

CONTACT INFORMATION 1-303, FIT Building, Phone: +86-13581700448
 Tsinghua University, E-mail: fanmiao.cslt.thu@gmail.com
 Beijing, 100084, P.R. China. Homepage: http://cslt.riit.tsinghua.edu.cn/mediawiki/index.php/Miao_Fan

RESEARCH INTERESTS **Text Mining & Social Computing**

- *Information Extraction*
- *Recommender Systems*

EDUCATION **New York University, New York, U.S.A. March 2015 - March 2016**

Joint-supervised Ph.D. Candidate in Computer Science.
Junior Research Scientist, Courant Institute of Mathematical Sciences.

- Research Topic: “Large-scale Entity Relation Extraction based on Low-dimensional Representations” (Proteus Project)².
- Advisor: Prof. Ralph Grishman³

Tsinghua University, Beijing, P.R. China. September 2012 - June 2017 (expected)

Ph.D. Candidate in Computer Science.

- Dissertation Topic: “Large-scale Knowledge Base Completion based on Low-dimensional Representations”.
- Co-advisors: Prof. Thomas Fang Zheng⁴ and Prof. Qiang Zhou⁵.
- GPA: **87.3**/100.

Beijing University of Posts and Telecommunications September 2008 - June 2012

B.Eng. in Software Engineering.

- Dissertation Topic: “Chinese Natural Question Generation Based on Wiki Knowledge”⁶, **Best B.Eng. Dissertation Award.**
- Advisor: Prof. Guoshi Wu⁷.
- GPA: **95.4**/100 (Rank **1**/97).

MY MONOGRAPHS **Miao Fan, Chao Li; *DIY Machine Learning Systems for Kaggle Competitions with Python Programming*, Printing, Tsinghua University Press.**

Miao Fan; *Distributed Machine Learning and Big Data Analysis with PySpark*, Drafting, Tsinghua University Press.

PATENTS **Miao Fan, Doo Soon Kim. Detecting Table Region in PDF Documents using Distant Supervision⁸. U.S. Patent submitted.**

Miao Fan, Qiang Zhou. The Approach to Social Tag Generation and Recommendation based on Monolingual Word Alignments. P.R.China Patent submitted.

¹<http://scholar.google.com/citations?user=aPlHReAAAAJ&hl=en>

²<http://nlp.cs.nyu.edu/index.shtml>

³<http://scholar.google.com/citations?user=blwKAKUAAAAJ>

⁴http://scholar.google.com/citations?user=H3MX_8IAAAAAJ&hl=en

⁵<http://cslt.riit.tsinghua.edu.cn/~qzhou/eng/index.htm>

⁶http://cslt.riit.tsinghua.edu.cn/mediawiki/images/a/a9/B.ENG_Dissertation_-Miao_Fan-.pdf

⁷<http://baike.baidu.com/view/9021485.htm>

⁸<http://arxiv.org/abs/1506.08891>

- Miao Fan**, Qiang Zhou, Thomas Fang Zheng, Ralph Grishman. Distributed Representation Learning for Knowledge Bases with Entity Descriptions. *Pattern Recognition Letters*⁹ (**IF:1.55**), Elsevier. SCIE Index. Full length article submitted.
- Miao Fan**, Qiang Zhou, Andrew Abel, Thomas Fang Zheng, Ralph Grishman. Probabilistic Belief Embedding for Large-scale Knowledge Population. *Cognitive Computation*¹⁰ (**IF: 1.44**), Springer. SCIE Index. Full length article submitted.
- Miao Fan**, Qiang Zhou, Thomas Fang Zheng. Learning Embedding Representations for Imperfect Knowledge Repository.¹¹ 2016 IEEE/WIC/ACM International Conference on Web Intelligence (*WI'16*). Full research paper submitted.
- Miao Fan**, Qiang Zhou, Thomas Fang Zheng. Distant Supervision for Entity Linking.¹² The 29th Pacific Asia Conference on Language, Information and Computation (*PACLIC'15*), pp. 79-86. Full paper, oral presentation.
- Miao Fan**, Kai Cao, Yifan He, Ralph Grishman. Jointly Embedding Relations and Mentions for Knowledge Population.¹³ The 10th Recent Advances in Natural Language Processing (*RANLP'15*), pp. 186-191. Poster paper.
- Miao Fan**, Qiang Zhou, Thomas Fang Zheng, Ralph Grishman. Large Margin Nearest Neighbor Embedding for Knowledge Representation.¹⁴ The 2015 IEEE/WIC/ACM Web Intelligence Conference (*WI'15*), pp. 53-59, 6-9 December 2015, Singapore. Full paper, oral presentation.
- Miao Fan**, Deli Zhao, Qiang Zhou, Zhiyuan Liu, Thomas Fang Zheng, Edward Y. Chang. Distant Supervision for Relation Extraction with Matrix Completion.¹⁵ The 52th Annual Meeting of the Association for Computational Linguistics (*ACL'14*), pp. 839-849. Full paper, oral presentation.
- Miao Fan**, Qiang Zhou, Emily Chang, Thomas Fang Zheng. Transition-based Knowledge Graph Embedding with Relational Mapping Properties.¹⁶ The 28th Pacific Asia Conference on Language, Information and Computing (*PACLIC'14*), pp. 328-337. Full paper, oral presentation.
- Miao Fan**, Qiang Zhou, Thomas Fang Zheng. Mining the Personal Interests of Microbloggers via Exploiting Wikipedia Knowledge.¹⁷ 15th International Conference on Intelligent Text Processing and Computational Linguistics (*CICLing'14*), pp. 188-200. Full paper, poster presentation.
- Miao Fan**, Qiang Zhou, Thomas Fang Zheng. Content-based Semantic Tag Ranking for Recommendation.¹⁸ The 2012 IEEE/WIC/ACM International Conference on Web Intelligence (*WI'12*), pp. 292-296. Short paper, oral presentation.
- Miao Fan**, Yingnan Xiao, Qiang Zhou. Bringing the Associative Ability to Social Tag Recommendation.¹⁹ *ACL'12 Workshop on Graph-based Methods for Natural Language Processing*, pp. 44-54. Workshop paper, oral presentation.
- Miao Fan**, Guoshi Wu. Opinion Summarization of Customer comments.²⁰ *Physics Procedia*, Vol 24 (2012) pp. 2222-2226, Elsevier.
- Miao Fan**, Guoshi Wu. Aspect Opinion Mining on Customer Reviews.²¹ *Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE'11)*

⁹<http://www.journals.elsevier.com/pattern-recognition-letters/>

¹⁰<http://link.springer.com/journal/12559>

¹¹<http://arxiv.org/pdf/1503.08155v1.pdf>

¹²<http://aclweb.org/anthology/Y15-1010>

¹³<http://arxiv.org/pdf/1504.01683v1.pdf>

¹⁴<http://arxiv.org/pdf/1504.01684v1.pdf>

¹⁵<http://arxiv.org/abs/1411.4455>

¹⁶<http://pan.baidu.com/s/1pJqLP1b>

¹⁷<http://pan.baidu.com/s/1bntmkpx>

¹⁸<http://pan.baidu.com/s/1kT41lyf>

¹⁹<http://pan.baidu.com/s/1o6wn7lg>

²⁰<http://www.sciencedirect.com/science/article/pii/S1875389212003690>

²¹<http://pan.baidu.com/s/1jGj47dG>

ACADEMIC
EXPERIENCE

November 19-20, 2011, Melbourne, Australia. *Advances in Intelligent and Soft Computing*, Vol 112, 2012, pp 27-33, Springer.

Enterprises Information Laboratory, Beijing University of Posts and Telecommunications.

Team leader

April, 2009 - May, 2011

Studying on developing *Feature-Opinion Recommender System*²² funded by the National Innovative Experimental Program.

Advisor: Prof. Guoshi Wu.

Natural Language Processing and Computational Social Science Lab, Tsinghua University.

Group member

May, 2011 - December, 2011

Studying on social tagging at Sina micro-blog platform.

Advisor: Prof. Maosong Sun²³ and Dr. Zhiyuan Liu²⁴.

The Association for Computational Linguistics.

Member

June, 2012 - Present

Center for Speech and Language Technology, Tsinghua University.

Ph. D. candidate

September, 2012 - Present

Exploring large scale knowledge extraction approaches which can be applied to build entity-based search engine or Google Knowledge Graph. Advisor: Prof. Thomas Fang Zheng and Prof. Qiang Zhou.

DolphinNLP Group & Social Card Project.

Founder

June, 2012 - December, 2013

ACM Transactions on Intelligent Systems and Technology²⁵ (IF²⁶: 9.39)

Reviewer

November, 2014

2014 International Doctoral Forum²⁷

*Technical Committee Track Chair*²⁸

December 5-7, 2014

C++ Programming, Tsinghua University.

*Teacher Assistant of Instructor Chao Li*²⁹

September, 2014 - January, 2015

Natural Language Processing³⁰, New York University.

*Teacher Assistant of Instructor Adam Meyers*³¹

September, 2015 - December, 2015

Special talks of "Statistical NLP: A Machine Learning Perspective"³².

²²<http://pan.baidu.com/s/1jGj47dG>

²³<http://scholar.google.com/citations?user=zIgTOHMAAAAJ&hl=en>

²⁴<http://scholar.google.com/citations?user=dTOv5uOAAAAJ>

²⁵<http://tist.acm.org/editors.html>

²⁶<http://www.spinellis.gr/blog/20140808/>

²⁷<http://www.nwpu-aslp.org/phdforum/2014/index.html>

²⁸<http://www.nwpu-aslp.org/phdforum/2014/content/orgcommittee.html>

²⁹<http://dbgroun.cs.tsinghua.edu.cn/lichao/>

³⁰<http://cs.nyu.edu/courses/fall15/CSCI-UA.0480-006/>

³¹<http://nlp.cs.nyu.edu/people/meyers.html>

³²<http://1drv.ms/1Saijf0>

The Eighth International Conference on Creative Content Technologies (CONTENT 2016)³³

*Technical Program Committee Member*³⁴

March 20-24, 2016

The 26th International Conference on Computational Linguistics (COLING 2016)³⁵

*Program Committee Member*³⁶

December 11-16, 2016

PROFESSIONAL
EXPERIENCE

Hulu Inc., Beijing, P.R. China.

Occupation: Research intern

May, 2013 - September, 2013

- *Brief Intro:* Researcher and computer model developer for extracting relation instances from movie plots.
- *Programming Language:* Python.
- *Supervisor:* Post Dr. Tao Xiong.

Microsoft Research Asia, Machine Learning Group³⁷, Beijing, P.R. China.

Occupation: Research intern

April, 2014 - July, 2014

- *Brief Intro:* Studying the link prediction in large-scale incomplete knowledge base, i.e. Freebase and WordNet, based on the low-dimensional embedding representation of entities and relationships without extra free texts. Developing parallel algorithms for large-scale knowledge embedding under the Microsoft Cosmos System³⁸.
- *Programming Language:* C++, C#, Microsoft-SCOPE (Cosmos).
- *Supervisor:* Researcher Jianwen Zhang³⁹.

Baidu Inc., Natural Language Group, Beijing, P.R. China.

Occupation: Research intern

November, 2014 - January, 2015

- *Brief Intro:* R&D for Baidu Chatbot (Smart Search) System⁴⁰. Modeling/designing multi-round dialogue algorithms/system for chatbot.
- *Programming Language:* C++.
- *Supervisor:* Senior Researcher Shiqi Zhao⁴¹ and Dr. Rui Yan⁴².

Bosch Research, Palo Alto, CA. U.S.

Occupation: Research intern authorized by NYU

June, 2015 - September, 2015

- *Brief Intro:* Table detection and extraction from PDF files, such as electronic manuals and academic articles⁴³. We also design the end-to-end system (PDFExtraction) for Bosch Research.
- *Programming Language:* Java.

³³<http://www.iaria.org/conferences2016/CONTENT16.html>

³⁴<http://www.iaria.org/conferences2016/ComCONTENT16.html>

³⁵<http://coling2016.anlp.jp/>

³⁶<http://coling2016.anlp.jp/>

³⁷<http://research.microsoft.com/en-us/groups/ml/>

³⁸<http://blogs.msdn.com/b/seliot/archive/2010/11/05/cosmos-petabytes-perfectly-processed-perfunctorily.aspx>

³⁹<http://research.microsoft.com/en-us/people/jiazhan/>

⁴⁰<http://baike.baidu.com/view/14785371.htm?fr=aladdin>

⁴¹<http://ir.hit.edu.cn/~zhaosq/>

⁴²<https://sites.google.com/site/ruiyan516/>

⁴³We collect all free published articles (9,466) from ACL Anthology: <http://aclweb.org/anthology/>

- *Supervisor*: Senior Research Engineer, Doo Soon Kim⁴⁴.

DailyCast, Beijing, P.R.China.

Occupation: Leading Researcher and Developer of DailyCast Lab

May, 2016 - Present

- *Brief Intro*: Leading the research and development of short-video recommender systems, and the “event-based” log collection and analytic system in Dailycast Lab⁴⁵.
- *Programming Language*: Python (Anaconda + Spark) and Java.

HONORS AND AWARDS

HUAGUANG 2nd Class Ph.D. Scholarship (1,000 CNY), 2015.

HUAWEI Ph.D. Fellowship (10,000 CNY), 2014.

RIIT Graduate Research Award (3,000 CNY), Tsinghua University, 2014.

Excellent Graduate Award in Tsinghua University, 2013.

National Scholarship for Undergraduates (8,000 CNY), 2011.

The IBM Chinese Excellent Undergraduate Scholarship (4,000 CNY), 2011.

Excellent Undergraduate Award in Beijing, China, 2011.

Excellent Academic Paper Award of 11th Creative Award in Beijing University of Posts and Telecommunications, 2011.

Great Leader Award of National Innovative Experimental Program, China, 2011.

MCM/ICM (Mathematical Contest in Modeling/ Interdisciplinary Contest in Modeling) **Meritorious Winner** (First Prize), 2011.

“Tang Jun&Sun Chunlan” Enterprise Scholarship (5,000 CNY), 2011.

PARTICIPATED CONTESTS

Kaggle Competitions

April, 2015 - Present

- *Profile*: <https://www.kaggle.com/michaelfan>
- *Rank*: 1,978th / 533,172 (Top 0.5%)
- *Awards*: Top25%(Bronze) * 4
- *Tools*: I use *NLTK*⁴⁶, *Scikit-learn*⁴⁷, *Pandas*⁴⁸, *Xgboost*⁴⁹, *TensorFlow*⁵⁰ and *Spark*⁵¹ to process the data, build new models and make predictions.

PROOFREADER

Yoshua Bengio, Ian Goodfellow and Aaron Courville; *Deep Learning*⁵², MIT Press book in preparation. I also contribute feedback on several chapters⁵³.

HOBBIES

Swimming, playing table tennis and the piano.

⁴⁴<https://sites.google.com/site/2soonk/>

⁴⁵<http://www.dailycast.tv>

⁴⁶Natural Language Toolkit: <http://www.nltk.org/>

⁴⁷Machine learning library in Python:<http://scikit-learn.org/stable/index.html>

⁴⁸Python data analysis library: <http://pandas.pydata.org/>

⁴⁹Extreme Gradient Boosting: <https://xgboost.readthedocs.org/en/latest/>

⁵⁰Google Deep Learning Framework: <http://www.tensorflow.org/>

⁵¹Lightning-fast cluster computing: <http://spark.apache.org/>

⁵²<http://www.iro.umontreal.ca/~bengioy/dlbook/>

⁵³<http://goodfeli.github.io/dlbook/contents/acknowledgements.html>