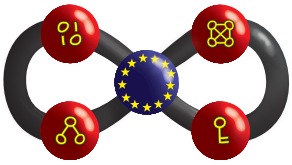
EMNLP-CoNLL 2012

Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning

Proceedings of the Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes

July 13, 2012

We wish to thank our sponsor: ETERNALS
https://www.eternals.eu/



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 978-1-937284-45-9

Introduction

This volume contains a description of the CoNLL-2012 Shared Task and the participating systems. The CoNLL-2012 shared task was on modeling multilingual unrestricted coreference in the OntoNotes data. This was an extension of the CoNLL-2011 shared task and involved automatic anaphoric mention detection and coreference resolution across three languages – English, Chinese and Arabic – using the OntoNotes 5.0 corpus, given predicted information on the syntax, proposition, word sense and named entity layers as input. The goal was to identify anaphoric mentions – both entities and events – and perform coreference resolution to create clusters of mentions representing the same entity or event in the text. The English and Chinese language portion of the OntoNotes data comprises roughly one million words per language from newswire, magazine articles, broadcast news, broadcast conversations, web data and conversational speech. The English corpus also contains a further 200k of the English translation of the New Testament. The Arabic portion is smaller, comprising 300k of newswire articles. One of the challenges for the shared task participants (though they were limited by the time constraints of the task) and also for continuing research going forward was to find effective ways to bring these multiple layers of information to bear on the coreference task to improve upon the current state of the art. An additional challenge for participants of this year's shared task was to develop systems that perform well across languages. We were happy to see many competitive systems in both English and Chinese. The results for Arabic are encouraging as well, in spite of the smaller data set.

As is traditional with CoNLL, we had two tracks – an open and a closed track. Since world knowledge is an important factor in coreference resolution, even in the closed task participants were allowed to use some limited, outside sources, including WordNet and a pre-computed table predicting number and gender information for noun phrases for the English task. This information is not available for Chinese and for Arabic due to lack of similar resources. For the open task, as usual, participants were allowed to use any other source of information, such as Wikipedia, gazetteers, etc., that did not violate the evaluation criteria designed to protect the test set. A total of 17 participants submitted system outputs and one participant withdrew because they found a bug in their system. Among the remaining 16 participants, 15 submitted system description papers. All 16 systems participated in the English task, 15 systems participated in the Chinese task and 8 systems participated in the Arabic task. There were 15 entries in the closed track and 3 in the open track. We hope that the data set of this year's shared task will provide a useful benchmark and spur further research in this important sub-field of language processing.

Sameer Pradhan, Alessandro Moschitti and Nianwen Xue
Organizers of the CoNLL-2012 Shared Task

Organizers:

Sameer Pradhan, Raytheon BBN Technologies
Alessandro Moschitti, University of Trento
Nianwen Xue, Brandeis University

Advisory Committee:

Mitchell Marcus, University of Pennsylvania
Martha Palmer, University of Colorado
Lance Ramshaw, Raytheon BBN Technologies
Ralph Weischedel, Raytheon BBN Technologies

Program Committee:

Jie Cai, HITS gGmbH
Kadri Hacioglu, Rosetta Stone
Veŕonique Hoste, University College Ghent
Dan Jurafsky, Stanford University
Sandra Kubler, Indiana University
Heeyoung Lee, Stanford University
Xiaoqiang Luo, IBM Research
Mitchell Marcus, University of Pennsylvania
Alessandro Moschitti, University of Trento
Vincent Ng, University of Texas at Dallas
Pierre Nugues, Lund University
Simone Ponzetto, University of Rome
Marta Recasens, University of Barcelona
Dan Roth, University of Illinois at Urbana-Champaign
Michael Strube, HITS gGmbH
Olga Uryupina, University of Trento
Nianwen Xue, Brandeis University

Table of Contents

Conference Program

Friday, July 13, 2012

11:00-12:30 Session I: Oral Presentation

11:00-11:30 *CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes*
Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina and Yuchen Zhang

11:30-11:45 *Latent Structure Perceptron with Feature Induction for Unrestricted Coreference Resolution*
Eraldo Fernandes, Cícero dos Santos and Ruy Milidiú

11:45-12:00 *Data-driven Multilingual Coreference Resolution using Resolver Stacking*
Anders Björkelund and Richárd Farkas

12:00-12:15 *Combining the Best of Two Worlds: A Hybrid Approach to Multilingual Coreference Resolution*
Chen Chen and Vincent Ng

12:15-12:30 *Using Syntactic Dependencies to Solve Coreferences*
Marcus Stamborg, Dennis Medved, Peter Exner and Pierre Nugues

12:30-13:45 Lunch

13:45-14:30 SIG's business meetings

14:30-15:30 Session 2: Poster Presentation

ICT: System Description for CoNLL-2012
Hao Xiong and Qun Liu

A Mixed Deterministic Model for Coreference Resolution
Bo Yuan, Qingcai Chen, Yang Xiang, Xiaolong Wang, Liping Ge, Zengjian Liu, Meng Liao and Xianbo Si

Simple Maximum Entropy Models for Multilingual Coreference Resolution
Xinxin Li, Xuan Wang and Xingwei Liao

UBIU for Multilingual Coreference Resolution in OntoNotes
Desislava Zhekova, Sandra Kübler, Joshua Bonner, Marwa Ragheb and Yu-Yin Hsu

Friday, July 13, 2012 (continued)

Chinese Coreference Resolution via Ordered Filtering
Xiaotian Zhang, Chunyang Wu and Hai Zhao

A Multigraph Model for Coreference Resolution
Sebastian Martschat, Jie Cai, Samuel Broscheit, Éva Mújdricza-Maydt and Michael Strube

Incorporating Rule-based and Statistic-based Techniques for Coreference Resolution
Ruifeng Xu, Jun Xu, Jie Liu, Chengxiang Liu, Chengtian Zou, Lin Gui, Yanzhen Zheng and Peng Qu

Illinois-Coref: The UI System in the CoNLL-2012 Shared Task
Kai-Wei Chang, Rajhans Samdani, Alla Rozovskaya, Mark Sammons and Dan Roth

System paper for CoNLL-2012 shared task: Hybrid Rule-based Algorithm for Coreference Resolution.
Heming Shou and Hai Zhao

BART goes multilingual: The UniTN / Essex submission to the CoNLL-2012 Shared Task
Olga Uryupina, Alessandro Moschitti and Massimo Poesio

Learning to Model Multilingual Unrestricted Coreference in OntoNotes
Baoli Li