

Fan Li

Email: hustlfhustlf@gmail.com
Cell: (412)535-1412
www.cs.cmu.edu/~hustlf

Expertise

- Machine Learning, Large-scale distributed Data Mining
- Personalized Recommendation, Web-Search algorithms, General Information Retrieval,
- Natural Language Processing and Text Mining

Education

- Ph.D., Carnegie Mellon University** 2001 – 2007
Language Technologies Institute, School of Computer Science
- Advisor: Prof. Yiming Yang
 - Dissertation Title: Structure learning with large sparse undirected graphs and its applications
- Master of Science, Carnegie Mellon University** 2001 – 2004
Language Technologies Institute, School of Computer Science
- Master of Science, Tsinghua University** 1998 - 2001
Department of Computer Science and Technology
- Advisor: Prof. Yuchang Lu
 - Dissertation Title: New approaches in text categorization
- Bachelor of Science, Huazhong Univ. of Science and Technology** 1994- 1998
Department of Computer Science and Technology

Honors and Awards

- **Yahoo Super Star Award 2009** (with other team members), Yahoo! Labs 2009
 - The highest yearly recognition given by the company to employees
- **Yahoo Patent Incentive Award 2009**, Yahoo! Labs Sep. 2009
- **Graduate Research Fellowship**, Carnegie Mellon University 2001 - 2007
- **Graduate Research Fellowship**, Tsinghua University 1998 - 2001
- **Best undergraduate thesis award**, Huazhong Univ. of Science and Technology 1998
 - Top 2.5%, 50 out of 2000 candidates

Experience

- Technical Lead and Senior Scientist in Yahoo! Labs** 2007–Present
- I work in Yahoo! Labs as a senior scientist and a technical lead, starting from designing algorithms for the Yahoo! core search engine, to leading a team building solutions for personalized recommendation products. I enjoy designing distributed algorithms on Hadoop to explore extremely large-scale datasets.

(Hadoop Map-reduce/Pig)

Some of the projects I lead include:

- Personalized news recommendation (currently used in yahoo homepage)
- Developing web-search algorithms for Yahoo search engine. (Related works published as a paper in WWW2010).
- Automatic metric Learning for Web-Search to maximize User Satisfaction. (Related works published as papers in CIKM 2009 and SIGIR 2009)
- Designing click-based query-URL features to improve Yahoo web search. (Related works published as a paper in EMNLP 2009).
- Semantic Text Mining. Related works lead to two (co-authored) patent applications:
 - Semantic text matching: Y05820US00
 - Auto-classify Webpage Segments: Y05503US00

Research Assistant in Information Retrieval Group, Carnegie Mellon University 2001 – 2007

- **Large-scale structure learning from relational data and its applications on text mining and bioinformatics.** I developed a new framework for learning complex networks from observed data. This technique can be used in many fields, including text mining, social network analysis, bioinformatics, marketing analysis, etc. My works are published as papers in SIGIR 2004 workshop and AAAI 2005.
- **Structure learning and sparse regression within non-linear models.** I developed this new technique for non-linear feature selection and pattern recognition in high dimensional space. Works in this line published as a paper in NIPS 2005.
- **Large-scale text categorization.** Developed various large-scale text categorization systems and conducted experiments on several benchmark corpuses. Our experimental results on the RCV1 dataset have become a popular baseline in the field of automated text categorization. Works in this line published as papers in ICML 2003 and Journal of Machine Learning Research 2004.
- **Feature selection Analysis.** Investigated different types of feature selection algorithms and developed a novel approach that has been applied successfully to extract disease-related genes from genome micro-array data. This approach has also been applied to improve the text categorization performance by eliminating redundant or irrelevant text features. Works in this line are published as papers in SAC 2005, SIGIR 2005 and Journal Bioinformatics 2005.

Research Intern in Yahoo! Lab, Santa Clara, CA Jul. 2006 – Sep. 2006

- Developed novel algorithms to classify web pages using hyperlinks and anchor text information within a graph regularization framework.

Research Assistant in the State Key Lab of Intelligent Technology at Tsinghua University 1998 – 2001

- Developed text-mining systems that extract association rules from text corpuses. My related works in this line published as papers in WECWIS 2002 and Tsinghua Academic Journal.

Professional Service

Program Committee member:

- ACM 17th Conference on Information and Knowledge Management, CIKM 2008/2009/2010/2011

Reviewer for Journals:

- Journal of Machine Learning Research
- IEEE Intelligent Systems on Web Mining

- IEEE Transactions on Pattern Analysis and Machine Intelligence
- Journal Bioinformatics
- IEEE/ACM Transactions on Computational Biology and Bioinformatics

Reviewer for Conferences:

- ACM SIGIR 2003,2004,2006
- ACM SIGKDD 2003

Publication

Fan Li, Xin Li, Jiang Bian, Zhaohui Zheng, Optimizing Unified Loss for Web Ranking Specialization, *CIKM 2010, accepted*

Jiang Bian, Xin Li, **Fan Li**, Hongyuan Zha, Zhaohui Zheng, **Ranking Specialization for Web Search: A Divide-and-Conquer Approach by Using Topical RankSVM**, *Proceedings of the 19th International World Wide Web Conference (WWW 2010) (full paper)*

Xin Li, **Fan Li**, Shihao Ji, Zhaohui Zheng, Yi Chang, Anlei Dong, **Incorporating Robustness into Web Ranking Function Selection**, *The 18th ACM Conference on Information and Knowledge Management (CIKM 2009) (short paper)*

Huihsin Tseng, Longbin Chen, **Fan Li**, Ziming Zhuang, Lei Duan, Belle Tseng, **Click-based Query-URL N-gram Features in Web Ranking**, *The 2009 Conference on Empirical Methods on Natural Language Processing (EMNLP 2009) (full paper)*

Fan Li, Xin Li, Shihao Ji, Zhaohui Zheng, **Comparing Both Relevance and Robustness in Selection of Web Ranking Functions**, *32th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR 2009) (poster)*

Fan Li, Yiming Yang and Eric P. Xing. **Inferring regulatory networks using a hierarchical Bayesian graphical Gaussian model**, *CMU-MLD Technical Report 06-117, 2006*

Fan Li. **Structure learning with large sparse undirected graphs and its applications**, *Thesis, Carnegie Mellon University, 2006*

Fan Li, Yiming Yang, Eric P. Xing. **From Lasso regression to feature vector machine**, *Advances in Neural Information Processing Systems 18 (NIPS2005) (full paper)*

Fan Li, Yiming Yang. **Use modified lasso regressions to learn large undirected graphs in a probabilistic framework**, *20th National Conference on Artificial Intelligence (AAAI 2005) (full paper)*

Fan Li, Yiming Yang. **An Analysis of Recursive Feature Elimination Methods for Statistical Classification**, *27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR 2005) (poster)*

Fan Li, Yiming Yang. **Using recursive Classification to discover predictive features**, *20th Annual ACM Symposium on Applied Computing (ACM SAC 2005) (full paper)*

Fan Li, Yiming Yang. **Analysis of recursive gene selection approaches from micro-array data**. *Bioinformatics, 2005 21(19):3741-3747(full journal paper)*

David D. Lewis, Yiming Yang, Tony Rose, **Fan Li**. **RCV1: A New Text Categorization Test Collection**, *Journal of Machine Learning Research* 5 (2004) 361-397 (full journal paper)

Fan Li, Yiming Yang. **Recovering genetic regulatory networks from micro-array data and location analysis data**, *Journal Genome Informatics*, 2004, Vol. 15, No.2 (full journal paper)

Fan Li, Yiming Yang. **Learning the Structure of Linear Bayesian Networks with a Large Number of Variables**. *Search and Discovery in Bioinformatics workshop, 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR 2004)* (workshop paper)

Fan Li, Yiming Yang. **A loss function analysis for classification methods in text categorization**. *20th International Conference on Machine Learning (ICML 2003)* (full paper)

Fan Li, Mingyu Lu, Yuchang Lu. **The Research on Some New Feature Selection Approaches in Text Categorization**, *Tsinghua Academic Journal (Natural Science Version)*, Vol.41, No.7, pp 98-101, 2001.7

Mingyu Lu, Qiang Zhou, **Fan Li**, Yuchang Lu, Lizhu Zhou. **Recommendation of Web Pages Based on Concept Association**. *WECWIS 2002*: 221-227