

Yi Su

20323 Pinntage Pkwy
Cupertino, CA 95014
(408) 772-4021

nuance@gmail.com
<http://nuance1979.bitbucket.io>

EDUCATION ◇ Johns Hopkins University, Baltimore, MD

- Ph.D. in Electrical and Computer Engineering Apr 2009
Thesis title: *Knowledge Integration into Language Models: A Random Forest Approach*.
Thesis advisor: Prof. Frederick Jelinek

- M.S.E. in Electrical and Computer Engineering May 2004

◇ Tsinghua University, Beijing, China

- B.E. in Computer Science and Technology Jul 2000
Thesis title: *Rule-based Semantic Analysis in Spoken Language System* (in Chinese).
Thesis advisor: Dr. Thomas Fang Zheng

EXPERIENCE ◇ Principal Research Engineer

July 2022 – Present

Apple Inc., Cupertino, CA

- Working on Siri perception.

◇ Staff Algorithm Engineer

Mar 2017 – July 2022

Ant Group, Sunnyvale, CA

- Architected from scratch, then led a small team to further develop, a low-code chatbot development platform based on our proprietary chatbot engine as the primary technical product of our team.
- Contributed, as a core member, to the development of our chatbot engine, powering the intelligent assistant for Alipay, the customer service chatbot for Alipay, the financial advisory chatbot for Ant Fortune and the interbank lending bot for MYbank, among others.
- Architected a reinforcement learning library for dialogue policy modeling based on *tianshou*, an open source general reinforcement learning library, consolidating team's previous work into a reusable package.
- Contributed 10 reinforcement learning model/technique implementations upstream to *tianshou*.
- Developed and deployed a highly efficient model for few-shot intent classification for Ant Fortune chatbot.
- Built a Python client library for our company's key management service to support internal open source effort of our codebases; the library itself is also recognized as one of the first internal open source projects.
- Organized a bi-weekly paper reading group to help the team stay current with the latest research.

- Mentored a summer intern on building a user simulator for dialog policy learning, which led to a successful hire.
- ◇ **Machine Learning Engineer** Nov 2015 – Feb 2017
Apple Inc., Cambridge, MA
- Built a large-scale n -gram language model trainer based on Apache Spark.
 - Contributed to launching iOS voicemail transcription feature by building custom language models.
 - Mentored Christophe J. Van Gysel on an internship project about a parameter-efficient way of learning word representations.
- ◇ **External Co-Advisor** Mar 2014 – Dec 2016
McGill University, Montréal, QC, Canada
- Co-advised Marc-Antoine Rondeau, a PhD student in the ECE department of McGill University to graduation, with Prof. Rick Rose.
 - Thesis title: *Neural Conditional Random Fields for Natural Language Understanding*.
- ◇ **Research Scientist Manager** Dec 2013 – Oct 2015
Nuance Communications, Inc., Montréal, QC, Canada
- Lead a team of 4 working on language modeling for Dragon TV, a commercial online service of speech recognition for smart TVs, releasing to production up to 10 new models every 3 months with provable improvements.
 - Lead the early adoption of deep learning models for natural language understanding, including domain classification and named entity recognition tasks.
 - Transitioned Nina, a virtual assistant for customer service, from grammar-based to statistical language model.
- ◇ **Senior Research Scientist** Feb 2009 – Nov 2013
Nuance Communications, Inc., Montréal, QC, Canada
- Improved discriminative language modeling recipe and tools, leading to significant gain in accuracy for our flagship products Dragon NaturallySpeaking, the leading desktop dictation software.
 - Built statistical models to improve list item formatting for PowerScribe 360, the leading radiology voice recognition software.
 - Developed a novel Bayesian class-based language model, inspired by LDA and random forest, with surprisingly good performance.
 - Participated Nuance Research Conference, an annual internal technology conference modeled after Interspeech/ICASSP, for *every single year*, giving multiple presentations.
- ◇ **Summer Intern** May 2008 – Aug 2008
Microsoft Research, Redmond, WA
- Developed a boosting algorithm for adapting a web search ranking model from high-resource languages to lower-resource languages.
- ◇ **Graduate Student Member**, CLSP Summer Workshop Jun 2004 – Aug 2004
Johns Hopkins University, Baltimore, MD

- Investigated various adaptation methods in building a Wu-dialectal Chinese speech recognition system using limited in-domain data.

◇ **Summer Intern** Jul 2002 – Aug 2002
d-Ear Technologies, Beijing, China

- Developed *d-Ear Pen*, a whole sentence Chinese input method engine based on n -gram language model on Palm OS.

◇ **Visiting Student**, Natural Language Computing Group Apr 2000 – Jul 2000
Microsoft Research China (now Microsoft Research Asia), Beijing, China

- Contributed to the team's participation of TREC 2000 evaluation by installing SMART, the Cornell vector-based information retrieval system on a Linux server.

PUBLICATION ◇ Shangshang Zheng, He Bai, Yizhe Zhang, **Yi Su**, Xiaochuan Niu, Navdeep Jaitly. 2024. KGLens: A Parameterized Knowledge Graph Solution to Assess What an LLM Does and Doesn't Know. In submission.

◇ Pranay Dighe, **Yi Su**, Shangshang Zheng, Yunshu Liu, Vineet Garg, Xiaochuan Niu, Ahmed Tewfik. 2024. Leveraging Large Language Models for Exploiting ASR Uncertainty. In *Proceedings of ICASSP*.

◇ Jiayi Weng, Huayu Chen, Dong Yan, Kaichao You, Alexis Duburcq, Minghao Zhang, **Yi Su**, Hang Su, Jun Zhu. 2022. Tianshou: A Highly Modularized Deep Reinforcement Learning Library. In *Journal of Machine Learning Research 23*.

◇ Xiang Hu, Haitao Mi, Zujie Wen, Yafang Wang, **Yi Su**, Jing Zheng and Gerard de Melo. 2021. R2D2: Recursive Transformer based on Differentiable Tree for Interpretable Hierarchical Language Modeling. In *Proceedings of ACL-IJCNLP*.

◇ Marc-Antoine Rondeau, **Yi Su**. 2016. LSTM-Based NeuroCRFs for Named Entity Recognition. In *Proceedings of Interspeech*.

◇ Marc-Antoine Rondeau, **Yi Su**. 2015. Recent Improvements to NeuroCRFs for Named Entity Recognition. In *Proceedings of ASRU*.

◇ Marc-Antoine Rondeau, **Yi Su**. 2015. Full-Rank Linear-Chain NeuroCRF for Sequence Labeling. In *Proceedings of ICASSP*.

◇ **Yi Su**. 2011. Bayesian Class-Based Language Models. In *Proceedings of ICASSP*.

◇ Anoop Deoras, Frederick Jelinek, **Yi Su**. 2010. Language Model Adaptation Using Random Forests. In *Proceedings of ICASSP*.

◇ Jianfeng Gao, Qiang Wu, Chris Burges, Krysta Svore, **Yi Su**, Nazan Khan, Shalin Shah, Hongyan Zhou. 2009. Model Adaptation via Model Interpolation and Boosting for Web Search Ranking. In *Proceedings of EMNLP*.

◇ **Yi Su**, Frederick Jelinek. 2008. Exploiting Prosodic Breaks in Language Modeling with Random Forests. In *Proceedings of Speech Prosody*.

◇ Jia Cui, **Yi Su**, Keith Hall, Frederick Jelinek. 2007. Investigating Linguistic Knowledge in a Maximum Entropy Token-Based Language Model. In *Proceedings of ASRU*.

◇ **Yi Su**, Frederick Jelinek, Sanjeev Khudanpur. 2007. Large-Scale Random Forest Language Models for Speech Recognition. In *Proceedings of Interspeech*.

- ◇ Yanli Zheng, Richard Sproat, Liang Gu, Izhak Shafran, Haolang Zhou, **Yi Su**, Dan Jurafsky, Rebecca Starr and Su-Youn Yoon. 2005. Accent Detection and Speech Recognition for Shanghai-Accented Mandarin. In *Proceedings of Interspeech*.
 - ◇ **Yi Su**, Wenhui Wu, Fang Zheng, Ditang Fang. 2001. Study on Support Vector Machine-based Speech Recognition. In *Proceedings of NCMMS-6* (in Chinese).
 - ◇ **Yi Su**, Fang Zheng, Yinfei Huang. 2001. Design of a Semantic Parser with Support to Ellipsis Resolution in a Chinese Spoken Language Dialogue System. In *Proceedings of Eurospeech*.
 - ◇ Jianfeng Gao, Jian-Yun Nie, Jian Zhang, Endong Xun, **Yi Su**, Ming Zhou, Changning Huang. 2000. TREC-9 CLIR Experiments at MSRCN. In *Proceedings of TREC-9*.
- PATENTS
- ◇ Zhiheng Zhou, Peng Xu, Jing Zheng, **Yi Su**, Zhongnan Shen. 2023. Methods and Apparatuses for Generating Dialogue Annotation Data. US Patent Pending. Application No. 18/363,504.
 - ◇ Christophe J. Van Gysel, **Yi Su**, Xiaochuan Niu, Ilya Oparin. 2017. Rank-reduced token representation for automatic speech recognition. US Patent No. 10,593,346.
 - ◇ Paul Vozila, Maximilian Bisani, **Yi Su**, Stephen Chu, Stanley Chen, Ruhi Sarikaya, Bbuvana Ramabhadran. 2011. Word-classing for language modeling. US Patent No. 9,367,526.
 - ◇ **Yi Su**. 2010. Class-based language model and use. US Patent No. 8,812,299.
 - ◇ Jianfeng Gao, **Yi Su**, Qiang Wu, Chris Burges, Krysta Svore, Elbio Abib. 2008. Boosting algorithm for ranking model adaptation. US Patent No. 8,255,412.
- SKILLS
- ◇ C/C++, Python, Bash, Swift, Perl, C#, Java, MATLAB;
 - ◇ PyTorch, Tensorflow, SRI LM Toolkit, HTK, SMART;
 - ◇ Apache Spark, iOS Programming;
 - ◇ Fluent in spoken/written English, native in Mandarin Chinese.
- HONORS
- ◇ **Awards**
 - Ant Group Inner Source Pioneer Award (2021)
 - Nuance Research Collaboration Award (2015)
 - ◇ **Grants**
 - ISCA Student Travel Grant for Speech Prosody (2008)
 - ◇ **Scholarships**
 - Tsinghua Practice Scholarship for outstanding graduates (2001)
 - Tsinghua - Seagate&Weishi Scholarship for outstanding undergraduates (1999)
 - Tsinghua First Class Scholarship for outstanding undergraduates (1998)
 - Tsinghua - Haolaixi Scholarship for outstanding freshmen (1996)
- REFERENCE Available upon request.