

# Liu Yang

Center for Intelligent Information Retrieval  
College of Information and Computer Sciences  
University of Massachusetts Amherst  
140 Governors Drive, Room 364, Amherst, MA 01003

Email: [lyang@cs.umass.edu](mailto:lyang@cs.umass.edu)  
Mobile: +1 (413) 404-9495  
Homepage: <https://people.cs.umass.edu/lyang/>  
Last Updated: 05/24/2018

---

## Research Interests

Information retrieval, natural language processing, data mining and machine learning, with an emphasis on text mining, deep learning, ranking and relevance, question answering and probabilistic graphical models

---

## Education

### University of Massachusetts Amherst

#### Ph.D. in Computer Science

Advisor: Prof. W. Bruce Croft, ACM Fellow

AMHERST, MA, USA

Aug. 2014 - May 2019 (expected)

### University of Massachusetts Amherst

#### M.S. in Computer Science (GPA: 3.919/4)

Advisor: Prof. W. Bruce Croft, ACM Fellow

AMHERST, MA, USA

Aug. 2014 - Feb. 2017

### Singapore Management University

#### Visiting Graduate Student in Computer Science

Advisor: Prof. Jing Jiang and Prof. Feida Zhu

SINGAPORE, SINGAPORE

Sep. 2012 - Jun. 2014

### Peking University

#### M.E. in Software Engineering

Advisor: Prof. Zhong Chen

BEIJING, P.R.CHINA

Sep. 2011 - Jul. 2014

### Northeastern University

#### B.E. in Software Engineering (Ranking: 1st of 337)

SHENYANG, P.R.CHINA

Sep. 2007 - Jul. 2011

---

## Internship Experience

### Microsoft Research Redmond

REDMOND, SEATTLE AREA, USA

#### Research Intern

Jun. 2016 - Sept. 2016

Mentors: Dr. Susan Dumais (ACM Fellow), Dr. Paul Bennett, Dr. Ahmed Hassan Awadallah from MSR  
**Characterize and predict user email reply behavior, Data Mining and Machine Learning; User Behavior Analysis and User Modeling; Classification; Feature Extraction and Analysis**

I worked on a project on characterizing and predicting user email reply behavior. The main task is to characterize factors affecting email replies and build models to predict user reply behavior including reply action/ reply time in enterprise emails. A research paper based on this project is published as a full oral paper in SIGIR'17.

### Microsoft Research Redmond/Bing

BELLEVUE, SEATTLE AREA, USA

#### Research Intern

May. 2015 - Aug. 2015

Mentors: Dr. Kieran McDonald, Dr. Qi Guo, Dr. Sha Meng from Bing and Dr. Yang Song, Dr. Milad Shokouhi from MSR

**Learning to rank for proactive ranking, Cosmos/SCOPE for machine learning with Big Data; Data Mining and Search Ranking; Mining user interests/user modeling from search/browse logs**

I worked on modeling user interests for proactive ranking of information cards of Microsoft Cortana. The main task is feature design and extraction from user behavior in search/browse logs to improve proactive ranking performance. Experiments performed with product data from a large commercial proactive search system show that the proposed method significantly outperforms the method deployed in the product. A research paper based on this project is published as a full oral paper in ECIR'16.

### Baidu Inc.

BEIJING, P.R.CHINA

#### Software Engineer Intern

Feb. 2011 - Aug. 2011

Mentors: R&D Manager Jianguo Liu and Senior Software Engineer Tianlong Zhang

**APP Inner Search Module of Baidu Search Engine, Linux, PHP, MySQL, JavaScript**

I developed the application inner search module of Baidu search engine. Main features include directly showing videos, music, web applications in the search results of Baidu. I also participated in the development of MIS system for the Baidu Open Platform. The URL for this service is [app.baidu.com](http://app.baidu.com).

---

## Research Experience

Center for Intelligent Information Retrieval, UMass Amherst

AMHERST, USA

### Research Assistant

Aug. 2014 - Present

Advisor: Prof. W. Bruce Croft

I am working with Prof. W. Bruce Croft on information retrieval, question answering and machine learning. I am also working closely with Prof. Jiafeng Guo on deep neural network models for question answering. We proposed an attention based neural matching model (aNMM) for ranking short answer text. aNMM adopts value-shared weighting scheme instead of position-shared weighting scheme for combining different matching signals and incorporates question term importance learning using question attention network. The paper on aNMM model is published in **CIKM'16** (Full Oral Paper). The implementation of aNMM using TensorFlow/Python is open sourced in [MatchZoo Toolkit](#).

Institute of Computing Technology, Chinese Academy of Science

BEIJING, CHINA

### Visiting PhD Student

Sep. 2017 - Nov. 2017

Advisor: Prof. Jiafeng Guo

I am working with Prof. Jiafeng Guo on deep learning, question answering and neural conversational models. The paper on response ranking with deep matching networks and external knowledge in information-seeking conversations is published in **SIGIR'18** (Full Oral Paper). I am also developing the [MatchZoo Toolkit](#) (about 1K Stars and 300 Forks on Github) with joint efforts of Prof. Guo's students. The [MatchZoo Toolkit](#) was developed with a focus on facilitating the designing, comparing and sharing of deep text matching models. There are a number of deep text matching models implemented, such as DRMM, MatchPyramid, MV-LSTM, aNMM, DUET, ARC-I, ARC-II, DSSM, CDSSM, K-NRM and CONV-KNRM. Potential tasks related to MatchZoo include document retrieval, question answering, conversational response ranking, paraphrase identification, textual entailment identification, etc.

Research Center, Singapore Management University

SINGAPORE, SINGAPORE

### Research Assistant

Feb. 2013 - Aug. 2013

Advisor: Prof. Jing Jiang and Prof. Feida Zhu

**Jointly Model Topics and Expertise in Community Question Answering, Probabilistic Graphical Model, Gaussian Mixture Model, Link Analysis, Research Project.**

We proposed a probabilistic generative model TEM with GMM hybrid to jointly model topical interests and topical expertise of CQA users. Based on TEM results, we proposed CQARank algorithm to combine textual content learning results and link structure analysis to enforce topical expertise learning. We published a paper in **CIKM'13** (Full Oral Paper, [Top 3 Cited Papers in CIKM 2013](#)).

---

## Publications

- 1) Minghui Qiu, **Liu Yang**, Feng Ji, Wei Zhou, Weipeng Zhao, Jun Huang, Haiqing Chen, W. Bruce Croft, Wei Lin. Transfer Learning for Context-Aware Question Matching in Information-seeking Conversation Systems in E-commerce. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (**ACL 2018**), Melbourne, Australia, July 15-20, 2018. **Short Paper**. Acceptance rate=24% (126 out of 526)
- 2) **Liu Yang**, Minghui Qiu, Chen Qu, Jiafeng Guo, Yongfeng Zhang, W. Bruce Croft, Jun Huang, Haiqing Chen. Response Ranking with Deep Matching Networks and External Knowledge in Information-seeking Conversation Systems, In Proceedings of the 41th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR 2018**), Ann Arbor, Michigan, U.S.A. July 8-12, 2018. **Full Oral Paper**. Acceptance rate=21% (86 out of 409)
- 3) Chen Qu, **Liu Yang**, W. Bruce Croft, Johanne R Trippas, Yongfeng Zhang, Minghui Qiu. Analyzing and Characterizing User Intent in Information-seeking Conversations, In Proceedings of the 41th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR 2018**), Ann Arbor, Michigan, U.S.A. July 8-12, 2018. **Short Paper**. Acceptance rate=30% (98 out of 327).
- 4) Daniel Cohen, **Liu Yang**, W. Bruce Croft. WikiPassageQA: A Benchmark Collection for Research on Non-factoid Answer Passage Retrieval, In Proceedings of the 41th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR 2018**), Ann Arbor, Michigan, U.S.A. July 8-12, 2018. **Short Paper**. Acceptance rate=30% (98 out of 327).
- 5) **Liu Yang**, Susan T. Dumais, Paul N. Bennett and Ahmed Hassan Awadallah, Characterizing and Predicting Enterprise Email Reply Behavior, In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR 2017**), Tokyo, Japan, August 7-11, 2017. **Full Oral Paper**. Acceptance rate=22% (78 out of 362).

- 6) **Liu Yang**, Hamed Zamani, Yongfeng Zhang, Jiafeng Guo, W. Bruce Croft. Neural Matching Models for Question Retrieval and Next Question Prediction in Conversation, In Neu-IR 2017: The SIGIR 2017 Workshop on Neural Information Retrieval (**SIGIR Neu-IR 2017**), Tokyo, Japan, August 7-11, 2017. **Oral Presentation**.
- 7) **Liu Yang**, Qingyao Ai, Jiafeng Guo, W. Bruce Croft, aNMM: Ranking Short Answer Texts with Attention-Based Neural Matching Model, In Proceedings of the 25th ACM International Conference on Information and Knowledge Management (**CIKM 2016**), Indianapolis, IN, USA. October 24-28, 2016. **Full Oral Paper**. Acceptance rate=17.6% (165 out of 935).
- 8) Qingyao Ai, **Liu Yang**, Jiafeng Guo, W. Bruce Croft, Analysis of the Paragraph Vector Model for Information Retrieval, In Proceedings of The 2nd ACM International Conference on the Theory of Information Retrieval (**ICTIR 2016**). Newark, DE, USA. September 12-16, 2016. **Full Oral Paper**.
- 9) Qingyao Ai, **Liu Yang**, Jiafeng Guo, W. Bruce Croft. Improving Language Estimation with the Paragraph Vector Model for Ad-hoc Retrieval. In Proceedings of the 39th Annual ACM SIGIR Conference on Research and Development in Information Retrieval (**SIGIR 2016**), Pisa, Italy. July 18-10, 2016. **Short Paper**.
- 10) **Liu Yang**, Qingyao Ai, Damiano Spina, Ruey-Cheng Chen, Liang Pang, W. Bruce Croft, Jiafeng Guo and Falk Scholer. Beyond Factoid QA: Effective Methods for Non-factoid Answer Sentence Retrieval. In Proceedings of the 38th European Conference on Information Retrieval (**ECIR 2016**), Padova, Italy, March 20-23, 2016. **Full Paper**. Acceptance rate = 21%.
- 11) **Liu Yang**, Qi Guo, Yang Song, Sha Meng, Milad Shokouhi, Kieran McDonald and W. Bruce Croft. Modelling User Interest for Zero-query Ranking. In Proceedings of the 38th European Conference on Information Retrieval (**ECIR 2016**), Padova, Italy, March 20-23, 2016. **Full Paper**. Acceptance rate = 21%.
- 12) **Liu Yang**, Jing Jiang, Lifu Huang, Minghui Qiu and Lizi Liao. Generating Supplementary Travel Guides from Social Media. In Proceedings of the 25th International Conference on Computational Linguistics (**COLING 2014**), Dublin, Ireland, August 2014. **Full Oral Paper**.
- 13) Jianguang Du, Jing Jiang, **Liu Yang**, Dandan Song, Lejian Liao. ShellMiner: Mining Organizational Phrases in Argumentative Texts in Social Media. In Proceedings of the 14th IEEE International Conference on Data Mining (**ICDM 2014**), Shenzhen, China, December 2014. **Short Paper**.
- 14) Swapna Gottipati, Minghui Qiu, **Liu Yang**, Feida Zhu, Jing Jiang. An Integrated Model for User Attribute Discovery: A Case Study on Political Affiliation Identification. In Proceedings of the 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining (**PAKDD 2014**), Tainan, Taiwan. May 2014. **Full Oral Paper**. acceptance rate=10.8% (40 out of 371).
- 15) **Liu Yang**, Minghui Qiu, Swapna Gottipati, Feida Zhu, Jing Jiang, Huiping Sun and Zhong Chen. CQARank: Jointly Model Topics and Expertise in Community Question Answering. In Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (**CIKM 2013**). San Francisco, California, USA. October 2013. **Full Oral Paper**, **Top 3 Cited Papers in CIKM 2013**, acceptance rate=16.8% (143 out of 848).
- 16) Minghui Qiu, **Liu Yang** and Jing Jiang. Modeling Interaction Features for Debate Side Clustering. In Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (**CIKM 2013**). San Francisco, California, USA. October 2013. **Short Paper**, acceptance rate=12.5% (106 out of 848).
- 17) Minghui Qiu, **Liu Yang** and Jing Jiang. Mining User Relations from Online Discussions using Sentiment Analysis and Probabilistic Matrix Factorization. In Proceedings of the 2013 Conference of North American Chapter of Association for Computational Linguistics (**NAACL 2013**). Atlanta, Georgia, USA. June 2013. **Long Paper**, acceptance rate=30% (88 out of 293).
- 18) Swapna Gottipati, Minghui Qiu, **Liu Yang**, Feida Zhu, Jing Jiang. Predicting User's Political Party using Ideological Stances. In Proceedings of the 5th International Conference on Social Informatics (**SocInfo 2013**), Kyoto, Japan. November 2013. **Full Paper**, **Best Paper Runner-ups**.

---

## Honors and Awards

**SIGIR Student Travel Award**, CIKM 2013, ECIR 2016, CIKM 2016, SIGIR 2017, SIGIR 2018 2013-2018  
 Awarded to selected student authors  
 ACM SIGIR, ACM SIGWEB, Facebook, etc.

**Best Paper Award Runner-ups** 2013  
 Awarded to the best five papers  
 SocInfo 2013

<b>National Scholarship</b> Top 1% in over 12,000 students Ministry of Education, P.R.China	2010
<b>IBM Excellent Student Scholarship</b> Awarded to only 75 undergraduates and postgraduates all over China IBM	2010
<b>International Mathematical Contest in Modelling(MCM)</b> Honorable Mention COMAP, SIAM and MAA, USA	2010
<b>Chinese Undergraduate Mathematical Contest in Modelling(CUMCM)</b> National Second Prize CSIAM and Department of Higher Education, MoE, P.R.China	2009
<b>ACM-ICPC Programming Contest</b> Second Prize Northeastern University	2011
<b>Top 10 Undergraduates of Liaoning Province</b> Top 10 in over 100,000 students Department of Education of Liaoning Province	2011
<b>Outstanding Bachelor Graduate</b> Top 2% in over 3,000 graduates Department of Education of Liaoning Province	2011

---

## Selected Open Source Projects

- [MatchZoo](#): MatchZoo is a toolkit for deep neural text matching, with model implementations include ARC-I/ ARC-II, DSSM, CDSSM, MatchPyramid, DRMM, aNMM, MV-LSTM, DUET, etc using TensorFlow.
  - [LDAGibbsSampling](#): LDAGibbsSampling is an open source package for Gibbs sampling inference of LDA model, which could be used for topic modeling in text mining.
- 

## Professional Activities

- PC Member, 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)
  - PC Member, The 27th ACM International Conference on Information and Knowledge Management (CIKM 2018)
  - PC Member, The 24rd SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2018)
  - PC Member, The 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018)
  - PC Member, Full Paper Track, The 4th ACM International Conference on the Theory of Information Retrieval (ICTIR 2018)
  - Reviewer, The 40th European Conference on Information Retrieval (ECIR 2018), 2018.
  - PC Member, Demo Track, The 11th ACM International Conference on Web Search and Data Mining (WSDM 2018), 2018
  - Reviewer, The 26th international World Wide Web Conference (WWW 2017), 2017.
  - PC Member, The 26th ACM International Conference on Information and Knowledge Management (CIKM 2017), 2017.
  - PC Member, Machine Learning Track, The 3rd ACM International Conference on the Theory of Information Retrieval (ICTIR 2017), 2017.
  - PC Member, The 24th ACM International Conference on Information and Knowledge Management (CIKM 2015), 2015.
  - ACM Student Member, [ACM](#), Jun. 2013 - Present.
-

## Technical Skills

- **Programming:**
    - Proficient: Java, Python, C, SQL
    - Skillful: C++, C#, Matlab, SCOPE, PHP, JSP, Shell Scripting,  $\LaTeX$
    - Experienced: ASP, JavaScript, Qt, Assembly Language
  - **Theory:** Foundation of algorithms, math, NLP, IR, machine learning and data mining.
  - **System:** Linux/Unix, Windows, Mac OS X, MySQL/SQL Server, COSMOS
  - **Tools:** Weka, LibSVM/ SVMlight, Mallet, Numpy&SciPy, Pandas, Scikit-learn, Theano, TensorFlow, Keras, PyTorch, MXNet, NLP pipelines (Stanford NLP, OpenNLP), NLTK, LP\_solver, CPLEX & many optimization toolkits.
- 

## Selected Courses

- Information Retrieval, Artificial Intelligence, Machine Learning, Data Mining, Deep Learning, Algorithm Analysis and Design, Advanced Algorithms, Optimization in Computer Science, Introduction to Numerical Computing with Python, Database Design and Implementation, Database System, Data Structure, Operating System, Distributed System, Computer Architecture, Computer Networks, Advanced Software Engineering, Compiler Method, Advanced Mathematics(Calculus), Linear Algebra, Discrete Mathematics, Probability Theory and Statistics.
- 

## Nature Languages

- **English** (professional), **Chinese** (mother tongue), **Japanese** (elementary).
- 

## References

Available Upon Request