

Diyi Yang

Assistant Professor
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA 30332

Phone: (412) 980-8543
Email: diyi.yang@cc.gatech.edu
Homepage: www.cc.gatech.edu/~dyang888

Education

- Ph.D., Computer Science - Language Technologies, **Carnegie Mellon University** 2015 - 2019.
Advisors: Robert Kraut and Eduard Hovy
Additional thesis committee members: Brandy L. Aven, Dan Jurafsky.
- Master in Language Technologies, **Carnegie Mellon University** 2013 - 2015.
- B.S. Computer Science, **Shanghai Jiao Tong University** 2009 - 2013.

Academic Employment

- Georgia Tech, Assistant Professor, School of Interactive Computing, 2019 Aug 1st - Present

Research Interests

- Computational Social Science (Online Communities)
- Natural Language Processing (Semantics, Discourse, Pragmatics, Text Generation)

Awards & Honors

- 30 Influential Women Advancing AI, RE·Work, 2019
- Best Reviewer Award, ICWSM, 2019
- Best Paper Honorable Mention, SIGCHI, 2019
- Selected Rising Stars in EECS, Stanford, 2017
- Facebook PhD Fellowship, 2017-2019
- Presidential Fellowship, CMU, 2016
- Best Paper Honorable Mention, ICWSM, 2016
- Inaugural Notable Data Set Award, EMNLP, 2015
- Google Anita Borg Scholarship, 2012
- Champion in KDD Cup, Track 1 (Team ACMClass@SJTU), 2012
- Best Paper Honorable Mention, International Conference on Web intelligence, 2012

Experience

- **Visiting Researcher** at *Google Inc.*
 - Google AI, Language Team, advised by Dipanjan Das Mar 2019 – July 2019
- **Visiting Researcher** at *Stanford University*
 - NLP Group, advised by Dan Jurafsky May 2018 – Aug 2018
- **Research Intern** at *Facebook Research*
 - Protect and Care Team, advised by Ryan Ritter and Mark Handel May 2017 – Aug 2017
- **Research Intern** at *Microsoft Research Redmond*
 - Social Computing Group, advised by Scott Counts May 2016 – Aug 2016
- **Research Intern** at *Wikimedia Foundation*
 - Wikimedia Foundation, advised by Aaron Halfaker June 2015 – Sep 2015
- **Research Intern** at *Microsoft Research Asia*
 - Artificial Intelligence Group, advised by Bin Gao and Tie-Yan Liu July 2012 – Feb 2013

Publications

Peer-Reviewed Conference

47. *Automatically Neutralizing Subjective Bias in Text*
Reid Pryzant, Richard Diehl Martinez, Nathan Dass, Sadao Kurohashi, Dan Jurafsky, **Diyi Yang**
The 34th AAAI Conference on Artificial Intelligence. AAAI 2020
46. *Proximate Social Factors in First-Time Contribution to Online Communities*
Joseph Seering, Jessica Hammer, Geoff Kaufman, **Diyi Yang**
The ACM CHI Conference on Human Factors in Computing Systems. SIGCHI 2020
45. *Multi-level Modeling of Social Roles in Online Micro-lending Platforms*
Lu Sun, Robert Kraut, and **Diyi Yang**
Proceedings of the ACM Conference on Computer-Supported Cooperative Work. CSCW 2019.
44. *Let's Make Your Request More Persuasive: Modeling Persuasive Strategies via Semi-Supervised Neural Nets on Crowdfunding Platforms*
Diyi Yang, Jiaao Chen, Zichao Yang, Dan Jurafsky, and Eduard Hovy
The 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics. NAACL 2019
43. *Seekers, Providers, Welcomers, and Storytellers: Modeling Social Roles in Online Health Communities*
Diyi Yang, Robert Kraut, Tenbroeck Smith, Elijah Mayfield, and Dan Jurafsky
The ACM CHI Conference on Human Factors in Computing Systems. CHI 2019.
Best Paper Honorable Mention
42. *The Channel Matters: Self-disclosure, Reciprocity and Social Support in Online Cancer Support Groups*
Diyi Yang, Zheng Yao, Joseph Seering, and Robert Kraut
The ACM CHI Conference on Human Factors in Computing Systems. CHI 2019.
Best Paper Honorable Mention

41. *Understanding Motivations behind Inaccurate Check-ins*
Fengli Xu, Guozhen Zhang, Jiaxin Huang, Yong Li, **Diyi Yang**, Ben Y. Zhao, Fanchao Meng
Proceedings of the ACM Conference on Computer-Supported Cooperative Work. CSCW 2018.
(Acceptance Rate: 27.3%)
40. *Understanding Self-narration of Personally Experienced Racism on Reddit*
Diyi Yang and Scott Counts
The 12nd International AAAI Conference on Weblogs and Social Media. ICWSM 2018.
39. *Persuading Teammates to Give: Systematic versus Heuristic Cues for Soliciting Loans*
Diyi Yang and Robert Kraut
Proceedings of the ACM Conference on Computer-Supported Cooperative Work. CSCW 2018.
(Acceptance Rate: 27.3%)
38. *Identifying Semantic Edit Intentions from Revisions in Wikipedia*
Diyi Yang, Aaron Halfaker, Robert Kraut and Eduard Hovy
The 2017 Conference on Empirical Methods on Natural Language Processing. EMNLP 2017 (Acceptance Rate: 24%)
37. *Self-disclosure and Channel Difference in Online Health Support Groups*
Diyi Yang, Zheng Yao, Robert Kraut
The 11th International AAAI Conference on Weblogs and Social Media. ICWSM 2017 (Acceptance Rate: 14%)
36. *Commitment of Newcomers and Old-timers to Online Health Support Communities*
Diyi Yang, Robert Kraut and John Levine
Proceedings of the 2017 Annual Conference on Human Factors in Computing Systems. CHI 2017.
(Acceptance Rate: 25%)
35. *Assessment of Discussion in Learning Contexts*
Carolyn Penstein Rose, Iris Howley, Miaomiao Wen, **Diyi Yang**, Oliver Ferschke
Innovative Assessment of Collaboration, 2017
34. *Who does What: Editor Role Identification in Wikipedia*
Diyi Yang, Aaron Halfaker, Robert Kraut and Eduard Hovy
The 10th International AAAI Conference on Weblogs and Social Media. ICWSM 2016 (Acceptance Rate: 17%)
Best Paper Honorable Mention
33. *Hierarchical Attention Networks for Document Classification*
Zichao Yang, **Diyi Yang**, Chris Dyer, Xiaodong He, Alex Smola and Eduard Hovy
The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics. NAACL 2016
32. *Edit Categories and Editor Role Identification in Wikipedia*
Diyi Yang, Aaron Halfaker, Robert Kraut and Eduard Hovy
The 10th International Conference on Language Resources and Evaluation. LREC 2016
31. *Humor Recognition and Humor Anchor Extraction*
Diyi Yang, Alon Lavie, Chris Dyer and Eduard Hovy
The 2015 Conference on Empirical Methods on Natural Language Processing. EMNLP 2015 (Acceptance Rate: 24%)
30. *That's So Annoying!!!: A Lexical and Frame-Semantic Embedding Based Data Augmentation Approach to Automatic Categorization of Annoying Behaviors using #petpeeve Tweets*

- William Yang Wang, **Diyi Yang**
The 2015 Conference on Empirical Methods on Natural Language Processing. EMNLP 2015 (Acceptance Rate: 24%)
Inaugural Notable Data Set Award, 2/1300, 0.2%
29. *Weakly Supervised Role Identification in Teamwork Interactions*
Diyi Yang, Miaomiao Wen, Carolyn Rosé
The 53rd Annual Meeting of the Association for Computational Linguistics. ACL 2015 (Acceptance Rate: 25%)
 28. *Exploring the Effect of Confusion in Discussion Forums of Massive Open Online Courses*
Diyi Yang, Miaomiao Wen, Iris Howley, Robert Kraut, Carolyn Rosé
The Second ACM Conference on Learning @ Scale. Learning at Scale 2015
 27. *Incorporating Word Correlation Knowledge into Topic Modeling*
Pengtao Xie, **Diyi Yang**, Eric Xing
The 14th Annual Conference of the North American Chapter of the Association for Computational Linguistics. NAACL 2015. (Acceptance Rate: 29%)
 26. *How does students cognitive behavior in MOOC discussion forums affect their learning gains?*
Xu Wang, **Diyi Yang**, Miaomiao Wen, Kenneth Koedinger, Carolyn Rosé
8th International Conference on Educational Data Mining. EDM 2015 (Acceptance Rate: 27.5%)
 25. *Virtual Teams in Massive Open Online Courses*
Miaomiao Wen, **Diyi Yang**, Carolyn Rosé
International Conference on Artificial Intelligence in Education. AIED 2015
 24. *CSCL and learning analytics: Opportunities to support social interaction, self-regulation and socially shared regulation*
Alyssa Friend Wise, Roger Azevedo, Karsten Stegmann, Jonna Malmberg, Carolyn Rosé, Nicholas Mudrick, Michelle Taub, Seth A Martin, Jesse Farnsworth, Jin Mu, Hanna Jarvenoja, Sanna Jarvela, Miaomiao Wen, **Diyi Yang**, Frank Fischer
International Society of the Learning Sciences.
 23. *Alleviating the negative effect of up and downvoting on help seeking in MOOC discussion forums*
Iris Howley, Gaurav Tomar, **Diyi Yang**, Oliver Ferschke, Carolyn Penstein Rosé
International Conference on Artificial Intelligence in Education. AIED 2015
 22. *Fostering discussion across communication media in massive open online courses*
Oliver Ferschke, Iris Howley, Gaurav Tomar, Diyi Yang, Yu Liu, Carolyn Penstein Rosé
International Society of the Learning Sciences, 2015
 21. *Positive Impact of Collaborative Chat Participation in an edX MOOC*
Oliver Ferschke, **Diyi Yang**, Gaurav Tomar, Carolyn Rosé
The 17th International Conference on Artificial Intelligence in Education. AIED, 2015
 20. *Challenges and opportunities of dual-layer MOOCs: Reflections from an edX deployment study*
Carolyn Penstein Rosé, Oliver Ferschke, Gaurav Tomar, Diyi Yang, Iris Howley, Vincent Alevan, George Siemens, Matthew Crosslin, Dragan Gasevic, Ryan Baker
Proceedings of the 11th International Conference on Computer Supported Collaborative Learning. CSCL 2015
 19. *Question Recommendation with Constraints for Massive Open Online Courses*
Diyi Yang, David Adamson, Carolyn Rosé
The 8th ACM Conference on Recommender Systems. RecSys 2014. (Acceptance Rate: 23%)

18. *Constrained Question Recommendation in MOOCs via Submodularity*
Diyi Yang, Jingbo Shang, Carolyn Rosé
The 2014 ACM International Conference on Information and Knowledge Management. CIKM 2014
(Acceptance Rate: 17%)
17. *Forum Thread Recommendation for Massive Open Online Courses*
Diyi Yang, Mario Piergallini, Iris Howley, Carolyn Rosé
7th International Conference on Educational Data Mining. EDM 2014 Short
16. *Peer Influence on Attrition in Massive Open Online Courses*
Diyi Yang, Miaomiao Wen, Carolyn Rosé
7th International Conference on Educational Data Mining. EDM 2014 Short
15. *Sentiment Analysis in MOOC Discussion Forums: What does it tell us?*
Miaomiao Wen, **Diyi Yang**, Carolyn Rosé
7th International Conference on Educational Data Mining. EDM 2014 (Acceptance Rate: 17%)
14. *Linguistic Reflections of Student Engagement in Massive Open Online Courses*
Miaomiao Wen, **Diyi Yang**, Carolyn Rosé
The 8th International AAAI Conference on Weblogs and Social Media. ICWSM 2014 (Acceptance Rate: 23%)
13. *Social Factors that Contribute to Attrition in MOOCs*
Carolyn Rosé, Ryan Carlson, **Diyi Yang**, Miaomiao Wen, Lauren Resnick, Pam Goldman, Jennifer Sheerer
Learning at Scale, 2014
12. *Predicting Advertiser Bidding Behaviors in Sponsored Search by Rationality Modeling*
Haifeng Xu, Bin Gao, **Diyi Yang**, Tie-Yan Liu
The 22nd International Conference on World Wide Web. WWW 2013 (Acceptance Rate: 16%)
11. *Collaborative Filtering with Short Term Preferences Mining*
Diyi Yang, Tianqi Chen, Weinan Zhang, Yong Yu
Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR 2012
10. *Local Implicit Feedback Mining for Music Recommendation*
Diyi Yang, Tianqi Chen, Weinan Zhang, Qiuxia Lu, Yong Yu
The 6th ACM Conference on Recommender Systems. RecSys 2012 (Acceptance Rate: 20%)
9. *Serendipitous Personalized Ranking for Top-N Recommendation*
Qiuxia Lu, Tianqi Chen, Weinan Zhang, **Diyi Yang**, Yong Yu
2012 IEEE/WIC/ACM International Conference on Web Intelligence. WI 2012
Best Paper Honorable Mention

Journal Publications

8. *Exploring the Effect of Student Confusion in Massive Open Online Courses*
Diyi Yang, Robert Kraut and Carolyn Rosé
Journal of Educational Data Mining (JEDM)
7. *Towards an Integration of Text and Graph Clustering Methods as Lens for Studying Social Interaction in MOOCs*
Diyi Yang, Miaomiao Wen, Abhimanu Kumar, Eric Xing, Carolyn Rosé
International Review of Research in Open and Distance Learning, Special Issue on the MOOC Research Initiative

Workshop Publications

6. *Semi-supervised Models via Data Augmentation for Classifying Interactive Affective Responses*
Jiaao Chen, Yuwei Wu, and **Diyi Yang**
AAAI 2020 Workshop on Affective Content Analysis
5. *A Lightly Supervised Approach to Role Identification in Wikipedia Talk Page Discussions*
Oliver Ferschke, **Diyi Yang**, Carolyn P. Rosé
ICWSM 2015 Workshop on Wikipedia
4. *Turn on, Tune in, Drop out: Anticipating student dropouts in Massive Open Online Courses*
Diyi Yang, Tanmay Sinha, David Adamson, Carolyn Rosé
Data Driven Education Workshop of the 27th Conference on Neural Information Processing Systems
3. *Towards Identifying the Resolvability of Threads in MOOCs*
Diyi Yang, Miaomiao Wen, Carolyn Rosé
MOOC Workshop of the 2014 Conference on Empirical Methods on Natural Language Processing (EMNLP)
2. *Combining Factorization Model and Additive Forest for Collaborative Followee Recommendation*
Tianqi Chen, Linpeng Tang, Qin Liu, **Diyi Yang**, Saining Xie, Xuezhi Cao, Chunyang Wu, Enpeng Yao, Zhengyang Liu, Zhansheng Jiang, Cheng Chen, Weihao Kong, Yong Yu
KDD Cup 2012, Champion
1. *Informative household recommendation with feature-based matrix factorization*
Qiuxia Lu, Diyi Yang, Tianqi Chen, Weinan Zhang, Yong Yu
Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation, RecSys, 2012

Grants and Sponsorship

- Data Curation Award, Institute for Data Engineering and Science: \$14,000 2020
- The Academic Data Science Alliance Seed Grant: \$49, 929 2019
- NVIDIA Accelerated Data Science GPU Grant: \$12,000 2019
- Google Gift Award: \$51,000 2019
- Google Gift Award: \$55,000 2019
- Google Cloud Platform Education Grant: \$20,000 2019

Advising

PhD Students

- Ian Stewart (PhD student, co-advised with Jacob Eisenstein)
- Jiaao Chen (PhD student)

Thesis Committees

- Sindhu Kiranmai Ernala (PhD Thesis Committee)
- Ian Stewart (PhD Thesis Committee)

Qualification Committees

- Prithviraj Ammanobrolu

Teaching

- Instructor, CS 4650/CS 7650: Natural Language Processing.
Georgia Institute of Technology Spring 2020
- Instructor, CS 8001: NLP and Social Interaction.
Georgia Institute of Technology Spring 2020
- Instructor, CS 8001: NLP and Social Interaction.
Georgia Institute of Technology Fall 2019
- Guest Lecturer, CSE/ISYE 6740 Computational Data Analysis.
Georgia Institute of Technology Fall 2019
- Guest Lecturer, CS 6474/CS 4803 Social Computing.
Georgia Institute of Technology Fall 2019
- Guest Lecturer, Natural Language Processing.
Carnegie Mellon University Fall 2018, Spring 2018
- Guest Lecturer, Natural Language Processing.
Carnegie Mellon University Spring 2018
- Guest Lecturer, Computational Ethnics in NLP.
Carnegie Mellon University Spring 2018
- Research Mentor, Opportunities for Undergraduate Research in Computer Science.
Carnegie Mellon University Fall 2017
- Teaching Assistant, Natural Language Processing.
Carnegie Mellon University Fall 2016
- Teaching Assistant, LTI Colloquium.
Carnegie Mellon University Spring 2016, Fall 2016
- Teaching Assistant, Computer Architecture.
Shanghai Jiao Tong University Spring 2012

Professional Activities

- Area Chair, the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020
- Area Chair, the 58th Annual Meeting of the Association for Computational Linguistics (ACL), 2020
- Area Chair, the ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW), 2020
- Area Chair, the 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019
- Area Chair, the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2019

- Associate Chair, CHI 2019 Late Breaking Work, 2019
- Journal Reviewer, ACM Transactions on Computer-Human Interaction (TOCHI)
- Co-Chair of WiNLP¹, colocated with ACL 2019
- Program committee member, the Third Workshop on Abusive Language Online, 2019.
- Program committee member, the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015-2019
- Program committee member, the SIGNLL Conference on Computational Natural Language Learning, 2019
- Best Paper Committee, The 4th Workshop on Noisy User-generated Text, EMNLP, 2018
- Co-Chair of WiNLP, colocated with NAACL 2018
- Program committee member, the AAAI Conference on Artificial Intelligence (AAAI) 2018
- Program committee member, the Annual Meeting of the Association for Computational Linguistics (ACL), 2015-2017
- Program committee member, the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015-2018
- Program committee member, the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT), 2015 - 2018
- Program committee, the International Conference on Computational Linguistics (COLING) 2016
- Poster session chair, the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018
- Program committee, the International Conference on Language Resources and Evaluation, 2016, 2018
- Program committee, the Workshop on Noisy User-generated Text (W-NUT), EMNLP. 2016-2018
- Reviewer, the ACM CHI Conference on Human Factors in Computing Systems (CHI), 2017, 2018, 2019, 2020
- Reviewer, the ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW) 2017,2018
- Reviewer, the International AAAI Conference on Weblogs and Social Media, 2018, 2019
- Reviewer, the 4th Annual International Conference on Computational Social Science, 2018
- Journal reviewer, Natural Language Engineering, 2018
- Journal reviewer, Transactions on Information Systems, 2018
- Journal reviewer, Computers Education, 2016

¹Women and Underrepresented Minorities in Natural Language Processing

Recent Presentations & Invited Talks

- “Building Language Technologies for Social Good”. Invited Talk at Salesforce, Palo Alto, Jan 31th, 2020
- “Building Language Technologies for Social Good”. Invited Speaker at RE·Work, San Francisco, Jan 30th, 2020
- “Building Language Technologies for Social Good”. GVU Brown Bag Seminar at Georgia Tech, Jan 9th, 2020
- “Building Language Technologies for Better Online Communities”. Invited talk at University of Massachusetts Amherst, Nov 21st, 2019
- “Building Language Technologies for Better Online Communities”. Invited talk at University of Texas at Austin, Nov 10th, 2019
- “Building Language Technologies for Better Online Communities”. Invited talk at the Workshop on Language, Technology, and Society at Georgia Tech, Oct 23rd, 2019
- “Building Language Technologies for Better Online Communities”. Invited talk at NYU Center of Data Science, Sept 26th, 2019
- “Building Language Technologies for Better Online Communities”. Invited talk at Bloomberg, July 18th, 2019
- “Modeling Persuasive Strategies via Semi-Supervised Neural Nets on Crowdfunding Platforms”. NAACL 2019, Minneapolis, USA, June 5th, 2019.
- “The Channel Matters: Self-disclosure, Reciprocity and Social Support in Online Cancer Support Groups”. CHI 2019, Glasgow, UK, May, 2019.
- “Modeling Social Roles in Online Health Communities”. CHI 2019, Glasgow, UK, May, 2019.
- “Computational Social Interaction Analysis for Better Online Communities”. Invited talk at Xian Jiao Tong University, Xi’an, China, Feb, 2019
- “Towards Understanding Human Communication via NLP”. Invited talk at National Taiwan University, Taipei, Taiwan, Feb, 2019
- “Towards Understanding Human Communication via NLP”. Invited talk at Google Research, NYC, USA, Dec, 2018
- “Persuading Teammates to Give: Systematic versus Heuristic Cues for Soliciting Loans”. CSCW, Jersey City, NJ, USA, Nov, 2018
- “Understanding Motivations behind Inaccurate Check-ins”. CSCW, Jersey City, NJ, USA, Nov, 2018
- “Computational Social Interaction Analysis for Better Online Communities”. Invited talk at Bocconi University, Milan, Italy, Oct, 2018
- “Modeling Members’ Social Roles and their Conversational Acts in Online Communities”. Invited talk at The 4th Workshop on Noisy User-generated Text (W-NUT) at EMNLP, Brussels, Belgium, Nov 1st, 2018
- “Computational Social Interaction Analysis for Better Online Communities”. Invited talk at Facebook’s Women in Research, March, 2018

- “Modeling Social Roles via Language Analysis in Online Cancer Support Groups”. Invited talk at Computational Linguistics and Lunch, University of Pennsylvania, Oct 3rd, 2018
- “Computational Social Interaction Analysis for Better Online Communities”. Stanford University, June 2018
- “Computational Social Interaction Analysis for Better Online Communities”. Shanghai Jiao Tong University, Shanghai, China, May 2018
- “Computational Social Interaction Analysis for Better Online Communities”. Carnegie Mellon University, Feb 2018
- “Computational Social Interaction Analysis for Better Online Communities”. Georgia Institute of Technology, Jan 2018
- “Identify, Recommend, and Configure in Social Roles in Online Communities ”. Rising Stars in EECS Workshop, Oct, 2017

Last updated: February 27, 2020