

CoNLL-2010: Shared Task

**Fourteenth Conference on
Computational Natural Language Learning**

Proceedings of the Shared Task

15-16 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 978-1-932432-84-8 / 1-932432-84-1

Introduction

This volume consists of the descriptions of the CoNLL-2010 Shared Task and the participating systems. The shared task was dedicated to the detection of uncertainty cues and their linguistic scope in natural language text. The motivation behind this task was that distinguishing factual and uncertain information in texts is of essential importance in information extraction.

The shared task addressed the detection of uncertainty in two domains. As uncertainty detection is extremely important for biomedical information extraction and most existing approaches have targeted such applications, participants were asked to develop systems for hedge detection in biological scientific articles. Uncertainty detection is also important, e.g. in encyclopedias, where the goal is to collect reliable world knowledge about real-world concepts and topics.

Two uncertainty detection tasks, sentence classification and in-sentence hedge scope detection were given to the participants. A total of 23 teams participated in the shared task. Those who participated in both tasks were invited to write a paper up to 8 pages. The page limit for those who participated only in the first task was 6 pages.

Although several approaches were introduced by the participants of the shared task and we believe that the ideas described in this proceedings can serve as an excellent starting point for the development of an uncertainty detector, there is a lot of room for improving such systems. The manually annotated datasets and software tools developed for the shared task may act as benchmarks for these future experiments and they are freely available at <http://www.inf.u-szeged.hu/rgai/conll2010st>.

Szeged, May, 2010

Richárd Farkas, for the Shared Task organizers

Organizers:

Richárd Farkas, Human Language Technology Group, University of Szeged
Veronika Vincze, Human Language Technology Group, University of Szeged
György Szarvas, Ubiquitous Knowledge Processing Lab, Technische Universität Darmstadt
György Móra, Human Language Technology Group, University of Szeged
János Csirik, Research Group on Artificial Intelligence, Hungarian Academy of Sciences

Reviewers:

Ekaterina Buyko, University of Jena
Kevin B. Cohen, University of Colorado
Hercules Dalianis, Stockholm University
Maria Georgescu, University of Geneva
Filip Ginter, University of Turku
Henk Harkema, University of Pittsburgh
Shen Jianping, Harbin Institute of Technology
Yoshinobu Kano, University of Tokyo
Jin-Dong Kim, Database Center for Life Science, Japan
Ruy Milidiú, Pontifícia Universidade Católica do Rio de Janeiro
Roser Morante, University of Antwerp
Lilja Øvrelid, University of Potsdam
Arzucan Özgür, University of Michigan
Vinodkumar Prabhakaran, Columbia University
Sampo Pyysalo, University of Tokyo
Marek Rei, Cambridge University
Buzhou Tang, Harbin Institute of Technology
Erik Tjong Kim Sang, University of Groningen
Katrín Tomanek, University of Jena
Erik Velldal, University of Oslo
Andreas Vlachos, University of Wisconsin-Madison
Xinglong Wang, University of Manchester
Torsten Zesch, Technische Universität Darmstadt
Qi Zhao, Harbin Institute of Technology
HuiWei Zhou, Dalian University of Technology

Table of Contents

<i>The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text</i> Richárd Farkas, Veronika Vincze, György Móra, János Csirik and György Szarvas	1
<i>A Cascade Method for Detecting Hedges and their Scope in Natural Language Text</i> Buzhou Tang, Xiaolong Wang, Xuan Wang, Bo Yuan and Shixi Fan	13
<i>Detecting Speculative Language Using Syntactic Dependencies and Logistic Regression</i> Andreas Vlachos and Mark Craven	18
<i>A Hedgehop over a Max-Margin Framework Using Hedge Cues</i> Maria Georgescu	26
<i>Detecting Hedge Cues and their Scopes with Average Perceptron</i> Feng Ji, Xipeng Qiu and Xuanjing Huang	32
<i>Memory-Based Resolution of In-Sentence Scopes of Hedge Cues</i> Roser Morante, Vincent Van Asch and Walter Daelemans	40
<i>Resolving Speculation: MaxEnt Cue Classification and Dependency-Based Scope Rules</i> Erik Velldal, Lilja Øvrelid and Stephan Oepen	48
<i>Combining Manual Rules and Supervised Learning for Hedge Cue and Scope Detection</i> Marek Rei and Ted Briscoe	56
<i>Hedge Detection Using the RelHunter Approach</i> Eraldo Fernandes, Carlos Crestana and Ruy Milidiú	64
<i>A High-Precision Approach to Detecting Hedges and their Scopes</i> Halil Kilicoglu and Sabine Bergler	70
<i>Exploiting Rich Features for Detecting Hedges and their Scope</i> Xinxin Li, Jianping Shen, Xiang Gao and Xuan Wang	78
<i>Uncertainty Detection as Approximate Max-Margin Sequence Labelling</i> Oscar Täckström, Sumithra Velupillai, Martin Hassel, Gunnar Eriksson, Hercules Dalianis and Jussi Karlgrén	84
<i>Hedge Detection and Scope Finding by Sequence Labeling with Procedural Feature Selection</i> Shaodian Zhang, Hai Zhao, Guodong Zhou and Bao-liang Lu	92
<i>Learning to Detect Hedges and their Scope Using CRF</i> Qi Zhao, Chengjie Sun, Bingquan Liu and Yong Cheng	100
<i>Exploiting Multi-Features to Detect Hedges and their Scope in Biomedical Texts</i> Huiwei Zhou, Xiaoyan Li, Degen Huang, Zezhong Li and Yuansheng Yang	106
<i>A Lucene and Maximum Entropy Model Based Hedge Detection System</i> Lin Chen and Barbara Di Eugenio	114
<i>HedgeHunter: A System for Hedge Detection and Uncertainty Classification</i> David Clausen	120

<i>Exploiting CCG Structures with Tree Kernels for Speculation Detection</i>	
Liliana Mamani Sánchez, Baoli Li and Carl Vogel	126
<i>Uncertainty Learning Using SVMs and CRFs</i>	
Vinodkumar Prabhakaran	132
<i>Features for Detecting Hedge Cues</i>	
Nobuyuki Shimizu and Hiroshi Nakagawa	138
<i>A Simple Ensemble Method for Hedge Identification</i>	
Ferenc Szidarovszky, Illés Solt and Domonkos Tikk	144
<i>A Baseline Approach for Detecting Sentences Containing Uncertainty</i>	
Erik Tjong Kim Sang	148
<i>Hedge Classification with Syntactic Dependency Features Based on an Ensemble Classifier</i>	
Yi Zheng, Qifeng Dai, Qiming Luo and Enhong Chen	151

Conference Program

Thursday, July 15, 2010

Shared Task Session 1: Overview and Oral Presentations (16:00-17:30)

- 16:00–16:20 *The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text*
Richárd Farkas, Veronika Vincze, György Móra, János Csirik and György Szarvas
- 16:20–16:30 *A Cascade Method for Detecting Hedges and their Scope in Natural Language Text*
Buzhou Tang, Xiaolong Wang, Xuan Wang, Bo Yuan and Shixi Fan
- 16:30–16:40 *Detecting Speculative Language Using Syntactic Dependencies and Logistic Regression*
Andreas Vlachos and Mark Craven
- 16:40–16:50 *A Hedgehop over a Max-Margin Framework Using Hedge Cues*
Maria Georgescu
- 16:50–17:00 *Detecting Hedge Cues and their Scopes with Average Perceptron*
Feng Ji, Xipeng Qiu and Xuanjing Huang
- 17:00–17:10 *Memory-Based Resolution of In-Sentence Scopes of Hedge Cues*
Roser Morante, Vincent Van Asch and Walter Daelemans
- 17:10–17:20 *Resolving Speculation: MaxEnt Cue Classification and Dependency-Based Scope Rules*
Erik Velldal, Lilja Øvrelid and Stephan Oepen
- 17:20–17:30 *Combining Manual Rules and Supervised Learning for Hedge Cue and Scope Detection*
Marek Rei and Ted Briscoe

Shared Task Discussion Panel (17:30-18:00)

Friday, July 16, 2010

Shared Task Session 2: Poster Session (11:00-12:30)

(Systems for both Task1 and Task2)

Hedge Detection Using the RelHunter Approach

Eraldo Fernandes, Carlos Crestana and Ruy Milidiú

A High-Precision Approach to Detecting Hedges and their Scopes

Halil Kilicoglu and Sabine Bergler

Exploiting Rich Features for Detecting Hedges and their Scope

Xinxin Li, Jianping Shen, Xiang Gao and Xuan Wang

Uncertainty Detection as Approximate Max-Margin Sequence Labelling

Oscar Täckström, Sumithra Velupillai, Martin Hassel, Gunnar Eriksson, Hercules Dalianis and Jussi Karlgren

Hedge Detection and Scope Finding by Sequence Labeling with Procedural Feature Selection

Shaodian Zhang, Hai Zhao, Guodong Zhou and Bao-liang Lu

Learning to Detect Hedges and their Scope Using CRF

Qi Zhao, Chengjie Sun, Bingquan Liu and Yong Cheng

Exploiting Multi-Features to Detect Hedges and their Scope in Biomedical Texts

Huiwei Zhou, Xiaoyan Li, Degen Huang, Zezhong Li and Yuansheng Yang

Friday, July 16, 2010 (continued)

(Systems for Task1)

A Lucene and Maximum Entropy Model Based Hedge Detection System

Lin Chen and Barbara Di Eugenio

HedgeHunter: A System for Hedge Detection and Uncertainty Classification

David Clausen

Exploiting CCG Structures with Tree Kernels for Speculation Detection

Liliana Mamani Sánchez, Baoli Li and Carl Vogel

Uncertainty Learning Using SVMs and CRFs

Vinodkumar Prabhakaran

Features for Detecting Hedge Cues

Nobuyuki Shimizu and Hiroshi Nakagawa

A Simple Ensemble Method for Hedge Identification

Ferenc Szidarovszky, Illés Solt and Domonkos Tikk

A Baseline Approach for Detecting Sentences Containing Uncertainty

Erik Tjong Kim Sang

Hedge Classification with Syntactic Dependency Features Based on an Ensemble Classifier

Yi Zheng, Qifeng Dai, Qiming Luo and Enhong Chen

