for Detecting Hedge Cues
Nobuyuki Shimizu and Hiroshi Nakagawa

A Simple Ensemble Method for Hedge Identification
Ferenc Szidarovszky, Illés Solt and Domonkos Tikk

A Baseline Approach for Detecting Sentences Containing Uncertainty
Erik Tjong Kim Sang

Hedge Classification with Syntactic Dependency Features based on an Ensemble Classifier
Yi Zheng, Qifeng Dai, Qiming Luo and Enhong Chen

12:30–14:00 Lunch

Session 3: Semantics and Information Extraction (14:00–15:15)

14:00–14:25 *Online Entropy-Based Model of Lexical Category Acquisition*
Grzegorz Chrupała and Afra Alishahi

14:25–14:50 *Tagging and Linking Web Forum Posts*
Su Nam Kim, Li Wang and Timothy Baldwin

14:50–15:15 *Joint Entity and Relation Extraction Using Card-Pyramid Parsing*
Rohit Kate and Raymond Mooney

15:30–16:00 Break

Friday, July 16, 2010 (continued)

Session 4: Machine learning (16:00–17:15)

16:00–16:25 *Distributed Asynchronous Online Learning for Natural Language Processing*
Kevin Gimpel, Dipanjan Das and Noah A. Smith

16:25–16:50 *On Reverse Feature Engineering of Syntactic Tree Kernels*
Daniele Pighin and Alessandro Moschitti

16:50–17:15 *Inspecting the Structural Biases of Dependency Parsing Algorithms*
Yoav Goldberg and Michael Elhadad

Closing Session (17:15–17:45)

17:15–17:45 SIGNLL Business Meeting and Best Paper Award