EMNLP 2011

Workshop on Unsupervised Learning in NLP

Proceedings of the Workshop

July 30, 2011
Edinburgh, Scotland, UK

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 978-1-937284-13-8 / 1-937284-13-1

Introduction

The rapid growth in the amount of computer-readable text in different languages, along with ever developing computational resources, raise much interest in fully automated algorithms for analyzing massive amounts of plain text without using any manually provided input. In addition to obviating the need for costly manual annotation, this line of research gives rise to exciting theoretical questions, exploring what information can be extracted purely by distributional analysis, and characterizing the theoretical significance of the output of such an automatic analysis.

Unsupervised learning is the main approach in NLP for addressing this challenge. Although this approach has grown in popularity over the past years and increasingly sophisticated methodology has been introduced, several fundamental challenges remain which need to be resolved and which cannot be effectively discussed in major conferences. This workshop aims to bridge this gap, by summarizing what has been achieved so far in unsupervised learning in NLP, by fostering discussions on these fundamental issues, and by discussing future trends.

The workshop encourages discussion on topics such as evaluation of unsupervised algorithms, comparison of different algorithmic approaches, and unsupervised learning across multiple languages. Our invited talk by Sharon Goldwater discusses the role unsupervised learning can play on shedding light on human cognition. The workshop program also includes papers that address unsupervised approaches for a broad variety of NLP tasks, ranging from syntactic parsing to lexical semantics. Finally, the workshop holds a panel discussion for exchanging ideas between leading researchers in the area, in order to gain some insight into how to best tackle the current big challenges in unsupervised NLP.

It is our hope that this workshop will provide a better understanding of this research area, and will initiate a series of workshops devoted to this important topic.

Omri Abend, Anna Korhonen, Ari Rappoport and Roi Reichart
UNSUP 2011 Organizers

Organizers:

Omri Abend (Hebrew University of Jerusalem, Israel)
Anna Korhonen (University of Cambridge, UK)
Ari Rappoport (Hebrew University of Jerusalem, Israel)
Roi Reichart (MIT, USA)

Program Committee:

Eneko Agirre (University of the Basque Country, Spain)
Jason Baldridge (University of Texas at Austin, USA)
Tim Baldwin (University of Melbourne, Australia)
Sam Brody (Columbia University, USA)
Alexander Clark (Royal Holloway, University of London, UK)
Shay Cohen (Carnegie Mellon University, USA)
Mona Diab (Columbia University, USA)
Gregory Druck (University of Massachusetts Amherst, USA)
Jason Eisner (Johns Hopkins University, USA)
Sharon Goldwater (University of Edinburgh, UK)
Joao Graca (University of Pennsylvania, USA)
Ioannis Klapaftis (University of York, UK)
Lillian Lee (Cornell University, USA)
Percy Liang (UC Berkeley, USA)
Diana McCarthy (Lexical Computing, Ltd., UK)
Preslav Nakov (National University of Singapore, Singapore)
Roberto Navigli (University of Rome, Italy)
Vincent Ng (UT Dallas, USA)
Ted Pedersen (University of Minnesota, USA)
Andrew Rosenberg (CUNY, USA)
Valentin Spitkovsky (Stanford University, USA)
Carlo Strapparava (FBK-irst, Italy)
Ben Taskar (University of Pennsylvania, USA)
Kristina Toutanova (Microsoft Research, USA)
Andreas Vlachos (University of Wisconsin-Madison, USA)

Invited Speaker:

Sharon Goldwater, Institute for Language, Cognition and Computation, University of Edinburg, UK

Table of Contents

Conference Program

July 30th, 2011

(9:00-9:15) Opening Words

(9:15-10:30) Invited Talk

Unsupervised NLP and Human Language Acquisition: Making Connections to Make Progress
Sharon Goldwater

(10:30-11:00) Coffee Break

(11:00:12:30) Morning Session

Structured Databases of Named Entities from Bayesian Nonparametrics
Jacob Eisenstein, Tae Yano, William Cohen, Noah Smith and Eric Xing

Unsupervised Cross-Lingual Lexical Substitution
Marianna Apidianaki

Reducing the Size of the Representation for the uDOP-Estimate
Christoph Teichmann

(12:30-14:00) Lunch Break

(14:00-14:30) Noon Session

Evaluating unsupervised learning for natural language processing tasks
Andreas Vlachos

July 30th, 2011 (continued)

(14:30-15:40) Panel Discussion

(15:40-16:10) Coffee Break

(16:10-17:15) Poster Session

Unsupervised Language-Independent Name Translation Mining from Wikipedia Infoboxes
Wen-Pin Lin, Matthew Snover and Heng Ji

Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph
Michael Speriosu, Nikita Sudan, Sid Upadhyay and Jason Baldridge

Unsupervised Bilingual POS Tagging with Markov Random Fields
Desai Chen, Chris Dyer, Shay Cohen and Noah Smith

Unsupervised Concept Annotation using Latent Dirichlet Allocation and Segmental Methods
Nathalie Camelin, Boris Detienne, Stéphane Huet, Dominique Quadri and Fabrice Lefèvre

Unsupervised Mining of Lexical Variants from Noisy Text
Stephan Gouws, Dirk Hovy and Donald Metzler

Lightly-Supervised Training for Hierarchical Phrase-Based Machine Translation
Matthias Huck, David Vilar, Daniel Stein and Hermann Ney

Unsupervised Alignment for Segmental-based Language Understanding
Stéphane Huet and Fabrice Lefèvre

Unsupervised Name Ambiguity Resolution Using A Generative Model
Zornitsa Kozareva and Sujith Ravi

Measuring the Impact of Sense Similarity on Word Sense Induction
David Jurgens and Keith Stevens