ACL 2010

TextGraphs-5

2010 Workshop on Graph-based Methods for Natural Language Processing

Proceedings of the Workshop

16 July 2010
Uppsala University
Uppsala, Sweden

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

©2010 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 978-1-932432-77-0 / 1-932432-77-9

Introduction

Recent years have shown an increased amount of interest in applying graph theoretic models to computational linguistics. Both graph theory and computational linguistics are well studied disciplines, which have traditionally been perceived as distinct, embracing different algorithms, different applications, and different potential end-users. However, as recent research work has shown, the two seemingly distinct disciplines 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 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.

In addition to the general goal of employing graph-theoretical methods for text processing tasks, this year we invited papers on a special theme "Graph Methods for Opinion analysis". One of the motivations for our special theme was that graphical approaches become very pertinent as the field of opinion mining advances towards deeper analysis and more complex systems. We wanted to encourage publication of early results, position papers and initiate discussions of issues in this area.

This volume contains papers accepted for publication for TextGraphs-5 2010 Workshop on Graph- Based Algorithms for Natural Language Processing. TextGraphs-5 was held on 16th of July 2010 in Uppsala, Sweden at ACL 2010, the 48th Annual Meeting of the Association for Computational Linguistics. This was the fifth workshop in this series, building on the successes of previous workshops that were held at HLT-NAACL (2006, 2007), Coling (2008) and ACL (2009).

We issued calls for both regular, short and position papers. Six regular and ten short papers were accepted for presentation, based on the careful reviews of our program committee. We are very thankful to the incredible program committee, whose detailed comments and useful suggestions have benefited the papers and helped us to create a great program. We are particularly grateful for the timely reviews, especially considering that we had a tight schedule.

The articles apply graph methods to a variety of NLP problems such as Word Sense Disambiguation (De Cao et al., Fagerlund et al., Biemannl), Topic Segmentation (Ambwani and Davis), Summarization (Jorge and Pardo, Fukumoto and Suzuki), Language Evolution (Enright), Language Acquisition (German et al.), Language Resources (Archer), Lexical Networks (Oliveira and Gomes, Görnerup and Karlgren) and Clustering (Wieling and Nerbonne, Chen and Ji). Additionally, we have selected three papers for our special theme: Zontone et al. use graph theoretic dominant set clustering algorithm for the annotation of images with sentiment scores, Amancio et al. employ complex network features to distinguish between positive and negative opinions, and Tatzl and Waldhauser offer a formalization that could help to automate opinion extraction within the Media Content Analysis framework.

Last but not the least, having a prominent researcher as the invited speaker significantly added to the success of our workshop. We would like to thank Professor Edwin Hancock from the

University of York for his captivating talk on graph-based machine learning algorithms. We are also grateful to the European Community project, EternalS: "Trustworthy Eternal Systems via Evolving Software, Data and Knowledge" (project number FP7 247758) for sponsoring our invited speaker.

Enjoy the workshop!

The organizers

Carmen Banea, Alessandro Moschitti, Swapna Somasundaran and Fabio Massimo Zanzotto

Upsala, July 2010

Organizers:

Carmen Banea, University of North Texas, US
Alessandro Moschitti, University of Trento, Italy
Swapna Somasundaran, University of Pittsburgh, US
Fabio Massimo Zanzotto, University of Rome, Italy

Program Committee:

Andras Csomai, Google Inc.
Andrea Esuli, Italian National Research Council, Italy
Andrea Passerini, Trento University, Italy
Andrew Goldberg, University of Wisconsin, US
Animesh Mukherjee, ISI Foundation, Turin, Italy
Carlo Strapparava, Istituto per la Ricerca Scientifica e Tecnologica, Italy
Dragomir R. Radev, University of Michigan, US
Eduard Hovy, Information Sciences Institute, US
Fabrizio Sebastiani, Istituto di Scienza e Tecnologia dell'Informazione, Italy
Giuseppe Carenini, University of British Columbia, Canada
Hiroya Takamura, Tokyo Institute of Technology, Japan
Lillian Lee, Cornell University, US
Lise Getoor, University of Maryland, US
Lluis Marquez Villodre, Universidad Politecnica de Catalunya, Spain
Michael Gamon, Microsoft Research, Redmond
Michael Strube, EML research, Germany
Monojit Choudhury, Microsoft Research, India
Richard Johansson, Trento University
Robero Basili, University of Rome, Italy
Roi Blanco, Yahoo!, Spain
Smaranda Muresan, Rutgers University, US
Sofus Macskassy, Fetch Technologies El Segundo, US
Stefan Siersdorfer, L3S Research Center, Hannover, Germany
Theresa Wilson, University of Edinburgh, Germany
Thomas Gartner, Fraunhofer Institute for Intelligent Analysis and Information Systems, Germany
Ulf Brefeld, Yahoo!, Spain
Veselin Stoyanov, Cornell University, US
William Cohen, Carnegie Mellon University, US
Xiaojin Zhu, University of Wisconsin, US
Yejin Choi, Cornell University, US

Invited Speaker:

Edwin R. Hancock, University of York, UK

Table of Contents

Graph-Based Clustering for Computational Linguistics: A Survey
Zheng Chen and Heng Ji 1

Towards the Automatic Creation of a Wordnet from a Term-Based Lexical Network
Hugo Gonçalo Oliveira and Paulo Gomes 10

An Investigation on the Influence of Frequency on the Lexical Organization of Verbs
Daniel German, Aline Villavicencio and Maity Siqueira 19

Robust and Efficient Page Rank for Word Sense Disambiguation
Diego De Cao, Roberto Basili, Matteo Luciani, Francesco Mesiano and Riccardo Rossi 24

Hierarchical Spectral Partitioning of Bipartite Graphs to Cluster Dialects and Identify Distinguishing Features
Martijn Wieling and John Nerbonne 33

A Character-Based Intersection Graph Approach to Linguistic Phylogeny
Jessica Enright 42

Spectral Approaches to Learning in the Graph Domain
Edwin Hancock 47

Cross-Lingual Comparison between Distributionally Determined Word Similarity Networks
Olof Görnerup and Jussi Karlgren 48

Co-Occurrence Cluster Features for Lexical Substitutions in Context
Chris Biemann 55

Contextually-Mediated Semantic Similarity Graphs for Topic Segmentation
Geetu Ambwani and Anthony Davis 60

MuLLinG: MultiLevel Linguistic Graphs for Knowledge Extraction
Vincent Archer 69

Experiments with CST-Based Multidocument Summarization
Maria Lucia Castro Jorge and Thiago Pardo 74

Distinguishing between Positive and Negative Opinions with Complex Network Features
Diego Raphael Amancio, Renato Fabbri, Osvaldo Novais Oliveira Jr., Maria das Graças Volpe Nunes and Luciano da Fontoura Costa 83

Image and Collateral Text in Support of Auto-Annotation and Sentiment Analysis
Pamela Zontone, Giulia Boato, Jonathon Hare, Paul Lewis, Stefan Siersdorfer and Enrico Minack 88

Aggregating Opinions: Explorations into Graphs and Media Content Analysis
Gabriele Tatzl and Christoph Waldhauser 93

Eliminating Redundancy by Spectral Relaxation for Multi-Document Summarization
Fumiyo Fukumoto, Akina Sakai and Yoshimi Suzuki 98

Computing Word Senses by Semantic Mirroring and Spectral Graph Partitioning
Martin Fagerlund, Magnus Merkel, Lars Eldén and Lars Ahrenberg 103

TextGraphs-5 Program

Friday July 16, 2010

09:00–09:10 Welcome to TextGraphs 5

Session 1: Lexical Clustering and Disambiguation

09:10–09:30 *Graph-Based Clustering for Computational Linguistics: A Survey*
Zheng Chen and Heng Ji

09:30–09:50 *Towards the Automatic Creation of a Wordnet from a Term-Based Lexical Network*
Hugo Gonçalo Oliveira and Paulo Gomes

09:50–10:10 *An Investigation on the Influence of Frequency on the Lexical Organization of Verbs*
Daniel German, Aline Villavicencio and Maity Siqueira

10:10–10:30 *Robust and Efficient Page Rank for Word Sense Disambiguation*
Diego De Cao, Roberto Basili, Matteo Luciani, Francesco Mesiano and Riccardo Rossi

10:30–11:00 Coffee Break

Session 2: Clustering Languages and Dialects

11:00–11:20 *Hierarchical Spectral Partitioning of Bipartite Graphs to Cluster Dialects and Identify Distinguishing Features*
Martijn Wieling and John Nerbonne

11:20–11:40 *A Character-Based Intersection Graph Approach to Linguistic Phylogeny*
Jessica Enright

11:40–12:40 Invited Talk

11:40–12:40 *Spectral Approaches to Learning in the Graph Domain*
Edwin Hancock

12:40–13:50 Lunch break

Session 3: Lexical Similarity and Its application

13:50–14:10 *Cross-Lingual Comparison between Distributionally Determined Word Similarity Networks*
Olof Görnerup and Jussi Karlgren

14:10–14:30 *Co-Occurrence Cluster Features for Lexical Substitutions in Context*
Chris Biemann

14:30–14:50 *Contextually-Mediated Semantic Similarity Graphs for Topic Segmentation*
Geetu Ambwani and Anthony Davis

14:50–15:10 *MuLLinG: MultiLevel Linguistic Graphs for Knowledge Extraction*
Vincent Archer

15:10–15:30 *Experiments with CST-Based Multidocument Summarization*
Maria Lucia Castro Jorge and Thiago Pardo

15:30–16:00 Coffee Break

Special Session on Opinion Mining

16:00–16:20 *Distinguishing between Positive and Negative Opinions with Complex Network Features*
Diego Raphael Amancio, Renato Fabbri, Osvaldo Novais Oliveira Jr., Maria das Graças Volpe Nunes and Luciano da Fontoura Costa

16:20–16:40 *Image and Collateral Text in Support of Auto-Annotation and Sentiment Analysis*
Pamela Zontone, Giulia Boato, Jonathon Hare, Paul Lewis, Stefan Siersdorfer and Enrico Minack

16:40–17:00 *Aggregating Opinions: Explorations into Graphs and Media Content Analysis*
Gabriele Tatzl and Christoph Waldhauser

(continued)

Session 5: Spectral Approaches

17:00–17:20 *Eliminating Redundancy by Spectral Relaxation for Multi-Document Summarization*
Fumiyo Fukumoto, Akina Sakai and Yoshimi Suzuki

17:20–17:40 *Computing Word Senses by Semantic Mirroring and Spectral Graph Partitioning*
Martin Fagerlund, Magnus Merkel, Lars Eldén and Lars Ahrenberg

17:40–18:00 Final Wrap-up