

# Diyi Yang

---

## Contact Information

Language Technologies Institute  
Carnegie Mellon University  
Pittsburgh, PA, 15213

*E-mail:* [diyiy@cs.cmu.edu](mailto:diyiy@cs.cmu.edu)  
*Phone:* +1 (412) 980-8543  
*Page:* [cs.cmu.edu/~diyiy](http://cs.cmu.edu/~diyiy)

## Research Interests

Social Computing, Natural Language Processing

## Education

### Carnegie Mellon University

Language Technologies Institute, School of Computer Science

- Ph.D., advised by [Robert Kraut](#) and [Eduard Hovy](#) 2013 – Present
- Additional thesis committee members: [Brandy L. Aven](#) (Tepper School of Business, Carnegie Mellon University), [Dan Jurafsky](#) (Department of Linguistics, Stanford University).

### Carnegie Mellon University

Language Technologies Institute, School of Computer Science

- M.Sc., advised by [Carolyn P. Rosé](#) 2013 – 2015

### Shanghai Jiao Tong University

Computer Science, School of Electronic Information and Electrical Engineering

- B.Sc., [ACM Honored Class](#), advised by [Yong Yu](#) 2009 – 2013

## Awards & Honors

[Selected Rising Stars in EECS, Stanford](#) 2017  
[Facebook PhD Fellowship](#) 2017-2019  
[Presidential Fellowship, CMU](#) 2016  
[ICWSM 2016 Best Paper Honorable Mention](#) 2016  
[EMNLP 2015 Notable Data Set Award](#) 2015  
[Excellent Undergraduate Thesis Award, SJTU](#) 2013  
[Google Anita Borg Scholarship](#) 2012  
[Champion in KDD Cup 2012, Track 1 \(Team ACMClass@SJTU\)](#) 2012  
[Best Paper Award Candidate, International Conference on Web intelligence](#) 2012

## Research Internships

### 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](#)

- [Wikipedia 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

## Conference Publications

[Google Scholar Profile](#) (Citation: >1000. H-index: 18. I10-index:20)

30. *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%)

29. *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%)
28. *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%)
27. *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%)
26. *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**
25. *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
24. *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
23. *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%)
22. *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%)  
**Notable Data Set Award: 2/1300, 0.2%**
21. *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%)
20. *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
19. *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%)
18. *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%)

17. *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
16. *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%)
15. *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%)
14. *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
13. *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
12. *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%)
11. *Linguistic Reflections of Student Engagement in Massive Open Online Courses*  
Miaomiao Wen, **Diyi Yang**, Carolyn Rosé  
The 8th International AAI Conference on Weblogs and Social Media. ICWSM 2014 (Acceptance Rate: 23%)
10. *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%)
9. *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
8. *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%)
7. *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 Award Candidate**
6. *Exploring the Effect of Student Confusion in Massive Open Online Courses*  
**Diyi Yang**, Robert Kraut and Carolyn Rosé  
Journal of Educational Data Mining (JEDM)
5. *Towards an Integration of Text and Graph Clustering Methods as a 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

**Journal  
Publications**

## Workshop Publications

4. *A Lightly Supervised Approach to Role Identification in Wikipedia Talk Page Discussions*  
Oliver Ferschke, **Diyi Yang**, Carolyn P. Rosé  
ICWSM 2015 Workshop on Wikipedia
3. *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 (NIPS)
2. *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)
1. *Combining Factorization Model and Additive Forest for Collaborative Follower 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

## Teaching

- 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 Service

- Co-Chair of WiNLP<sup>1</sup>, colocated with NAACL 2018 2018
- Reviewer for CHI, NAACL, LREC, ICWSM 2018
- Poster Session Chair for EMNLP 2017
- Reviewer for ACL, EMNLP 2017
- Reviewer for ACL, EMNLP, NAACL, LREC 2016

---

<sup>1</sup>Women and Underrepresented Minorities in Natural Language Processing