

University of Houston, Computer Science Department
COSC 7397: Text Mining and its Applications, Spring 2012

Reading List

1. Adwait Ratnaparkhi. (1996). A Maximum Entropy Model for Part-Of-Speech Tagging.
2. Mikheev, A., Claire Grover, and Marc Moens. 1998. Description of the LTG system used for MUC-7. In Proceedings of the Seventh Message Understanding Conference. (on named entity recognition (NER))
3. *Locating Complex Named Entities in Web Text*, by Doug Downey and Oren Etzioni: It provides a very intuitive approach for identifying complex named entities in web documents using a simple probabilistic approach. A complex problem solved in a very simple way.

Keyphrase Extraction

- *Keyphrase Extraction. SZTERGAK paper from SemEval 2010.*
- *Conundrums in Unsupervised Keyphrase Extraction: Making Sense of the State-of-the-Art*, by K.S. Hasan and V. Ng. COLING 2010.
- 4. *Automatic Keyphrase Extraction By Bridging Vocabulary Gap*, Zhiyuan Liu et al. COLING 2011.
- 5. *Chen Wang, Sujian Li: CoRankBayes: bayesian learning to rank under the co-training framework and its application in keyphrase extraction.* CIKM 2011:2241-2244

Parsing

- Klein & Manning: "Accurate Unlexicalized Parsing" (shows that lexicalization is not necessary to achieve reasonably good parsing accuracy)
- Klein & Manning: "Corpus-Based Induction of Syntactic Structure: Models of Dependency and Constituency" (a revolution in unsupervised dependency parsing)
- Nivre "Deterministic Dependency Parsing of English Text" (shows that deterministic parsing actually works quite well)
- McDonald et al. "Non-Projective Dependency Parsing using Spanning-Tree Algorithms" (the other main method of dependency parsing, MST parsing)


Language Modeling

- Goodman "A bit of progress in language modeling" (describes just about everything related to n-gram language models)
- Teh "A Bayesian interpretation of Interpolated Kneser-Ney" (shows how to get state-of-the art accuracy in a Bayesian framework, opening the path for other applications)



Machine Learning for NLP

6. Sutton & McCallum "An introduction to conditional random fields for relational learning" (everyone should know CRFs, and this paper is the easiest to understand)
- Knight "Bayesian Inference with Tears" (explains the general idea of bayesian techniques quite well)
8. Berg-Kirkpatrick et al. "Painless Unsupervised Learning with Features" (this is from this year and thus a bit of gamble, but this has the potential to bring the power of discriminative methods to unsupervised learning)
7. Michael Collins (2002), "Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms"
- John D. Lafferty, Andrew McCallum, Fernando C. N. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. (Got test of time award at ICML 2011)





Information Extraction

- 9  Hearst. Automatic Acquisition of Hyponyms from Large Text Corpora. COLING 1992. (The very first paper for all the bootstrapping methods for NLP. It is a hypothetical work in a sense that it doesn't give experimental results, but it influenced it's followers a lot.)
- Collins and Singer. Unsupervised Models for Named Entity Classification. EMNLP 1999. (It applies several variants of co-training like IE methods to NER task and gives the motivation why they did so. Students can learn the logic from this work for writing a good research paper in NLP.)

Computational Semantics

- 10  Gildea and Jurafsky. Automatic Labeling of Semantic Roles. Computational Linguistics 2002. (It opened up the ends in NLP for semantic role labeling, followed by several CoNLL shared tasks dedicated for SRL. It shows how linguistics and engineering can collaborate with each other. It has a shorter version in ACL 2000.)
- 11  Pantel and Lin. Discovering Word Senses from Text. KDD 2002. (Supervised WSD has been explored a lot in the early 00's thanks to the senseval workshop, but a few system actually benefits from WSD because manually crafted sense mappings are hard to obtain. These days we see a lot of evidence that unsupervised clustering improves NLP tasks such as NER, parsing, SRL, etc, and this work is one of the roots of unsupervised clustering of words)

Automatic Text Summarization

- 12  J. Clarke and M. Lapata. Modeling Compression with Discourse Constraints. EMNLP-CoNLL 2007. (shows importance of joint inference)
- 13  K. Knight and D. Marcu. Summarization beyond sentence extraction. Artificial Intelligence 139, 2002. (opens the door to statistical approach to sentence compression)
- 14  R. McDonald. A Study of Global Inference Algorithms in Multi-Document Summarization ECIR 2007. (formulates summarization task as global optimization problem using integer linear programming)
- 15  W. Yih et al. Multi-Document Summarization by Maximizing Informative Content-Words. IJCAI 2007. (introduces stack decoding to this field)

Machine Translation

- Knight "A statistical MT tutorial workbook" (easy to understand, use instead of the original Brown paper in Computational Linguistics Journal 1990)
- Och "The Alignment-Template Approach to Statistical Machine Translation" (foundations of phrase based systems)
- Wu "Inversion Transduction Grammars and the Bilingual Parsing of Parallel Corpora" (arguably the first realistic method for biparsing, which is used in many systems)
- Chiang "Hierarchical Phrase-Based Translation" (significantly improves accuracy by allowing for gappy phrases)

Useful Website:

<http://www.cs.jhu.edu/~jason/405/>