Yi Yang

478 Lindbergh Pl NE, Atlanta, GA 30324
www.cc.gatech.edu/~yyang319/

Education

Doctor of Philosophy — Expected in December 2016
School of Interactive Computing, College of Computing — GPA: 3.92/4.0
Georgia Institute Technology, Atlanta, GA

Bachelor of Science — Sep.2007 - Jun.2011
School of Software — GPA: 3.70/4.0, Rank: 2/144
Beihang University, Beijing, China

Research Experience

Graduate Research Assistant — Jan. 2013 - present
Computational Linguistics Lab — Advisor: Jacob Eisenstein
School of Interactive Computing
Georgia Institute of Technology, Atlanta, GA

- Socially Adapted Natural Language Processing
  - Community-specific embedding projections for Twitter sentiment analysis.
  - Sentiment analysis with social attention.
  - Socially-infused information extraction by embedding users, mentions, and entities.
- Unsupervised Domain Adaptation for Structured Prediction
  - Feature embedding for unsupervised multi-domain adaptation.
  - Representation learning through marginalized denoising autoencoders with structured dropout noise.
  - Part-of-Speech tagging for historical English.
- Token-based Unsupervised Text Normalization
  - Defined a log-linear model for unsupervised text normalization.
  - Approximated target distributions with Sequential Monte Carlo.

Work Experience

Research Intern — May. 2015 - Aug. 2015
Machine Learning Department — Mentor: Scott Wen-tau Yih
Microsoft Research Redmond, WA

- Presented WikiQA, a new dataset for open-domain question answering.
- Compared several deep learning algorithms for answer sentence selection.

Research Intern — May. 2014 - Aug. 2014
Internet Services Research Center (ISRC) Group — Mentor: Ming-Wei Chang
Microsoft Research Redmond, WA

- Proposed Structured Multiple Additive Regression Trees (S-MART) for fast non-linear structured learning.
- Proposed a global joint model for tweet entity linking.

Research Scientist Intern — May. 2013 - Jul. 2013 & May. 2012 - Aug. 2012
Speech Group & NLU Group — Mentor: Bjoern Hoffmeister & Imre Kiss
Amazon, inc. Seattle, WA

- Improved confusion network decoding for Automatic Speech Recognition (ASR).
- Improved confidence modeling for language understanding with ASR output.
- Proposed a SMT based approach for named entity recognition.

Publications

Yi Yang and Jacob Eisenstein. Overcoming Language Variation in Sentiment Analysis with Social Attention. *Accepted to Transactions of the Association for Computational Linguistics (TACL), 2016* (Full Paper)

Yi Yang Ming-Wei Chang and Jacob Eisenstein. Toward Socially-Infused Information Extraction: Embedding Authors, Mentions, and Entities. *To appear in Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016* (Full Paper)

Yi Yang and Jacob Eisenstein. Part-of-Speech Tagging for Historical English. *In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2016* (Oral Full Paper)

Yi Yang and Jacob Eisenstein. Putting Things in Context: Community-specific Embedding Projections for Sentiment Analysis. Technical Report, 2015

Yi Yang, Wen-tau Yih and Christopher Meek. WikiQA: A Challenge Dataset for Open-Domain Question Answering. *In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015* (Short Oral Paper)

Yi Yang and Ming-Wei Chang. S-MART: Novel Tree-based Structured Learning Algorithms Applied to Tweet Entity Linking. *In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL), 2015* (Full Oral Paper)

Yi Yang and Jacob Eisenstein. Unsupervised Multi-Domain Adaptation with Feature Embeddings. *In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2015* (Full Paper)

Yi Yang and Jacob Eisenstein. Unsupervised Domain Adaptation with Feature Embeddings. *In Proceedings of the 3rd International Conference on Learning Representations (ICLR), 2015* (Workshop Paper)

Yi Yang and Jacob Eisenstein. Fast Easy Unsupervised Domain Adaptation with Marginalized Structured Dropout. *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014.* (Short Paper)

Yi Yang and Jacob Eisenstein. A Log-Linear Model for Unsupervised Text Normalization. *In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP), 2013.* (Full Oral Paper)

Yi Yang and Lambert Mathias. A statistical machine translation approach to entity recognition. *In Proceedings of Amazon Conference on Machine Learning, 2013*

Fangtao Li, Minlie Huang, **Yi Yang** and Xiaoyan Zhu. Learning to Identify Review Spam. *In Proceedings of the Twenty-Second International Joint Conference On Artificial Intelligence (IJCAI), 2011.* (Full Oral Paper)

Minlie Huang, **Yi Yang** and Xiaoyan Zhu. Quality-biased Ranking of Short Texts in Microblogging Services. *In Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP), 2011* (Full Oral Paper)

Rezarta Islamaj Doan, **Yi Yang**, Aurlie Nvol, Minlie Huang and Zhiyong Lu. Identifying protein-protein interactions in biomedical text articles. *In Proceedings of the Third BioCreative Challenge Evaluation Workshop, 2010*

Selected Talks

Invited Talk, June 2016. "Toward Robust Text Analysis for Language Variation" *University of Pennsylvania; Bloomberg L.P.; Baidu Research; Google Research*

Conference Talk, June 2016. "Part-of-Speech Tagging for Historical English." *NAACL 2016*

Invited Talk, May 2016. "Efficiently Learning and Applying Dense Feature Representations for Natural Language Processing" *IBM research; Allen Institute for Artificial Intelligence*

Guest lecture, December 2015. "Domain Adaptation for Natural Language Processing." *CS 4650/7650, Natural Language Processing*

Conference Talk, July 2015. "S-MART: Novel Tree-based Structured Learning Algorithms Applied to Tweet Entity Linking." *ACL 2015*

Guest lecture, March 2015. "Deep Learning of Representations for Domain Adaptation." *CS8803DL, Deep Learning for Perception*

Intern Talk, August 2014. "Improved Tweet Entity Linking with Structured Gradient Tree Boosting." *NLP reading group, MSR; ISRC group meeting, MSR*

Conference Talk, October 2013. "A Log-Linear Model for Unsupervised Text Normalization." *EMNLP 2013*

Intern Talk, August 2012. "A statistical machine translation approach to entity recognition." *Amazon Tech Talk series*

Teaching Experience

- Computational Journalism — Spring 2016
- Intro to Enterprise Computing — Spring 2013
- J2EE Architecture — Spring 2011
- Java Programming Language — Fall 2010
- Algorithm Analysis & Design — Fall 2009

Computer Skills

- **Programming Language:** Python, Java, Matlab, C++, C#, R
- **Web Technologies:** HTML, CSS, ASP, JavaScript, JSP
- **Databases:** SQL Server, Oracle, MySql
- **Tools:** Theano, Hadoop, Mallet, Indri, Lucene

Professional Services

- Program Committee Member: EACL 2017, RepL4NLP 2017, ACL 2016, WNUT 2016, SocialNLP 2016, Workshop on Human-Computer QA 2016, EMNLP 2015, ACL 2015, ACL 2014
- Conference Review: CONLL 2014, EMNLP 2013, ICDCS 2012, WWW 2011

Selected Honors

- ACL Student Volunteer — 2015
- Georgia Tech GSC Travel Grant — 2015
- ACL Outstanding Reviewer — 2015
- NAACL Student Volunteer — 2015
- ACL Student Travel Awards — 2014
- EMNLP Student Volunteer — 2013
- Excellent Student of Beihang University (Top 1%) — 2010
- First Prize for Academic Excellent Student (Top 2%), Beihang University — 2010
- Ranked 7th in ACM Programming Contest, Beihang University — 2009
- First Prize for Innovative Student (Top 2%), Beihang University — 2009
- National Scholarship (Top 2%), P.R.China — 2009