

Topic: Theoretical models for NLP

Themes:

For each model, I will select 4 to 6 representative papers. A paper will be deemed representative if the application of the model to the task at hand is interesting and complements other applications of the model. 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 pool:

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. Parsing
 - Eugene Charniak (2000) *A maximum-entropy inspired parser*. NAACL
 6. 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
 7. 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
8. User Modeling
- S. Sarawagi. (2000) *User-adaptive exploration of multidimensional data*. In Proc. of the 26th Int'l Conference on Very Large Databases (VLDB), 2000
9. Text classification
- Kamal Nigam, John Lafferty, and Andrew McCallum. Using maximum entropy for text classification. In IJCAI-99 Workshop on Machine Learning for Information Filtering, pages 61--67, 1999
- Support papers (will not be selected)
- Ratnaparkhi, Maximum Entropy Models for Natural Language Ambiguity Resolution, Ph.D. thesis, University of Pennsylvania, Philadelphia, PA, 1998
 - Mark McLauchlan (2001) *Maximum Entropy Models and Prepositional Phrase Ambiguity*. Master Thesis

Bayesian Belief Networks

10. Selectional preference
- Massimiliano Ciaramita; Mark Johnson (2000) *Explaining away ambiguity: Learning verb selectional preference with Bayesian networks*. COLING
 - Gerald Chao; Michael G. Dyer (2000) *Word Sense Disambiguation of Adjectives Using Probabilistic Networks*. COLING
11. Sentence processing
- Narayanan, Srini, and Daniel Jurafsky. 1998. Bayesian models of human sentence processing. In Proceedings of CogSci-98
12. Discourse Structure
- Michel 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
13. Dialogue Act Tagging
- Simon Keizer; Riëks op den Akker; Anton Nijholt (2002) *Dialogue Act Recognition with Bayesian Networks for Dutch Dialogue*. ACL Workshop
14. 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
15. User modeling
- Akiba, Tomoyosi; Tanaka, Hozumi (1994) *A Bayesian Approach for User Modeling in Dialogue Systems*. COLING
16. 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
17. Argumentation systems
- Zukerman, I., McConachy, R., Kórb, 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.

- Richard McConachy; Kevin B. Korb; Ingrid Zukerman (1998) *A Bayesian Approach to Automating Argumentation*. ACL Workshop on New Methods in Language Processing and Computational Natural Language Learning
- Support paper (will probably not be selected)
- POS (Dynamic BBN) - Leonid Peshkin; Avi Pfeffer; Virginia Savova (2003) *Bayesian Nets for Syntactic Categorization of Novel Word*. NAACL (probably short papers)
 - Plan Recognition (Dynamic BBN) - D. W. Albrecht, I. Zukerman, and A. E. Nicholson. (1998) *Bayesian models for keyhole plan recognition in an adventure game*. User Modeling and User-Adapted Interaction, 8:5--47, 1998
 - User Modeling - Oliver Lemon; Prashant Parikh; Stanley Peters (2002) *Probabilistic Dialogue Modelling*. SIGDIAL
 - QA - S. Narayanan, S. Harabagiu (2004). *Answering Questions using Advanced Semantics and Probabilistic Inference*, Workshop on Pragmatics of Question Answering, HLT-NAACL

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.
18. Word Segmentation
- Hubert Hin-Cheung Law; Chorkin Chan (1996) *N-th Order Ergodic Multigram HMM for Modeling of Languages without Marked Word Boundaries*. COLING
19. POS
- Chao-Huang Chang; Cheng-der Chen (1993) *HMM-Based Part-of-Speech Tagging for Chinese Corpora*. ACL Workshop on VLC
 - T. Brants (2000) *TnT - A Statistical Part-of-Speech Tagger*. ANLP
 - Cutting, D., J. Kupiec, J. Pedersen and P. Sibun. (1992) *A Practical Part-of-Speech Tagger*. ANLP
20. Named Entity Recognition
- GuoDong Zhou; Jian SU (2002) *Named Entity Recognition using an HMM-based Chunk Tagger*. ACL
 - Kuniko Saito; Masaaki Nagata (2003) *Multi-Language Named-Entity Recognition System based on HMM*. ACL Workshop on Multilingual and Mixed-language Named Entity Recognition
21. Text Chunking
- GuoDong Zhou; Jian Su (2000) *Error-driven HMM-based Chunk Tagger with Context-dependent Lexicon*. Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora
22. Parsing
- Wide R. Hogenhout; Yuji Matsumoto (1998) *Robust Parsing Using a Hidden Markov Model*. ACL Workshop on Finite State Methods in Natural Language Processing
23. WSD
- Antonio Molina, Ferran Pla and Encarna Segar (2004) *WSD system based on specialized Hidden Markov Model*. ACL SENSEVAL
24. Content structure
- Regina Barzilay and Lillian Lee (2004) *Catching the Drift: Probabilistic Content Models, with Applications to Generation and Summarization*. HLT-NAACL
25. Word Alignment for Statistical Machine Translation

- Stephan Vogel; Hermann Ney; Christoph Tillmann (1996) *HMM-Based Word Alignment in Statistical Translation*. COLING
- Kristina Toutanova; H. Tolga Ilhan; Christopher Manning (2002) *Extensions to HMM-based Statistical Word Alignment Models*. EMNLP
- 26. Information Retrieval
 - Jinxi Xu; Ralph Weischedel (2000) *Cross-lingual Information Retrieval Using Hidden Markov Models*. Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora
- 27. Phoneme-to-grapheme conversion
 - Panagiotis A. Rentzepopoulos; George K. Kokkinakis (1996) *Efficient Multilingual Phoneme-to-Grapheme Conversion Based on HMM*. CL
- 28. Summary Decomposition
 - Hongyan Jing (2002) *Using Hidden Markov Modeling to Decompose Human-Written Summaries*. CL
- HMM extensions for NLP tasks (will probably not be selected)
 - POS - Jin-Dong Kim; Hae-Chang Rim; Jun'ichi Tsujii (2003) *Self-Organizing Markov Models and Their Application to Part-of-Speech Tagging*. ACL
 - POS - H. Schutze and Y. Singer. (1994) Part-of-Speech tagging using a variable memory Markov model. In Proceedings of ACL
 - POS - Sang-Zoo Lee; Jun-ichi Tsujii; Hae-Chang Rim (2000) *Lexicalized Hidden Markov Models for Part-of-Speech Tagging*. COLING
 - Shallow Parsing - V. Punyakanok and D. Roth. (2000) *Shallow parsing by inferencing with classifiers*. In CoNLL
 - Conversion to FST - Andre Kempe (1996) *Finite State Transducers Approximating Hidden Markov Models*. ACL

ME, HMM and BBN Extensions

- 29. 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
- 30. 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
- 31. 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