

NAACL HLT 2010

Computational Linguistics in a World of Social Media

Proceedings of the Workshop

June 6, 2010
Los Angeles, California

USB memory sticks produced by
Omnipress Inc.
2600 Anderson Street
Madison, WI 53707
USA

©2010 The Association for Computational Linguistics

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

Introduction

Social Media eg Twitter, Blogs, Forums, FaceBook, Google Buzz has exploded over the last few years. FaceBook is now the most visited site in the US, overtaking Google in the first quarter of 2010. These sites contain the aggregated beliefs and opinions of millions of people on an epic range of topics, in a multitude of languages.

Social Media presents many challenges and opportunities to the ACL community, with this workshop being the first of its kind at a computational linguistics venue. Accepted papers range from story detection and tracking to discourse, applied across new and old media including company announcements, news, forums, blogs and micro-blogs. A notable aspect is the predominance of Twitter as a Social Media resource.

We experimented with a new kind of workshop based on a philosophy that ACL workshops should serve a different purpose than the main conference. To encourage submission of new ideas, we restricted papers to just two sides. And, to create a fast-paced and highly interactive workshop, each accepted paper was allotted a short talk and a poster.

Our invited talks touch upon various aspects of Social Media; distilling collective beliefs and making them concrete (Noah Smith); new technologies (Casey Whitelaw); the relationships between old and new media (Jochen Leidner). They give a balance between industry and academia and highlight the relationships between Human Language Technologies and Social Media.

We are grateful to Google Research for sponsoring the workshop. We used part of the sponsorship to award a prize to the best presentation (be it poster or short talk). This is a conscious decision to reward people for putting effort into communicating their ideas. At the time of writing this preface we have not made the award. But by the time you are reading this, it may well be you!

Organizers:

Ben Hachey, Capital Markets CRC and Macquarie University
Miles Osborne, University of Edinburgh

Program Committee:

Beatrice Alex, University of Edinburgh
Regina Barzilay, Massachusetts Institute of Technology
James Curran, University of Sydney
Murray Z. Frank, University of Minnesota
Michael Gamon, Microsoft Research
Nikesh Garera, Kosmix
Keith Hall, Google
John Henderson, MITRE
Bill Hu, Arkansas State University
Ben Hutchinson, Google
Rebecca Hwa, University of Pittsburgh
Mirella Lapata, University of Edinburgh
Victor Lavrenko, University of Edinburgh
Jochen Leidner, Thomson Reuters
Adam Lopez, University of Edinburgh
Craig Macaulay, Ernst & Young
Rob Malouf, San Diego State University
Yuval Marom, Pacific Brands
Rada Mihalcea, University of North Texas
Maria Milosavljevic, Macquarie University
Gabriel Murray, University of British Columbia
Deepak Ravichandran, Google
Calum Robertson, Sirca
Anoop Sarkar, Simon Fraser University
Robert P. Schumaker, Iona College
Noah Smith, Carnegie Mellon University
Tae Yano, Carnegie Mellon University

Invited Speakers:

Jochen Leidner, Thomson Reuters
Noah Smith, Carnegie Mellon University
Casey Whitelaw, Google

Table of Contents

<i>The “Nays” Have It: Exploring Effects of Sentiment in Collaborative Knowledge Sharing</i> Ablimit Aji and Eugene Agichtein	1
<i>An Analysis of Verbs in Financial News Articles and their Impact on Stock Price</i> Robert Schumaker	3
<i>Detecting Word Misuse in Chinese</i> Wei Liu	5
<i>An Information-Retrieval Approach to Language Modeling: Applications to Social Data</i> Juan Huerta	7
<i>Towards Automatic Question Answering over Social Media by Learning Question Equivalence Patterns</i> Tianyong Hao, Wenyin Liu and Eugene Agichtein	9
<i>Modeling Message Roles and Influence in Q&A Forums</i> Jeonhyung Kang and Jihie Kim	11
<i>Towards Modeling Social and Content Dynamics in Discussion Forums</i> Jihie Kim and Aram Galstyan	13
<i>Intelligent Linux Information Access by Data Mining: the ILIAD Project</i> Timothy Baldwin, David Martinez, Richard Penman, Su Nam Kim, Marco Lui, Li Wang and Andrew MacKinlay	15
<i>Mining User Experiences from Online Forums: An Exploration</i> Valentin Jijkoun, Wouter Weerkamp, Maarten de Rijke, Paul Ackermans and Gijs Geleijnse ...	17
<i>Social Links from Latent Topics in Microblogs</i> Kriti Puniyani, Jacob Eisenstein, Shay B. Cohen and Eric Xing	19
<i>Automatic Detection of Tags for Political Blogs</i> Khairun-nisa Hassanali and Vasileios Hatzivassiloglou	21
<i>Twitter in Mass Emergency: What NLP Can Contribute</i> William J. Corvey, Sarah Vieweg, Travis Rood and Martha Palmer	23
<i>The Edinburgh Twitter Corpus</i> Saša Petrović, Miles Osborne and Victor Lavrenko	25
<i>Labelling and Spatio-Temporal Grounding of News Events</i> Bea Alex and Claire Grover	27
<i>Tracking Information Flow between Primary and Secondary News Sources</i> Will Radford, Ben Hachey, James Curran and Maria Milosavljevic	29

Detecting controversies in Twitter: a first study

Marco Pennacchiotti and Ana-Maria Popescu 31

Workshop Program

Sunday, June 6, 2010

Session 1: Applications in Social Media

- 9:00 Invited Talk: *Text-Driven Forecasting*
Noah Smith, Carnegie Mellon University
- 9:45 *The “Nays” Have It: Exploring Effects of Sentiment in Collaborative Knowledge Sharing*
Ablimit Aji and Eugene Agichtein
- 9:51 *An Analysis of Verbs in Financial News Articles and their Impact on Stock Price*
Robert Schumaker
- 9:57 *Detecting Word Misuse in Chinese*
Wei Liu
- 10:03 *An Information-Retrieval Approach to Language Modeling: Applications to Social Data*
Juan Huerta
- 10:09 *Towards Automatic Question Answering over Social Media by Learning Question Equivalence Patterns*
Tianyong Hao, Wenyin Liu and Eugene Agichtein
- 10:15 Posters
- 10:30 Posters & Coffee

Sunday, June 6, 2010 (continued)

Session 2: Forums and Networks

- 11:00 Invited Talk: *Google Wave as a Computational Linguistic Platform*
Casey Whitelaw, Google
- 11:45 *Modeling Message Roles and Influence in Q&A Forums*
Jeonhyung Kang and Jihie Kim
- 11:51 *Towards Modeling Social and Content Dynamics in Discussion Forums*
Jihie Kim and Aram Galstyan
- 11:57 *Intelligent Linux Information Access by Data Mining: the ILIAD Project*
Timothy Baldwin, David Martinez, Richard Penman, Su Nam Kim, Marco Lui, Li Wang
and Andrew MacKinlay
- 12:03 *Mining User Experiences from Online Forums: An Exploration*
Valentin Jijkoun, Wouter Weerkamp, Maarten de Rijke, Paul Ackermans and Gijs Geleij-
jnse
- 12:09 *Social Links from Latent Topics in Microblogs*
Kriti Puniyani, Jacob Eisenstein, Shay B. Cohen and Eric Xing

12:15 Posters

12:30 Lunch Break

Session 3: (Micro)-Blogs and Information Tracking

- 13:30 Invited Talk: *The Interaction between News and Social Media*
Jochen Leidner, Thomson Reuters
- 14:15 *Automatic Detection of Tags for Political Blogs*
Khairun-nisa Hassanali and Vasileios Hatzivassiloglou
- 14:21 *Twitter in Mass Emergency: What NLP Can Contribute*
William J. Corvey, Sarah Vieweg, Travis Rood and Martha Palmer
- 14:27 *The Edinburgh Twitter Corpus*
Saša Petrović, Miles Osborne and Victor Lavrenko

Sunday, June 6, 2010 (continued)

- 14:33 *Labelling and Spatio-Temporal Grounding of News Events*
Bea Alex and Claire Grover
- 14:39 *Tracking Information Flow between Primary and Secondary News Sources*
Will Radford, Ben Hachey, James Curran and Maria Milosavljevic
- 14:45 *Detecting controversies in Twitter: a first study*
Marco Pennacchiotti and Ana-Maria Popescu
- 14:51 Posters
- 15:00 Posters & Coffee
- 16:00 Finish

