ACL 2010

DANLP 2010

2010 Workshop on
Domain Adaptation for Natural Language Processing

Proceedings of the Workshop

15 July 2010
Uppsala University
Uppsala, Sweden

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

DANLP 2010 Invited Speaker Sponsor:

©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 978-1-932432-80-0 / 1-932432-80-9

Preface

Most modern Natural Language Processing (NLP) systems are subject to the well known problem of lack of portability to new domains and genres: there is a substantial drop in their performance when tested on data from a new domain, i.e., when test data is drawn from a related but different distribution from training data. This problem is inherent in the assumption of independent and identically distributed (i.i.d.) variables for machine learning systems, but has started to get attention only in recent years. The need for domain adaptation arises in almost all NLP tasks – the goal of this workshop is to provide a meeting point for research that approaches the problem of adaptation from the varied perspectives of machine learning and a variety of NLP tasks. We believe there is much to gain by treating domain adaptation as a general learning strategy that utilizes prior knowledge of a specific or a general domain in learning about a new domain. Sharing insights, methodologies and successes across tasks will contribute towards a better understanding of this problem. To this end, this workshop presents original research in areas such as parsing, machine translation, dialog act tagging, entity recognition, summarization, etc. with the common theme of domain adaptation. We received sixteen submissions in all, out of which eight were selected for inclusion in the workshop.

We thank the members of the Program Committee for timely and insightful reviews, and the invited speaker John Blitzer for his talk.

Hal Daumé III, Tejaswini Deoskar, David McClosky, Barbara Plank and Jörg Tiedemann

Organizers:

Hal Daumé III, University of Utah, USA
Tejaswini Deoskar, University of Amsterdam, The Netherlands
David McClosky, Stanford University, USA
Barbara Plank, University of Groningen, The Netherlands
Jörg Tiedemann, Uppsala University, Sweden

Program Committee:

Eneko Agirre, University of the Basque Country, Spain
John Blitzer, University of California, USA
Walter Daelemans, University of Antwerp, Belgium
Mark Dredze, Johns Hopkins University, USA
Philipp Koehn, University of Edinburgh, United Kingdom
Kevin Duh, NTT Communication Science Laboratories, Japan
Jing Jiang, Singapore Management University, Singapore
Oier Lopez de Lacalle, University of the Basque Country, Spain
Robert Malouf, San Diego State University, USA
Ray Mooney, University Texas, USA
Hwee Tou Ng, National University of Singapore, Singapore
Khalil Sima'an, University of Amsterdam, The Netherlands
Michel Simard, National Research Council of Canada, Canada
Jun'ichi Tsujii, University of Tokyo, Japan
Antal van den Bosch, Tilburg University, The Netherlands
Josef van Genabith, Dublin City University, Ireland
Yi Zhang, German Research Centre for Artificial Intelligence (DFKI GmbH) and Saarland University, Germany

Invited Speaker:

John Blitzer, University of California, USA

Table of Contents

Adaptive Parameters for Entity Recognition with Perceptron HMMs
Massimiliano Ciaramita and Olivier Chapelle 1

Context Adaptation in Statistical Machine Translation Using Models with Exponentially Decaying Cache
Jörg Tiedemann 8

Domain Adaptation to Summarize Human Conversations
Oana Sandu, Giuseppe Carenini, Gabriel Murray and Raymond Ng 16

Exploring Representation-Learning Approaches to Domain Adaptation
Fei Huang and Alexander Yates 23

Using Domain Similarity for Performance Estimation
Vincent Van Asch and Walter Daelemans 31

Self-Training without Reranking for Parser Domain Adaptation and Its Impact on Semantic Role Labeling
Kenji Sagae 37

Domain Adaptation with Unlabeled Data for Dialog Act Tagging
Anna Margolis, Karen Livescu and Mari Ostendorf 45

Frustratingly Easy Semi-Supervised Domain Adaptation
Hal Daumé III, Abhishek Kumar and Avishek Saha 53

Conference Program

Thursday 15 July 2010

9:15-9:30 Opening by Barbara Plank

9:30-10:30 Invited Talk "Semi-supervised Domain Adaptation: From Practice to Theory" by John Blitzer

10:30-11:00 Morning Break

Session I:

11:00–11:25 *Adaptive Parameters for Entity Recognition with Perceptron HMMs*
Massimiliano Ciaramita and Olivier Chapelle

11:30–11:55 *Context Adaptation in Statistical Machine Translation Using Models with Exponentially Decaying Cache*
Jörg Tiedemann

12:00–12:25 *Domain Adaptation to Summarize Human Conversations*
Oana Sandu, Giuseppe Carenini, Gabriel Murray and Raymond Ng

12:30-14:00 Lunch

Session II:

14:00–14:25 *Exploring Representation-Learning Approaches to Domain Adaptation*
Fei Huang and Alexander Yates

14:30–14:55 *Using Domain Similarity for Performance Estimation*
Vincent Van Asch and Walter Daelemans

15:00–15:25 *Self-Training without Reranking for Parser Domain Adaptation and Its Impact on Semantic Role Labeling*
Kenji Sagae

15:30-16:00 Afternoon Break

Thursday 15 July 2010 (continued)

Session III:

16:00–16:25 *Domain Adaptation with Unlabeled Data for Dialog Act Tagging*
Anna Margolis, Karen Livescu and Mari Ostendorf

16:30–16:55 *Frustratingly Easy Semi-Supervised Domain Adaptation*
Hal Daumé III, Abhishek Kumar and Avishek Saha

17:00-17:45 Panel Discussion by John Blitzer, Walter Daelemans, Hal Daumé III, Jing Jiang, Khalil Sima'an