NAACL HLT 2010

First International Workshop on Formalisms and Methodology for Learning by Reading (FAM-LbR)

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

It has been a long term vision of Artificial Intelligence to develop Learning by Reading systems that can capture knowledge from naturally occurring texts, convert it into a deep logical notation and perform some inferences/reasoning on them. Such systems directly build on relatively mature areas of research, including Information Extraction (for picking out relevant information from the text), Commonsense and AI Reasoning (for deriving inferences from the knowledge acquired), Bootstrapped Learning (for using the learned knowledge to expand the knowledge base) and Question Answering (for providing evaluation mechanisms for Learning by Reading systems). In Natural Language Processing, statistical learning techniques have provided new solutions and breakthroughs in various areas over the last decade. In Knowledge Representation and Reasoning, systems have achieved impressive performance and scale in far more complex problems than the past.

Learning by Reading is a two-part process. One part deals with extracting interesting information from naturally occurring texts, and the other is to use this extracted knowledge to expand the knowledge base and consequently the system's inference capabilities. Previous systems have chosen either a "broad and shallow" or a "narrow and deep" knowledge acquisition and reasoning strategy. These techniques are constrained by either their limited reasoning ability or their extreme domain dependence.

The goal of this workshop is to draw together researchers to explore the nature and degree of integration possible between symbolic and statistical techniques for knowledge acquisition and reasoning. In particular, given these developments, what is the role of commonsense knowledge and reasoning in language understanding? What are the limitations of each style of processing, and how can they be overcome by complementary strengths of the other? What are appropriate evaluation metrics for Learning by Reading systems?

Organizers:

Rutu Mulkar-Mehta (University of Southern California)
James Allen (University of Rochester)
Jerry Hobbs (University of Southern California)
Eduard Hovy (University of Southern California)
Bernardo Magnini (Povo-Trento, Italy)
Chris Manning (Stanford University)

Program Committee:

Ken Barker (University of Texas, Austin)
Peter Clark (Boeing)
Ido Dagan (Bar Ilan University)
Oren Etzioni (University of Washington)
David Ferrucci (IBM)
Marjorie Freedman (BBN)
Tom Mitchell (Carnegie Mellon University)
Dan Moldovan (University of Texas at Dallas)
Alessandro Moschitti (University of Trento)
Sergei Nirenburg (University of Maryland, Baltimore County)
Tim Oates (University of Maryland, Baltimore County)
Dan Oblinger (DARPA)
Ellen Riloff (University of Utah)
Stuart Shapiro (SUNY Buffalo)
Len Schubert (Rochester University)
Ralph Weischedel (BBN)
Michael Witbrock (Cyc Corp.)

Table of Contents

Machine Reading as a Process of Partial Question-Answering
Peter Clark and Phil Harrison 1

Building an end-to-end text reading system based on a packed representation
Doo Soon Kim, Ken Barker and Bruce Porter 10

Semantic Enrichment of Text with Background Knowledge
Anselmo Peñas and Eduard Hovy 15

Large Scale Relation Detection
Chris Welty, James Fan, David Gondek and Andrew Schlaikjer 24

Mining Script-Like Structures from the Web
Niels Kasch and Tim Oates 34

Open-domain Commonsense Reasoning Using Discourse Relations from a Corpus of Weblog Stories
Matthew Gerber, Andrew Gordon and Kenji Sagae 43

Semantic Role Labeling for Open Information Extraction
Janara Christensen, Mausam, Stephen Soderland and Oren Etzioni 52

Empirical Studies in Learning to Read
Marjorie Freedman, Edward Loper, Elizabeth Boschee and Ralph Weischedel 61

Learning Rules from Incomplete Examples: A Pragmatic Approach
Janardhan Rao Doppa, Mohammad NasrEsfahani, Mohammad Sorower, Thomas G. Dietterich, Xiaoli Fern and Prasad Tadepalli 70

Unsupervised techniques for discovering ontology elements from Wikipedia article links
Zareen Syed and Tim Finin 78

Machine Reading at the University of Washington
Hoifung Poon, Janara Christensen, Pedro Domingos, Oren Etzioni, Raphael Hoffmann, Chloe Kiddon, Thomas Lin, Xiao Ling, Mausam, Alan Ritter, Stefan Schoenmackers, Stephen Soderland, Dan Weld, Fei Wu and Congle Zhang 87

Analogical Dialogue Acts: Supporting Learning by Reading Analogies
David Barbella and Kenneth Forbus 96

A Hybrid Approach to Unsupervised Relation Discovery Based on Linguistic Analysis and Semantic Typing
Zareen Syed and Evelyne Viegas 105

Supporting rule-based representations with corpus-derived lexical information.
Annie Zaenen, Cleo Condoravdi, Daniel Bobrow and Raphael Hoffmann 114

PRISMATIC: Inducing Knowledge from a Large Scale Lexicalized Relation Resource
James Fan, David Ferrucci, David Gondek and Aditya Kalyanpur 122

Workshop Program

Sunday, June 6, 2010

9:00–9:10 **Opening**

Session 1

9:10–9:30 *Machine Reading as a Process of Partial Question-Answering*
Peter Clark and Phil Harrison

9:30–9:50 *Building an end-to-end text reading system based on a packed representation*
Doo Soon Kim, Ken Barker and Bruce Porter

9:50–10:10 *Semantic Enrichment of Text with Background Knowledge*
Anselmo Peñas and Eduard Hovy

10:10–10:30 Discussion

10:30–11:00 **Break**

Session 2

11:00–11:20 *Large Scale Relation Detection*
Chris Welty, James Fan, David Gondek and Andrew Schlaikjer

11:20–11:40 *Mining Script-Like Structures from the Web*
Niels Kasch and Tim Oates

11:40–12:00 *Open-domain Commonsense Reasoning Using Discourse Relations from a Corpus of Weblog Stories*
Matthew Gerber, Andrew Gordon and Kenji Sagae

12:00–12:20 Discussion

12:20–2:00 **Lunch**

Sunday, June 6, 2010 (continued)

Session 3

2:00–2:20 *Semantic Role Labeling for Open Information Extraction*
Janara Christensen, Mausam, Stephen Soderland and Oren Etzioni

2:20–2:40 *Empirical Studies in Learning to Read*
Marjorie Freedman, Edward Loper, Elizabeth Boschee and Ralph Weischedel

2:40–3:00 *Learning Rules from Incomplete Examples: A Pragmatic Approach*
Janardhan Rao Doppa, Mohammad NasrEsfahani, Mohammad Sorower, Thomas G. Dietterich, Xiaoli Fern and Prasad Tadepalli

3:00–3:30 **Break**

3:30–4:30 **Poster Session**

Unsupervised techniques for discovering ontology elements from Wikipedia article links
Zareen Syed and Tim Finin

Machine Reading at the University of Washington
Hoifung Poon, Janara Christensen, Pedro Domingos, Oren Etzioni, Raphael Hoffmann, Chloe Kiddon, Thomas Lin, Xiao Ling, Mausam, Alan Ritter, Stefan Schoenmackers, Stephen Soderland, Dan Weld, Fei Wu and Congle Zhang

Analogical Dialogue Acts: Supporting Learning by Reading Analogies
David Barbella and Kenneth Forbus

A Hybrid Approach to Unsupervised Relation Discovery Based on Linguistic Analysis and Semantic Typing
Zareen Syed and Evelyne Viegas

Supporting rule-based representations with corpus-derived lexical information.
Annie Zaenen, Cleo Condoravdi, Daniel Bobrow and Raphael Hoffmann

PRISMATIC: Inducing Knowledge from a Large Scale Lexicalized Relation Resource
James Fan, David Ferrucci, David Gondek and Aditya Kalyanpur

4:30–5:00 **Discussion**