Curriculum Vitae

Jun Wu

Department of Computer Science
Johns Hopkins University
Baltimore MD 21218

Telephone: (410)516-7052
Email: junwu@cs.jhu.edu
http://www.cs.jhu.edu/~junwu

Education

Ph.D. candidate	1996 - Present	Dept. of Computer Science, Johns Hopkins University, (GPA 3.98/4.0.). Expected Date of Graduation: Summer 2002.
M.S.E	1998	Johns Hopkins University, Computer Science.
M.E	1993	Tsinghua University (China), Electronic Engineering.
B.E	1989	Tsinghua university (China), Computer Science.

Professional Experience

September 1996-Present	**Research Assistant** in Center for Speech and Language Processing, Johns Hopkins University. Participated in the NSF STIMULATE project. Exploiting non-local dependencies in language modeling for speech recognition. Proposed maximum entropy language models with syntactic and topic dependencies (with Sanjeev Khudanpur). Successfully applied this model to speech recognition and obtained significantly improvement in performance. Proposed fast training algorithms for maximum entropy models. Proposed a classifier combination method for part-of-speech tagging (with Eric Brill). **Teaching Assistant** in courses Natural Language Processing, Computer Architecture, Computer Systems and Computer Networks, Johns Hopkins University.
June - August 1997	**Intern**, AT&T Labs-Research, Florham Park, NJ. Participated in the development of a speech recognition system for telephone dialogues. (Mentors: Roberto Pieraccini and Esther Levin.)
July 1993 - August 1996	**Lecturer and Research Associate**, Department of Electronic Engineering, Tsinghua University, China. Principle investigator of the project on telephone speech recognition of Mandarin Chinese sponsored by Nortel Research. Individual investigator of the project on the dictation machine sponsored by National 863 Foundation of China.
May 1993 - May 1995	**Part time consultant**, Avante Electronics Corp., China. Worked as the project leader for the JANET system, a large vocabulary Mandarin Chinese speech recognition and handwriting recognition system.
July 1989 - August 1991	**Software Engineer** in Chinese Educational Electronics Corporation, China.

Honors and Awards

The ELSNET Prize for the Best Student Paper of Eurospeech99 (1999).

Honor for Excellent Young Faculty Members, by Tsinghua, China (1996).
Honor for Excellent Young Faculty Members, by Tsinghua, China (1995).
Best Paper Awards of the 3rd National Conference on Man-Machine Speech Communication, China (1994).
Excellent Graduate Awards, by Tsinghua (1993).
Chia-Chiao Lin Applied Mathematics Awards, Tsinghua University (1993).
Motorola fellowship, Tsinghua University, China (1992).

Target

Looking for a permanent research/development position in industry (perfer research labs).

Skills

Programming languages: C++, C, Perl, Fortran, Pascal, Lisp, Prolog, Assembly (X86 and MIPS).
Operating systems: Unix, Windows95/98/NT, DOS and more.
Other skills: Oracle/SQL, HTML.

Research Interests

Speech Recognition, Pattern Recognition.
Natural Language Processing.
Machine Learning.
Statistical Modeling.

Affiliation

Member of IEEE, Society of Signal Processing.
Member of International Speech Communication Association (ISCA).

Selected Publications

1. **Jun Wu** and Sanjeev Khudanpur, 'Building A Topic-Dependent Maximum Entropy Language Model for Very Large Corpora'", To appear in ICASSP2002.

2. Woosung Kim, Sanjeev Khudanpur and **Jun Wu**, "Smoothing Issues in the Structured Language Model," Eurospeech 2001, September 2001, Denmark.

3. **Jun Wu** and Sanjeev Khudanpur, "Efficient Training Methods for Maximum Entropy Language Modeling", Proceedings of ICSLP2000, Vol.3. pp 114-117, Oct. 2000, Beijing, China

4. Sanjeev Khudanpur and **Jun Wu** "Maximum Entropy Techniques for Exploiting Syntactic, Semantic and Collocational Dependencies in Language Modeling". Computer Speech and Language, pp. 355-372, Oct. 2000.

5. **Jun Wu** and Sanjeev Khudanpur, "Syntactic Heads in Statistical Language Modeling", Proceedings of ICASSP2000, pp. 1699-1702, June 4-9 2000, Istanbul, Turkey.

6. **Jun Wu** and Sanjeev Khudanpur, "Combining Nonlocal, Syntactic and N-Gram Dependencies in Language Modeling". Proceedings of Eurospeech'99, vol 5, pp2179-2182, September 6-10, 1999, Budapest, Hungary. **Winner Eurospeech99 ELSNET Best Student Paper Award.**

7. Sanjeev Khudanpur and **Jun Wu**, "A Maximum Entropy Language Model to Integrate N-Grams and Topic Dependencies for Conversational Speech Recognition". Proceedings of ICASSP'99, pp. 553-556, March 14-19, 1999, Phoenix.

8. Eric Brill and **Jun Wu**, "Classifier Combination for Improved Lexical Disambiguation", Processings of COLING-ACL'98, pp 191-195, August 10-14, 1998, Montreal Canada.

9. **Jun Wu** and Zuoying Wang, "Entropy of Chinese and the Perplexity of the Language Models", ACTA Electronica Sinica, v24 n10 Oct 1996, p69-71.

10. Zuoying Wang, **Jun Wu**, et al., "Methods Towards the Very Large Vocabulary Chinese Speech Recognition", Eurospeech' 95, pp.215-218, 1995.9, Madrid, Spain.

11. **Jun Wu**, Zuoying Wang, Jiasong Sun, Jin Guo, "Chinese Speech Understanding and Spelling-Word Translation Based on the Statistics of Corpus", ICSLP'94 , September 18-22, Yokohama, Japan.

12. **Jun Wu**, Jiasong Sun, Huizhong Yang and Yan Zhang, "IBM OS/2 Programming Guide" (Translation from English edition to Chinese edition), Tsinghua Press, 1994.

Other Publications

13. **Jun Wu**, Zuoying Wang and Jiasong Sun, "A Method of Speech Understanding for Chinese Speech Recognition", Chinese Software Journal, Special Issue, Oct 1996, pp188-193

14. **Jun Wu**, et al., "Entropy for Chinese Text", Chinese Information Journal, Vol. 6 N.2, pp. 17-23, 1996.

15. **Jun Wu**, Z. Wang, Y. Feng, "Automatic Classification of Chinese Texts", Chinese Information Journal, Vol. 5 N.4, pp. 34-41, 1995.

16. **Jun Wu**, Z. Wang, Y. Ren, "Stochastic Language Models for Chinese Speech Recognition Based on Chinese Spelling", 1994 International Symposium on Speech, Image Processing and Neural Networks, pp. 674-677., April 1994, HongKong.

17. **Jun Wu**, et al., "Large Vocabulary Telephone Speech Recognition System", The 4th National Conference on Man-Machine Speech Communication, 1996, 10, Beijing China.

18. **Jun Wu**, Z. Wang, et al., "A Strategy for Large Vocabulary Continuous Speech Recognition", Proc. of ICCC, Singapore, June, 1996.

19. **Jun Wu**, Z. Wang, J. Yu, "Research on the Perplexity of Language Models", The 2nd National Conference on Intelligent Computer Interface and Applications of Artificial Intelligence, pp. 169-174, July, 1995, Weihai, China.

20. **Jun Wu**, Z. Wang, J. Sun, J. Guo, "An Approach toward Speech Understanding and Spelling-word Translation", The 3rd National Conference on Man-Machine Speech Communication, pp. 214 -220, 1994.10, Chongqing, China.

21. Zuoying Wang, **Jun Wu**, et al., "JANET - A New Chinese Speech Recognition, Understanding, and Synthesis System", The 3rd National Conference on Man-Machine Speech Communication, pp.482-486, 1994.10, Chongqing, China.

22. **Jun Wu**, Z. Wang, Y. Ren, "A New Approach to Speech Understanding Based on Corpus", The 2nd National Conference of Computer Linguistics, pp. 96-101, Nov. 1993. Xiamen China.

23. **Jun Wu**, Z. Wang, "Stochastic Language Models for Eliminating Chinese Speech Recognition Error", The 1st National Conference of Intelligent Computer Interface and Application of Artificial Intelligence, pp. 233-238, July 22-24, 1993, Jinbo Lake, China.

Presentations/Seminars

Speechworks, Boston, July 2001.
AT&T Shannon Labs of Research, July 2001.
Speechworks, New York, June 2001.
IBM Almaden Research Labs, June 2001.
IBM T.J. Watson Research Labs, May 2001.
Johns Hopkins University, April 2001.
ICSLP, Beijing, October, 2000.
Nuance, August 2000.
SRI International, August 2000.
ICASSP, Istanbul, June, 2000.
IBM T.J. Watson Research Labs, April 1999.

References

Sanjeev Khudanpur, sanjeev@clsp.jhu.edu, 410-516-7052.
David Yarowsky, yarowsky@cs.jhu.edu, 410-516-5372.
Frederick Jelinek, jelinek@clsp.jhu.edu, 410-516-7730
Gerald Masson, masson@jhu.edu, 410-516-7013