Chuancong Gao

Phone: +1 (604) 368-2443
Email: chuancong@gmail.com

8888 University Drive
Burnaby, BC, Canada V5A 1S6

Work Experience

Simon Fraser University — Burnaby, BC, Canada

Research Assistant in Intelligent Data Engineering and Analytic Lab — Jan 2013 – Now

- Devised an unsupervised algorithm framework for automatically parsing log-like text to data table. Empirical results show great or even near-perfect accuracies for various kinds of data.
- Developed multiple unsupervised algorithm frameworks for automatically joining data tables without schema and joining similar objects without similarity threshold, by utilizing different user result set preferences. Empirical results show significant improvements on both accuracy and running time, comparing to state-of-the-art baselines.

Lead Research Developer in Genomics and Networks Analysis Lab — Jan 2014 – Now

- Lead the development of an interactive data analytical system, for collecting, indexing, mining, and visualizing very large scale data from PubMed, Twitter, NCAA, GoFundMe.com, or user upload.
  - Design and implement the system infrastructure. It features a unified HTML5-based front-end, multiple back-ends for different data sources, Redis-based job queue, and Elasticsearch for full-text indexing and search.
  - Devise and implement various state-of-the-art analyzing/mining algorithms.
  - Implement and deploy data crawlers for NCAA, GoFundMe.com, etc.
  - Lead, mentor, and recruit other developers.

Yahoo! Research — Sunnyvale, CA, United States

Intern Scientist in Search Science Group — Jun 2015 – Sep 2015

- Worked on mining distinct multi-dimensional hierarchical behavior patterns for mobile search. Pattern involve time, location, and entity, and are further applied to help predict the future queries.

Nanyang Technological University — Singapore

Project Officer (Researcher) in Center for Advanced Information Systems — Mar 2011 – Jul 2012

- Worked on predicting moving objects' future trajectories, with significant improvement on accuracy comparing to state-of-the-art baselines.

Max-Planck Institute for Informatics — Saarbrücken, Saarland, Germany

Visiting Scholar in Database Department — Aug 2010 – Jan 2011

- Devised an algorithm for computing the top-k interesting phrases w.r.t. any ad-hoc query (from millions of documents) almost instantly, with half the index size and faster querying time comparing to baselines.

Tsinghua University — Beijing, China

Research Assistant in Database Lab — Sep 2007 – Jul 2010

- Devised several pattern mining algorithms, for generator sequence, generator itemset on sliding window, discriminative itemset on uncertain data, closed sequence on sliding window, etc.
- Devised and evaluated multiple applications on various types of patterns, including query completion, classification, etc., with the accuracies improved significantly.

Education

Simon Fraser University — Burnaby, BC, Canada

Ph.D. in Computing Science — Jan 2013 – Now

- Graduate Student Fellowship (Fall 2013, Summer 2015, Summer 2016, Fall 2016)
- Helmut & Hugo Eppich Family Scholarship (Spring 2017)
- Borden Ladner Gervais Scholarship (Spring 2016)

Tsinghua University — Beijing, China

Master in Computer Science & Technology — Sep 2007 – Jul 2010

- Outstanding Master Graduate (University Level) & Outstanding Master Thesis (University Level)
- Morgan Stanley Scholarship (University Level, 2008)

Shandong University — Jinan, Shandong, China

Bachelor in Computer Science & Technology — Sep 2003 – Jun 2007

- Outstanding Graduate (Department Level) & Outstanding Dissertation (University Level)
- Undergraduate Student Scholarship (2005, 2006)
- Entrance Scholarship (University Level)

PUBLICATIONS

Conference Papers

1. Chuancong Gao, Jian Pei, Jiannan Wang. *Parsing Text to Relational Table via Hidden Schema Discovery*. (Preparing for Submission)
2. Chuancong Gao, Jian Pei, Jiannan Wang, Yi Chang. *Schema-less Join for Result Set Preferences*. (Under Review)
3. Chuancong Gao, Jiannan Wang, Jian Pei, Rui Li, Yi Chang. *Preference-driven Similarity Join*. 17th IEEE/WIC/ACM International Conference on Web Intelligence, 2017.
4. Chuancong Gao, Sebastian Michel. *Top-k interesting phrase mining in ad-hoc collections using sequence pattern indexing*. 15th International Conference on Extending Database Technology, 2012.
5. Chuancong Gao, Jianyong Wang, Qingyan Yang. *Efficient Discovery of Closed Sequential Patterns over Stream Sliding Window*. 11th IEEE International Conference on Data Mining, 2011.
6. Chuancong Gao, Qingyan Yang, Jianyong Wang. *SEQUEL: Query Completion via Frequent Pattern Mining on Multi-Column Structural Data*. 19th ACM Conference on Information and Knowledge Management, 2010. (Demo Paper)
7. Chuancong Gao, Jianyong Wang. *Direct Mining of Discriminative Patterns for Classifying Uncertain Data*. 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010.
8. Chuancong Gao, Jianyong Wang. *Efficient Itemset Generator Discovery Over a Stream Sliding Window*. 18th ACM Conference on Information and Knowledge Management, 2009.
9. Chuancong Gao, Jianyong Wang, Yukai He, Lizhu Zhou. *Efficient Mining of Frequent Sequence Generators*. 17th International Conference on World Wide Web, 2008. (Poster Paper, **Best Poster Award**)

Journal Papers

1. Xiaoning Xu, Chuancong Gao, Jian Pei, Ke Wang, Abdullah Al-Barakati. *Continuous Similarity Search for Evolving Queries*. Knowledge and Information Systems, 2016.

PROFESSIONAL SERVICES

Journal Reviewer

2013, 2016	Data Mining and Knowledge Discovery
2013 – 2014, 2016	IEEE Transactions on Knowledge and Data Engineering
2015 – 2016	Journal of Computer Science and Technology
2014 – 2017	Knowledge and Information Systems
2015 – 2016	Neurocomputing
2015, 2017	Pattern Recognition Letters

External Conference Reviewer for Multiple Top-tier Conferences

Publication Committee

2011 Pacific Asia Conference on Language, Information and Computation

ACTIVITIES

ACM SIGKDD KDD Cup 2008

7th among Teams in Challenge 1: *Early Detection of Breast Cancer from X-ray Images*

- Devised a boosting-based classification model directly optimized towards the evaluation metric (modified ROC).

TECHNICAL SKILLS

- C#, C++, HTML & CSS, JavaScript, LaTeX, Python, Shell, SQL
- Apache Hive, Elasticsearch, MySQL & SQLite, Redis