Jinying Chen

University of Pennsylvania
Department of Computer and Information Science
3330 Walnut Street, Levine Hall, CIS Dept.
Philadelphia, PA, 19104
Office: (215) 573-7736
http://www.cis.upenn.edu/~jinying

Home Address:
32 Marvin Lane
Piscataway, NJ, 08854
Cell phone: (732) 668-7728
jinying@cis.upenn.edu

Education

University of Pennsylvania, Philadelphia, PA, USA Ph.D. candidate 2001 – Pres.
Department of Computer and Information Science

Dissertation: Towards High-performance Word Sense Disambiguation by Combining Rich Linguistic Knowledge and Machine Learning Approaches (to be defended in July, 2006)
Advisor: Martha S. Palmer
Committee Members: Joshi K. Aravind, Claire Cardie, (external examiner), Mitch P. Marcus (chair), Lyle H. Ungar

University of Pennsylvania, Philadelphia, PA, USA M.S. 2000 – 2001
Department of Computer and Information Science

Tsinghua University, Beijing, China M.E. 1998 – 2000, B.S. 1994 – 1998
Department of Computer Science and Technology

Research Interests

Machine learning and feature engineering for natural language processing (NLP). Automatic word sense disambiguation; clustering semantically coherent words and automatic acquisition of large-scale semantic taxonomies. NLP applications to information retrieval, information extraction, machine translation and bioinformatics.

Research Experience

Department of Computer and Information Science, University of Pennsylvania
Ph.D. Candidate ***2001 – 2006***

High performance supervised word sense disambiguation (WSD) through combining linguistically motivated features and a smoothed Maximum Entropy (MaxEnt) model. The system achieved higher accuracy than previous best systems on the SENSEVAL2 English verb data.

Unsupervised and active learning methods for WSD. EM clustering for Chinese verb senses and active learning for English verb senses. Clustering-based feature selection for WSD. Noun clustering and semi-automatically created noun taxonomies, used for semantic features for WSD.

Nominal entity detection for the Chinese Automatic Content Extraction (CACE) project (summer 2003, summer 2004). Boosting and TAG (Tree Adjoining Grammar) Supertagging for template relation detection, a subtask of the MUC-7 information extraction task (fall, 2001).

Department of Computer Science & Technology, Tsinghua University, Beijing, China
Master Student, Senior College Student ***1997 – 2000***

Visualization, dimension reduction, and classification algorithms for Chinese character recognition. A classification algorithm, based on Mahalanobis distance and dimension reduction, for distinguishing well-similar handwritten Chinese characters.

Honors and Awards

- Graudate student research fellowship from the Department of Computer and Information Science, University of Pennsylvania. Sept. 2000 – pres.
- Tsinghua-Motorola Outstanding Student Scholarship, top 3 among over 50 graduate students in the Department of Computer Science and Technology, Tsinghua University. Oct. 1999
- Honor of Excellent Student of Tsinghua University, top 10 among over 150 undergraduate students in the Department of Computer Science and Technology, Tsinghua University. Nov. 1997
- Tsinghua-Daren Chen Scholarship, top 5 among over 150 undergraduate students in the Department of Computer Science and Technology, Tsinghua University. Nov. 1996
- Honor of Excellent Student of Tsinghua University, top 10 among over 150 undergraduate students in the Department of Computer Science and Technology, Tsinghua University. Nov. 1995
- First Prize in the Tenth National High School Student Contest in Physics in Tianjin, sponsored by Chinese Physical Society and Tianjin Physical Society. Top 10 among over 1,000 competition participants in Tianjin area. Nov. 1993.

Publications

- Nianwen Xue, Jinying Chen and Martha Palmer. Aligning Features with Sense Distinction Dimensions. Submitted.
- Jinying Chen, Andrew Schein, Lyle Ungar and Martha Palmer. An Empirical Study of the Behavior of Active Learning for Word Sense Disambiguation. Accepted by *Human Language Technology conference - North American chapter of the Association for Computational Linguistics annual meeting (HLT-NAACL) 2006.* New York City.
- Jinying Chen and Martha Palmer. Clustering-based Feature Selection for Verb Sense Disambiguation. In *Proceedings of the 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE 2005)*, pp. 36-41. Oct. 30- Nov. 1, Wuhan, China, 2005.
- Jinying Chen and Martha Palmer. Towards Robust High Performance Word Sense Disambiguation of English Verbs Using Rich Linguistic Features, In *Proceedings of the 2nd International Joint Conference on Natural Language Processing (IJCNLP2005)*, pp. 933-944. Oct. 11-13, Jeju, Korea, 2005.
- Martha Palmer, Nianwen Xue, Olga B Babko-Malaya, Jinying Chen and Benjamin Snyder. A Parallel Proposotion Bank II for Chinese and English, in *Proceedings of the 2005 ACL Workshop in Frontiers in Annotation II: Pie in the Sky*, pp. 61-68. June 29, Ann Arbor, Michigan, 2005.
- Jinying Chen and Martha Palmer. Unsupervised Learning of Chinese Verb Senses by Using an EM Clustering Model with Rich Linguistic Features. In *Proceedings of the 42nd Annual Meeting of Computational Linguists (ACL-04),* pp. 295-302. July 21-23, Barcelona, Spain. 2004.
- Jinying Chen, Nianwen Xue and Martha Palmer. Using a Smoothing Maximum Entropy Model for Chinese Nominal Tagging (poster presentation), In *Proceedings of the 1st International Joint Conference on Natural Language Processing*, pp. 493-500. March 22-24, Hainan Island, China, 2004.
- Libin Shen and Jinying Chen. Using Supertag in MUC-7 Template Relation Task, *Technical Report*, MS-CIS-02-26, CIS Dept., University of Pennsylvania, 2002.

- Jinying Chen, Yijiang Jin and Shaoping Ma. The Visualization Analysis of Handwritten Chinese Characters in Their Feature Space. *Journal of Chinese Information Processing*. Vol.14, No. 5, pp42~48, 2000.
- Jinying Chen, Yijiang Jin and Shaoping Ma. A Learning Algorithm Detecting the Similar Chinese Characters' Boundary Based on Unequal-Contraction of Dimension. In *Proceedings of the 3rd World Congress on Intelligent Control and Automation*, pp. 2765-2769, vol. 4. June 28-July 02, Hefei, China, 2000.

Oral Presentations (2001-2006)

- "Towards Robust High Performance Word Sense Disambiguation by Combining Rich Linguistic Knowledge and Machine Learning Methods", in the 7th Penn Engineering Graduate Research Symposium, Feb. 15, 2006.
- "What We Learned from Supervised Word Sense Disambiguation for English Verbs" in a visit to the Center for Spoken Language Research at University of Colorado, Dec. 7, 2005.
- "Clustering-based Feature Selection for Verb Sense Disambiguation" in the 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE 2005) in Wuhan, China, Oct. 30, 2005.
- "Towards Robust High Performance Word Sense Disambiguation of English Verbs Using Rich Linguistic Features" in the 2nd International Joint Conference on Natural Language Processing (IJCNLP2005) in Jeju, Korea, Oct. 13, 2005.
- "Unsupervised Learning of Chinese Verb Senses by Using an EM Clustering Model with Rich Linguistic Features" in the 42nd Annual Meeting of Computational Linguists (ACL-04), in Barcelona, Spain, July 23, 2004.
- "Fine-grained and Coarse-grained Supervised Word Sense Disambiguation" in ARDA (Advanced Research and Development Activity)'s visit at the Computer and Information Science Department at the University of Pennsylvania, Aug. 22, 2003.

Other Professional Activities

- Organizer of the weekly seminar, the Computational Linguists' Lunch (CLUNCH), attended by about 30 faculty members and students mainly from the Department of Computer Science and Information and the Department of Linguistics, University of Pennsylvania. Spring, 2003
- Teaching assistant for the graduate-level course CIT594 II – Programming Languages and Techniques, which is oriented to master students in the Department of Computer and Information Science, University of Pennsylvania. Spring, 2002
- Teaching assistant for the graduate-level course CIS500 – Programming Languages, which is oriented to Ph.D. students in the Department of Computer and Information Science, University of Pennsylvania. Fall, 2001
- Teaching assistant for the undergraduate-level course "Introduction to Artificial Intelligence" in the Department of Computer Science and Technology, Tsinghua University. Fall, 1998
- Participation in the editorial work (collecting and editing about 200 vocabulary entries) for a major computer dictionary – *English-Chinese Dictionary of Computers and Multimedia* (published by Tsinghua University Press, 2003) in Tsinghua University under the supervision of Dr. Fuzong Lin. Summer, 1998

Reference

Joshi K. Aravind, PhD (joshi@linc.cis.upenn.edu, 215-898-8540)

Martha S. Palmer, PhD (Martha.Palmer@colorado.edu, 303-492-1300)
Lyle H. Ungar, PhD (ungar@cis.upenn.edu, 215-898-7449)