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
Mobile: +1 (413) 404-9495
Homepage: https://people.cs.umass.edu/lyang/
Last Updated: 05/15/2017

Research Interests

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

Education

University of Massachusetts Amherst — Amherst, MA, USA
Ph.D. in Computer Science — *Aug. 2014 - Present*
Advisor: Prof. W. Bruce Croft, ACM Fellow

University of Massachusetts Amherst — Amherst, MA, USA
M.S. in Computer Science — *Aug. 2014 - Feb. 2017*
Advisor: Prof. W. Bruce Croft, ACM Fellow

Singapore Management University — Singapore, SINGAPORE
Visiting Graduate Student in Computer Science — *Sep. 2012 - Jun. 2014*
Advisor: Prof. Jing Jiang and Prof. Feida Zhu

Peking University — Beijing, P.R.China
M.E. in Software Engineering — *Sep. 2011 - Jul. 2014*
Advisor: Prof. Zhong Chen

Northeastern University — Shenyang, P.R.China
B.E. in Software Engineering (Summa Cum Laude) — *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 predit 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/browze 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*
Mentor: 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.

Research Experience

Center for Intelligent Information Retrieval, UMass Amherst — Amherst, USA

Research Assistant — *Aug. 2014 - Present*

Working with Prof. W. Bruce Croft on information retrieval, question answering and machine learning.

Research Center, Singapore Management University — Singapore, SINGAPORE

Research Assistant — *Nov. 2013 - Jun. 2014*

Advisor:Prof. Jing Jiang

Generating Supplementary Travel Guides from Social Media, *Probabilistic Graphical Model, Integer Linear Programming, Summarization*, Research Project.

We studied how to summarize travel-related information in forum threads to generate supplementary travel guides and proposed a summarization framework based on a latent variable model and Integer Linear Programming. We published a paper in **COLING'14** (Full Oral Paper).

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 topics and expertise of CQA users. Based on TEM results, we proposed CQARank algorithm to combine textual content learning results and link structure analysis to enforce topic expertise learning. We published a paper in **CIKM'13** (Full Oral Paper).

Research Assistant — *Dec. 2012 - Feb. 2013*

Advisor: Prof. Jing Jiang

Modeling Interaction Features for Finding Viewpoint, *Probabilistic Graphical Model, Topic Model, JAVA, Matlab*, Research Project.

We proposed a two-stage solution based on latent variable models to mine interaction features from online debate forums. Empirical evaluation shows that the learned interaction features provide good insights into user interactions. We published a paper in **CIKM'13** (Short Paper).

Research Assistant — *Sep. 2012 - Dec. 2012*

Advisor: Prof. Jing Jiang

Mining User Relations with Sentiment Analysis and Probabilistic Matrix Factorization, *Opinion Mining, Probabilistic Matrix Factorization, JAVA, Matlab*, Research Project.

We identified aspects and opinions in online discussion and extracted relations between them to construct a User-Aspect and User-User rate matrix. Then we mined user relations based on Probabilistic Matrix Factorization to tackle the data sparsity. We published a paper in **NAACL'13** (Long Paper).

Publications

1) **Liu Yang**, Susan T. Dumais, Paul N. Bennett and Ahmed Hassan Awadallah, Characterizing and Predicting Enterprise Email Reply Behavior, To appear 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).
2) **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).
3) 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 Retrieva (**ICTIR 2016**). Newark, DE, USA. September 12-16, 2016. **Full Oral Paper**.
4) 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**.
5) **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%.

6) **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%.

7) **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**.

8) 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**.

9) 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).

10) **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).

11) 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).

12) 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).

13) 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 *2013-2017*
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

National Scholarship *2010*
Top 1% in over 12,000 students
Ministry of Education, P.R.China

IBM Excellent Student Scholarship *2010*
Awarded to only 75 undergraduates and postgraduates all over China
IBM

International Mathematical Contest in Modelling(MCM) *2010*
Honorable Mention
COMAP, SIAM and MAA, USA

Chinese Undergraduate Mathematical Contest in Modelling(CUMCM) *2009*
National Second Prize
CSIAM and Department of Higher Education, MoE, P.R.China

ACM-ICPC Programming Contest *2011*
Second Prize
Northeastern University

Top 10 Undergraduates of Liaoning Province *2011*
Top 10 in over 100,000 students
Department of Education of Liaoning Province

Outstanding Bachelor Graduate *2011*
Top 2% in over 3,000 graduates
Department of Education of Liaoning Province

Academic Excellent Award *2010*
Top 10 in over 12,000 students
Northeastern University

Suzhou Industrial Park Scholarship *2009*
Top 2% students
Suzhou Industrial Park

Outstanding Student Scholarship *2008-2011*
First Grade, Top 3% students
Northeastern University

Recent Presentations and Invited Talks

1) **CQARank: Jointly Model Topics and Expertise in Community Question Answering**
at the 22nd ACM International Conference on Information and Knowledge Management (CIKM 2013), Burlingame, CA, USA, Oct. 29th, 2013.
2) **Modelling User Interest for Zero-query Ranking**
at the 38th European Conference on Information Retrieval (ECIR 2016), Padova, Italy, March 20-23, 2016.
3) **Beyond Factoid QA: Effective Methods for Non-factoid Answer Sentence Retrieval**
at the 38th European Conference on Information Retrieval (ECIR 2016), Padova, Italy, March 20-23, 2016.
4) **aNMM: Ranking Short Answer Texts with Attention-Based Neural Matching Model**
at the 25th ACM International Conference on Information and Knowledge Management (CIKM 2016), Indianapolis, IN, USA. October 24-28, 2016.
5) **aNMM: Ranking Short Answer Texts with Attention-Based Neural Matching Model**
at the Data Science Tea of CICS, UMass Amherst, Amherst, MA, USA. November 28, 2016.
6) **aNMM: Ranking Short Answer Texts with Attention-Based Neural Matching Model**
at the guest lecture in CS646 Information Retrieval course, UMass Amherst, Amherst, MA, USA. December 5, 2016.

Professional Activities

- Reviewer, The 26th international World Wide Web Conference (WWW 2017), 2017.
- Program Committee Member and Reviewer, The 26th ACM International Conference on Information and Knowledge Management (CIKM 2017), 2017.
- Program Committee Member, The Sixth Conference on Natural Language Processing and Chinese Computing (NLPCC 2017), 2017.
- Program Committee Member, Machine Learning Track, The 3rd ACM International Conference on the Theory of Information Retrieval (ICTIR 2017), 2017.
- Program Committee Member and Reviewer, The 24th ACM International Conference on Information and Knowledge Management (CIKM 2015), 2015.
- Journal Reviewer, ACM Transactions on Internet Technology (ACM TOIT), 2013.
- 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, Torch, 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