

Topic: Theoretical models for NLP

Themes:

The write-up will address the following questions:

- How was the model applied to the NLP tasks?
- How was the model learned?
- What was the model performance in comparison with other approaches to the same task?
- What are the pros and cons of applying this model to the selected tasks and in general?
- What are other NLP tasks where this model has potential for successful results?

Reading list:

Maximum entropy

Introduction/Tutorial

- Adam L. Berger, Stephen A. Della Pietra, and Vincent J. Della Pietra. (1996). *A maximum entropy approach to natural language processing*. Computational Linguistics, 22(1):39–71.
 - Ronald Rosenfeld (1994) *Adaptive Statistical Language Modeling: A Maximum Entropy Approach*. Ph.D. Thesis. Chapter 4 and 5
1. Language Modeling
 - R. Rosenfeld. (1997) *A whole sentence maximum entropy language model*. In Proc. of the IEEE Workshop on Automatic Speech Recognition and Understanding, 1997
 2. Part of Speech Tagging
 - A. Ratnaparkhi (1996) *A maximum entropy model for part-of-speech tagging*. In Proceedings of Conference on Empirical Methods in Natural Language Processing
 3. Acronyms
 - Serguei Pakhomov (2002) *Semi-Supervised Maximum Entropy Based Approach to Acronym and Abbreviation Normalization in Medical Texts*. ACL
 4. Word Sense Disambiguation
 - Gerald Chao; Michael G. Dyer. (2002) *Maximum Entropy Models for Word Sense Disambiguation*. COLING
 5. Sentence Boundaries
 - Reynar, J. C. and Ratnaparkhi, A. (1997). *A Maximum Entropy Approach to Identifying Sentence Boundaries*. In Proceedings of the Fifth Conference on Applied Natural Language Processing, pages 16—19
 6. Machine Translation
 - Adam L. Berger, Stephen A. Della Pietra, and Vincent J. Della Pietra. (1996). *A maximum entropy approach to natural language processing*. Computational Linguistics, 22(1):39–71.
 - Och, F.J., Ney, H. (2002) *Discriminative training and maximum entropy models for statistical machine translation*. ACL
 - Kengo Sato; Masakazu Nakanishi. (1998) *Maximum Entropy Model Learning of the Translation Rule*. ACL

Bayesian Belief Networks

1. Selectional preferences
 - Massimiliano Ciaramita; Mark Johnson (2000) *Explaining away ambiguity: Learning verb selectional preference with Bayesian networks*. COLING
2. Word sense disambiguation
 - Gerald Chao; Michael G. Dyer (2000) *Word Sense Disambiguation of Adjectives Using Probabilistic Networks*. COLING
3. Models of human sentence processing
 - Narayanan, Sridhar, and Daniel Jurafsky. 1998. Bayesian models of human sentence processing. In Proceedings of CogSci-98
4. Discourse Structure
 - Michael Galley, Kathleen McKeown, Julia Hirschberg and Elizabeth Shriberg (2004) *Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies*. ACL
5. Dialogue Act Tagging
 - Simon Keizer; Riems op den Akker; Anton Nijholt (2002) *Dialogue Act Recognition with Bayesian Networks for Dutch Dialogue*. ACL Workshop
6. Dialogue Modeling
 - C. Wai; H.M. Meng; R. Pieraccini (2001) *Scalability and Portability of a Belief Network-based Dialog Model for Different Application Domains*. HLT
7. Question Answering
 - Ganesh Ramakrishnan; Apurva Jadhav; Ashutosh Joshi; Soumen Chakrabarti; Pushpak Bhattacharyya (2003) *Question Answering via Bayesian Inference on Lexical Relations*. ACL Workshop on Multilingual Summarization and Question Answering
8. Argumentation systems
 - Zukerman, I., McConachy, R., Korb, K. B., and Pickett, D. A. (1999). *Exploratory interaction with a Bayesian argumentation system*. In IJCAI99 - Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, pages 1294-1299, Stockholm, Sweden.
 - I. Zukerman, R. McConachy, and K. Korb, (1998) "Bayesian Reasoning in an Abductive Mechanism for Argument Generation and Analysis," AAAI.

Hidden Markov Models

Introduction/Tutorial

- Lawrence R. Rabiner (1989) *A tutorial on Hidden Markov Models and selected applications in speech recognition*. In Proceedings of the IEEE, volume 77(2), pages 257-285.
- 1. Word Segmentation
 - Hubert Hin-Cheung Law; Chorkin Chan (1996) *N-th Order Ergodic Multigram HMM for Modeling of Languages without Marked Word Boundaries*. COLING
- 2. POS
 - T. Brants (2000) *TnT - A Statistical Part-of-Speech Tagger*. ANLP
- 3. Text Chunking
 - Skut, W., Brants, T. (1998). *Chunk Tagger - Statistical Recognition of Noun Phrases*. ESSLLI.
- 4. Named Entity Recognition
 - Bikel, D. M., Schwartz, R., Weischedel, R. M. (1999). *An Algorithm that Learns What's in a Name*. Machine Learning (Special Issue on NLP).
 - GuoDong Zhou; Jian SU (2002) *Named Entity Recognition using an HMM-based Chunk Tagger*. ACL
- 5. Content structure
 - Regina Barzilay and Lillian Lee (2004) *Catching the Drift: Probabilistic Content Models, with Applications to Generation and Summarization*. HLT-NAACL
- 6. Word Alignment for Statistical Machine Translation
 - Stephan Vogel; Hermann Ney; Christoph Tillmann (1996) *HMM-Based Word Alignment in Statistical Translation*. COLING

ME, HMM and BBN Extensions

- 7. Maximum Entropy Markov Models
 - A. McCallum, D. Freitag, and F. Pereira (2000) *Maximum entropy markov models for information extraction and segmentation*. In International Conf. on Machine Learning.
 - Ciprian Chelba and Alex Acero (2004) *Adaptation of Maximum Entropy Capitalizer: Little Data Can Help a Lot*. EMNLP
- 8. Conditional Random Fields
 - John Lafferty, Andrew McCallum, and Fernando Pereira (2001) *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*. In Proc. 18th International Conf. on Machine Learning.
 - Fuchun Peng and Andrew McCallum (2004) *Accurate Information Extraction from Research Papers using Conditional Random Fields*. HLT-NAACL
 - Fei Sha; Fernando Pereira (2003) *Shallow Parsing with Conditional Random Fields*. HLT-NAACL 2003
- 9. Relational Markov Networks
 - B. Tasker, A. Pieter, and D. Koller. (2002) *Discriminative probabilistic models for relational data*. In Uncertainty in Artificial Intelligence
 - Razvan C. Bunescu, Raymond J. Mooney (2004) *Collective Information Extraction with Relational Markov Networks*. ACL