

Research Interests

Machine Learning, Data Mining, Artificial Intelligence

Education

- 2010.09 - 2013.03 School of Computer Science and Technology, Beihang University (BUAA)
Master of Engineering in Computer Science and Technology
Overall GPA: 3.58/4, Top 15% Ranking
- 2006.09 - 2010.07 School of Computer Science and Technology, Shandong Jianzhu University (SDJZU)
Bachelor of Engineering in Network Engineering;
Overall GPA: 3.62/4; Top 1% Ranking

Research Experience

- 2013.09 - Present **R&D (Research & Development) Engineer at Department of LBS (Location Based Service), Baidu Inc.**
Clustering, Machine Learning, Topic Model, Recommender System
Advisor: Senior R&D Engineer Zhengxing Zhuo
- *Mobile Trajectory Mining*
 - a) *Reviewed and experimented* large-scale clustering methods for trajectory data under MapReduce.
 - b) *Investigated and updated* incremental clustering methods for user's streaming trajectory data.
 - c) *Modeled* user's life patterns for POI (place of interest) recommendation by mining his/her significant places and computing temporal distribution on those places.
- 2010.10 – 2012.07 **Research Assistant at Department of Computer Science, BUAA**
Clustering, Topic Model, Tagging, Recommender System, Complex network
Advisor: Prof. Deyi Li
- *Social Media Recommendation*
 - a) *Investigated* statistical features of user's social tagging behaviors on dataset crawled from Flickr and *verified* that power law plays an important role in depicting those social behaviors.
 - b) *Proposed* a topic discovery method using clustering strategy particular for social tags, then evaluated it.
 - c) *Introduced* a framework to recommend social media, such as tags, photos and places, which combined collaborative filtering, content-based recommendation and random walk model.
- 2008.08 – 2009.06 **Research Assistant at Department of Computer Science, SDJZU**
Text Mining, Data Mining, Ontology
Advisor: Associate Prof. Ling Song
- *Measure Similarity between Two XML Documents*
 - a) *Reviewed* different kinds of similarity measures (Euclidean distance, Cosine distance, KL divergence etc.) to study the similarity between XML Documents
 - b) *Implemented* a variation of LCS (Longest Common Substring) algorithm to measure the similarity between two XML documents' structures.
 - c) *Designed* experiments to compare the proposed approach with state-of-art methods.

Work Experience

- 2013.09 - Present **R&D Engineer, Baidu Inc., Beijing**
- *Responsible* for daily processing massive mobile app's (such as Baidu Map, Baidu Travel) location-related log (about billions of records daily) to mine the regions where users stay by tools like Hadoop and Spark.
 - *Designed and developed* a framework, based on DBSCAN and Hierarchical Clustering, to mine users' personal significant places. Then built classification models to predict user's significant places' types (Home, Company, Favorite Restaurants, etc.).
 - *Proposed and implemented* a combined strategy to match coordinates of user's trajectory with POI in the context of one's query and Wi-Fi connection information.
 - *Evaluated and improved* LDA topic model's performance on POI recommendation.
- 2012.03 - 2012.07 **Intern R&D Engineer, 37Degree Inc., Beijing**
- *Designed and developed* a micro-blog crawler particular for Weibo based on Gevent (a coroutine-based network library for Python).
 - *Responsible* for some NLP works on micro-blog, like word segmentation and feature selection.
 - *Explored* topic models and classification models' performance in predicting micro-blog's classes.
- 2011.10 - 2012.02 **Intern SDE (Software Development Engineer), Amazon, Beijing**
- *Developed* some new features of Amazon GDMS (Global Delivery Management System) corporately.
 - *Upgraded* GDMS 'a customer feedback module independently.
- 2011.07 -2011.09 **Intern R&D Engineer, Sogou Inc., Beijing**
- *Responsible* for daily parsing query and click data from log and storing them into database.
 - *Developed* a simple display platform which shows the statistics of query and websites visit.
 - *Explored* methods on personal news recommendation.

Publications

- Jinpeng Chen, Yu Liu, Jun Hu, **Wei He**, Deyi Li, "A Novel Framework for Improving Recommender Diversity", *BSI-UCBCN 2013 Workshop (conjunction with the 2013 Pacific-Asia Conference on Data Mining and Knowledge Discovery-PAKDD 2013)*, vol 8178, no., pp 129-138, Brisbane, Australia, April 14, 2013
- **Wei He**; Runsheng Li; Zhenyu Wu; Jun Hu; Yu Liu, "Generating landmark overviews with geo-tagged web photos," *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*, vol., no., pp.2644,2649, 14-17 Oct. 2012
- Runsheng Li; **Wei He**; Zhenyu Wu; Jun Hu; Yu Liu, "Modeling user's temporal dynamic profile in micro-blogging using clustering method," *Information Science and Service Science and Data Mining (ISSDM), 2012 6th International Conference on New Trends in*, vol., no., pp.808,812, 23-25 Oct. 2012
- Zhaowen Tao; Jun Hu; **Wei He**; Runsheng Li; Deyi Li, "Modeling user's preference in folksonomy for personalized search," *Cloud and Service Computing (CSC), 2011 International Conference on*, vol., no., pp.55,59, 12-14 Dec. 2011
- Song Ling; **He Wei**; Yang Tongjiang; Liu Zhendong, "A Study on XML Path Similarity," *Management and Service Science, 2009. MASS '09. International Conference on*, vol., no., pp.1,4, 20-22 Sept. 2009
- Song Ling; Li Shengen; Lv Qiang; **He Wei**; Yang Tongjiang, "An Approach for Measuring Similarity between XML Documents," *Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09. Sixth International Conference on*, vol.7, no., pp.410,414, 14-16 Aug. 2009

Technical Skills

Languages: **Skilled** in *Python, C, java, C++*; **Familiar with** *Shell, Php, Js, Html, Css, Sql*;
Keen learner on *Scala, Go*;

Large-scale Data: **Skilled** in *Hadoop, Hdfs, Hive*; **Keen Learner** on *Spark, Storm, Neo4j*.

Operating System: **Skilled in** *Linux, Windows*.

DataBase: **Familiar with** *Mysql, LevelDB, Redis, RockDB, MongoDB*.

Miscellaneous: **IDE**: *Vim*(very often), *Eclipse*(sometimes); **Web Dev**: *Django, Tornado, WordPress, JQuery*.
ML Libaray: *Numpy, Scipy, Scikit-Learn, Gensim, Matplotlib, GibbsLDA*

Honors and Awards

- 2014.05 Season Star of LBS-DA (Data Analysis) Group of *Baidu Inc.*
- 2011.09 Second-Class Scholarship of *BUAA* (30%)
- 2010.05 Top 10 Most Outstanding Undergraduates of *Shandong Province* (only 10 in *Shandong province* annually)
- 2009.10 Third-prize of “College Cup” (*national academic science and technology Olympic-like competition*).
- 2008.09 National Scholarship awarded by *China Education Ministry* (2%)
- 2007.10 First-Class Scholarship of *SDJZU* (10%)

References

Ling,Song, Associate Professor

School of Computer Science and Technology
Shandong Jianzhu University
(086) 136 2541 9436
song_ling@sina.com

JinPeng Chen, Ph.D student

School of Computer Science and Technology
Beihang University
(086) 138 1141 5176
chenjinpeng@nlsde.buaa.edu.cn

JiaGuang Jin, R&D Engineer

LBS Department
Baidu Inc.
(086) 152 0161 6149
jinjiaguang@baidu.com