

Yinggong Zhao

CONTACT

Room 911, Building of Computer Science and Technology

INFORMATION

Mobile: +86-13921426252

Department of Computer Science and Technology

Email: zero1986618@gmail.com or zhaoyg@nlp.nju.edu.cn

Nanjing University

No.163, XianLin Avenue, Nanjing, Jiangsu Province, China, 210046

RESEARCH

Natural Language Processing

INTERESTS

Statistical Machine Translation

EDUCATION

Department of Computer Science and Technology

Nanjing University

Doctoral Student

Sept 2008 - Present

◆ Master-Doctorate Program

Bachelor of Science

Sept 2003 – Jun 2008

◆ Admitted to grad school w/o entrance exam

HONOURS

First Prize of Jiangsu Province, National Mathematical Contest in Modeling 2005

AWARDS

Second Class People's Scholarship, 2006

Second Class People's Scholarship, 2007

Pan XuePing Encouragement Scholarship, 2007

Dongshi Dongfang Scholarship, 2011

PUBLICATIONS

Yinggong Zhao, Yangsheng Ji, Ning Xi, Shujian Huang and Jiajun Chen. Language Model Weight Adaptation Based on Cross-entropy for Statistical Machine Translation. Accepted by PACLIC 2011.

Yinggong Zhao, Shujie Liu, Yangsheng Ji and Jiajun Chen. Transductive Minimum Error Rate Training for Statistical Machine Translation. To appear on IJCNLP 2011.

Henry Li, Nan Duan, **Yinggong Zhao**, Shujie Liu, Lei Cui, M. Hwang, A. Axelrod, Jianfeng Gao, Y. Zhang, and Li Deng. The MSRA Machine Translation System for IWSLT-2010, In Proc. the 7th International Workshop on Spoken Language Translation (IWSLT2010), Paris, France, December 2010

Mu Li, **Yinggong Zhao**, Dongdong Zhang and Ming Zhou. 2010. Adaptive Development Data Selection for Log-linear Model in Statistical Machine Translation. In Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)

Shujian Huang, Ning Xi, **Yinggong Zhao**, Xinyu Dai, and Jiajun Chen. 2009. An Error-Sensitive Metric for Word Alignment in Phrase-based SMT. Journal of Chinese Information Processing, 23(3)

PROFESSIONAL

Teaching Assistant for Numerical Computing, Department of Computer

EXPERIENCE

Science and Technology, Nanjing University

Feb 2011 – Jun 2011

Teaching Assistant for C++ Programming language, Department of Computer Science and Technology, Nanjing University *Sept 2008 – Jan 2009*
Internship, Microsoft Research Asia, NLC group *Dec 2009 – Sept 2010*

PROJECT & RESEARCH

EXPERIENCE

Bilingual Subtitle Corpus Mining *April 2010 – June 2010*

Description: Mining English-Chinese parallel data from movie subtitle files downloaded from website. The final bilingual dataset contains totally 17 million parallel sentences. In this work, I also implemented a sentence alignment tool in C# based on Champollion, which also incorporates many components like tokenization and Chinese segmentation.

Sentence Punctuation Prediction *June 2010 – Aug 2010*

Description: This task is for IWSLT evaluation and Speech Translation task. I try use different features and tool. And I apply a two-pass approach that could improve the prediction accuracy and final SMT performance.

Participation in IWSLT2010 *June 2010 – Sept 2010*

Description: In this evaluation task, I worked on English-Chinese part, including data processing, translation, named entity translation, MBR re-ranking and system combination. Finally E-C part won 2nd in ASR part and 4th in CER part.

Implementation of Hierarchical phrase-based Decoder *Sept 2010 – Nov 2010*

Description: implement a hierarchical decoder under hypergraph framework, including both cube-pruning and cube-growing algorithm. Besides, I also implement other related components, including recaser(case restore for English), phrase&rule extract and a tool that convert ARPA format language model file into binary format.

Participation in CWMT2011 *June 2011 – Aug 2011*

Description: In this evaluation task, I am responsible for all the tracks we participate, including Chinese-English, English-Chinese and Japanese-Chinese. My work including data processing, named entity recognition and translation, model training and final evaluation. Our system ranks 2nd in Japanese-Chinese track, 3rd in English-Chinese track and 4th in Chinese-English track.

Implementation of String-dependency Decoder *Jun 2011 – Now*

Description: implement a string-dependency decoder based on Shen et al(2008). It is on progress and about 80% of total work is completed now.

PROGAMMING

C, C++, C#, JAVA, Perl, Matlab, Latex

COURSE TAKEN

Graduate

Data Mining, Distributed Systems, Introduction to the Theory of Computation, Computer Graphics, Embedded System, Natural Language Processing

Undergraduate

Principle of Compiler, Operating System, Data Structure, Mathematical Logic, Discrete Mathematics, Introduction to Programming, Probability and Statistics, Digital Image Processing, Introduction to Database, Software Engineering