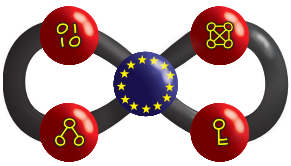
ACL HLT 2011

Workshop on Graph-based Methods for Natural Language Processing TextGraphs-6

Proceedings of the Workshop

23 June, 2011
Portland, Oregon, USA

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



©2011 The Association for Computational Linguistics

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-13 9781937284008

Preface

TextGraphs is at its SIXTH edition! This confirms that two seemingly distinct disciplines, graph theoretic models and computational linguistics, are in fact intimately connected, with a large variety of Natural Language Processing (NLP) applications adopting efficient and elegant solutions from graph-theoretical framework.

The TextGraphs workshop series addresses a broad spectrum of research areas and brings together specialists working on graph-based models and algorithms for natural language processing and computational linguistics, as well as on the theoretical foundations of related graph-based methods.

This workshop series is aimed at fostering an exchange of ideas by facilitating a discussion about both the techniques and the theoretical justification of the empirical results among the NLP community members. Spawning a deeper understanding of the basic theoretical principles involved, such interaction is vital to the further progress of graph-based NLP applications.

The submissions to this year workshop were high quality and also the selection process was more competitive than in previous editions. We selected 9 out of 16 papers for an acceptance rate of about 55%. The predominant topics of such contributions are, as usual, semantic similarity and word sense disambiguation. However, thanks also to the special theme of this year in the area of machine learning, i.e. *Graphs in Structured Input/Output Learning*, a larger use of principled statistical approaches can be observed. This trend will be nicely supported by the very interesting invited talk by Prof. Hal Daumé III on advanced and practical machine learning, entitled: *Structured Prediction need not be Slow*.

Finally, we are grateful to the European Community project, EternalS: "Trustworthy Eternal Systems via Evolving Software, Data and Knowledge" (project number FP7 247758) for continuing to sponsor our workshop.

The organizers
Irina Matveeva, Lluís Màrquez, Alessandro Moschitti and Fabio Massimo Zanzotto
Portland, June 2011

Structured Prediction need not be Slow

Invited talk

Hal Daumé III

University of Maryland – College Park

hal@umiacs.umd.edu

Abstract

Classic algorithms for predicting structured data (eg., graphs, trees, etc.) rely on expensive (sometimes intractable) inference at test time. In this talk, I'll discuss several recent approaches that enable computationally efficient (eg., linear-time) prediction at test time. These approaches fall in the category of learning algorithms that optimize accuracy for some fixed notion of efficiency. I'll conclude by considering the question: can a learning algorithm figure out how to make fast predictions on its own?

Organizers:

Irina Matveeva, Dieselpoint Inc., USA
Alessandro Moschitti, University of Trento, Italy
Lluís Màrquez, Technical University of Catalonia, Spain
Fabio Massimo Zanzotto, University of Rome "Tor Vergata", Italy

Program Committee:

Eneko Agirre, University of the Basque Country, Spain
Roberto Basili, University of Rome "Tor Vergata", Italy
Ulf Brefeld, Yahoo! Barcelona, Spain
Razvan Bunescu, Ohio University, USA
Nicola Cancedda, Xerox Research Centre Europe, France
William Cohen, Carnegie Mellon University, USA
Andras Csomai, Google USA
Mona Diab, Columbia University, USA
Gael Dias, Universidade da Beira Interior, Portugal
Michael Gamon, Microsoft Research, Redmond, USA
Thomas Gaertner, University of Bonn and Fraunhofer IAIS, Germany
Andrew Goldberg, University of Wisconsin, USA
Richard Johansson, Trento University, Italy
Lillian Lee, Cornell University, USA
Ryan McDonald, Google Research, USA
Rada Mihalcea, University of North Texas, USA
Animesh Mukherjee, CSL Lab, ISI Foundation, Torino, Italy
Bo Pang, Yahoo! Research, USA
Patrick Pantel, USC Information Sciences Institute, USA
Daniele Pighin, Technical University of Catalonia, Spain
Uwe Quasthoff, University of Leipzig, Germany
Dragomir Radev, University of Michigan, USA
Dan Roth, University of Illinois at Urbana Champaign, USA
Aitor Soroa, University of the Basque Country, Spain
Veselin Stoyanov, Johns Hopkins University, USA
Swapna Somasundaran, Siemens Corporate Research, USA

Invited Speaker:

Hal Daumé III, University of Maryland, USA

Official Sponsor:

ETERNALS: European Coordinate Action on Trustworthy Eternal Systems via Evolving Software, Data and Knowledge (project number FP7 247758)

Table of Contents

A Combination of Topic Models with Max-margin Learning for Relation Detection
Dingcheng Li, Swapna Somasundaran and Amit Chakraborty 1

Nonparametric Bayesian Word Sense Induction
Xuchen Yao and Benjamin Van Durme 10

Invariants and Variability of Synonymy Networks: Self Mediated Agreement by Confluence
Benoit Gaillard, Bruno Gaume and Emmanuel Navarro 15

Word Sense Induction by Community Detection
David Jurgens 24

Using a Wikipedia-based Semantic Relatedness Measure for Document Clustering
Majid Yazdani and Andrei Popescu-Belis 29

GrawlTCQ: Terminology and Corpora Building by Ranking Simultaneously Terms, Queries and Documents using Graph Random Walks
Clément de Groc, Xavier Tannier and Javier Couto 37

Simultaneous Similarity Learning and Feature-Weight Learning for Document Clustering
Pradeep Muthukrishnan, Dragomir Radev and Qiaozhu Mei 42

Unrestricted Quantifier Scope Disambiguation
Mehdi Manshadi and James Allen 51

From ranked words to dependency trees: two-stage unsupervised non-projective dependency parsing
Anders Søgaard 60

TextGraphs-6 Program

Thursday, June 23, 2011

9:00–9:15 Opening Remarks

Special Track Session: "Graphs in Structured Input/Output Learning"

9:15–9:40 *A Combination of Topic Models with Max-margin Learning for Relation Detection*
Dingcheng Li, Swapna Somasundaran and Amit Chakraborty

9:40–10:05 *Nonparametric Bayesian Word Sense Induction*
Xuchen Yao and Benjamin Van Durme

10:05–10:30 *Invariants and Variability of Synonymy Networks: Self Mediated Agreement by Confluence*
Benoit Gaillard, Bruno Gaume and Emmanuel Navarro

10:30–11:00 Coffee Break

Session 1

11:00–11:25 *Word Sense Induction by Community Detection*
David Jurgens

11:25–12:30 Invited talk by Hal Daumé III: *Structured Prediction need not be Slow*

12:30–14:00 Lunch Break

Thursday, June 23, 2011 (continued)

Session 2

14:00–14:25 *Using a Wikipedia-based Semantic Relatedness Measure for Document Clustering*
Majid Yazdani and Andrei Popescu-Belis

14:25–14:50 *GrawlTCQ: Terminology and Corpora Building by Ranking Simultaneously Terms, Queries and Documents using Graph Random Walks*
Clément de Groc, Xavier Tannier and Javier Couto

14:50–15:15 *Simultaneous Similarity Learning and Feature-Weight Learning for Document Clustering*
Pradeep Muthukrishnan, Dragomir Radev and Qiaozhu Mei

15:15–15:45 Coffee Break

Session 3

15:45–16:10 *Unrestricted Quantifier Scope Disambiguation*
Mehdi Manshadi and James Allen

16:10–16:35 *From ranked words to dependency trees: two-stage unsupervised non-projective dependency parsing*
Anders Søgaard

16:35–17:30 Panel Discussion

17:30–17:45 Closing Session