Junbo Zhang

Units 525-530, Core Building 2   (852) 5519-5158
Hong Kong Science Park   jbzhang@cse.cuhk.edu.hk
Shatin, Hong Kong   http://www.lucktroy.org

Education

09/2009 - present   **Southwest Jiaotong University**, Chengdu, Sichuan, China

Ph.D. candidate in Compute Science, advised by Prof. Tianrui Li & Prof. Yi Pan

Special Grade Scholarship for 5 years

09/2005 - 06/2009   **Southwest Jiaotong University**, Chengdu, Sichuan, China

Bachelor of Telecommunication Engineering in **the Mao Yisheng Honors Class**

Undergraduate Scholarship for 4 years

Experience

05/2013 - present   **The Chinese University of Hong Kong**, Shatin, New Territories, Hong Kong

Research Assistant with Prof. John C.S. Lui

Research and develop large-scale deep learning and feature engineering algorithms.

05/2013 - present   **Huawei Noah's Ark Lab**, Shatin, New Territories, Hong Kong

Research Intern with Dr. Wei Fan

Research and develop large-scale deep learning and feature engineering algorithms.

- Research on Transformational Learning (TL, the generation of deep learning). Using deep learning techniques with decision tree, model-based search tree or graphical models, TL can learn structure & high-level knowledge representations from most kinds of data, included sequences, graphs, images, etc. `The draft is being prepared.`

- Proposed a novel deep learning model named SUGAR. SUGAR regularizes the network construction by utilizing similarity or dissimilarity constraints between data pairs, rather than sample-specific annotations. Such kind of side information is more flexible and greatly mitigates the workload of annotators. Secondly, unlike prior works, SUGAR decouples the supervision information and intrinsic data structure. SUGAR includes two heterogeneous networks, each of which encodes either supervision or unsupervised data structure respectively. `The work has been presented at` *KDD 2014*. `Code:` https://github.com/lucktroy/sugar.

- Proposed a novel algorithm (FSP-MbT) for mining discriminative and essential frequent sequential patterns via model-based search tree. It is the core component of the subsequent prediction model, which is used for the consumer complaint problem in the telecommunications industry. The related prediction system is being deployed and tested for real-world application, which includes millions of consumers. `Patent Pending.`

- Proposed a locally linear deep learning model (LLDL) for large-scale stellar spectrum recognition. The characteristics of stellar spectra: the massive volume, ultrahigh dimensionality ($>5K$), and significantly low signal/noise ratio (<10). Using GPU, stochastic gradient descent (SGD) and Dropout, LLDL can model the locally linear structures found in the stellar spectra. Experimental results show that LLDL achieves up to 100% improvements over both popular deep neural networks and shallow SVM and Logistic Regression. `Paper is under review.`

02/2012 - 02/2013   **Georgia State University**, Atlanta, GA, USA

Research Assistant with Prof. Yi Pan

Research and develop parallel algorithms, large-scale algorithms in cloud computing and GPU cluster.

- Participated as a mentor in 8-week NSF REU Undergraduate Summer Research Program.

- Proposed large-scale feature selection algorithms, provide HADOOP and SPARK implementations.

- Proposed a composite relation for data fusion & a novel parallel matrix algorithm using multi-GPU.

08/2011 - 09/2011 Belgian Nuclear Research Centre (SCK-CEN), Mol, Belgium

Intern with Dr. Klaas van der Meer

Research and develop incremenal and parallel algorithm for feature selection and knowledge acquisition.

Publications

Conference Papers

1. **Junbo Zhang**, Guangjian Tian, Yadong Mu, Wei Fan.
   Supervised Deep Learning with Auxiliary Networks.
   Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD 2014**), New York, USA, 2014, pp. 353-361.
   (AR: $151/1036 = 14.6\%$)

2. **Junbo Zhang**, Tianrui Li, Yi Pan.
   PLAR: Parallel Large-scale Attribute Reduction on Cloud Systems.
   The 14th International Conference on Parallel and Distributed Computing, Applications and Technologies (**PDCAT 2013**), Taipei, 2013, pp. 184-191.

3. **Junbo Zhang**, Tianrui Li, Yi Pan.
   Parallel Rough Set Based Knowledge Acquisition Using MapReduce from Big Data.
   ACM SIGKDD 2012 Workshop on Big Data Mining (**BigMine 2012**), Beijing, China, pp. 20-27,2012.

Journal Papers

1. **Junbo Zhang**, Yun Zhu, Yi Pan, Tianrui Li.
   Efficient Parallel Boolean Matrix Based Algorithms for Computing Composite Rough Set Approximations.
   Information Sciences, 1st revision.

2. **Junbo Zhang**, Jian-Syuan Wong, Yi Pan, Tianrui Li.
   A Parallel Matrix-based Method for Computing Approximations in Incomplete Information Systems.
   Accepted for publication in IEEE Transactions on Knowledge and Data Engineering (**TKDE**).

3. **Junbo Zhang**, Tianrui Li, Yi Pan, Chuan Luo, Fei Teng.
   A Parallel and Incremental Algorithm for Updating Knowledge Based on Rough Sets in Cloud Computing Platform.
   Accepted for publication in **Journal of Software**. (in Chinese)

4. **Junbo Zhang**, Jian-Syuan Wong, Tianrui Li, Yi Pan.
   A Comparison of Parallel Large-scale Knowledge Acquisition Using Rough Set Theory on Different MapReduce Runtime Systems.
   International Journal of Approximate Reasoning, vol. 55, no. 3, pp. 896-907, 2014.

5. **Junbo Zhang**, Tianrui Li, Hongmei Chen.
   Composite Rough Sets for Dynamic Data Mining.
   Information Sciences, vol. 257, pp. 81-100, 2014.

6. **Junbo Zhang**, Dong Xiang, Tianrui Li, Yi Pan.
   M2M: A Simple Matlab-to-MapReduce Translator for Cloud Computing.
   Tsinghua Science and Technology, vol 18, no. 1, pp. 1-9, 2013.

7. **Junbo Zhang**, Tianrui Li, Da Ruan, Zizhe Gao, Chengbing Zhao.
   A Parallel Method for Computing Rough Set Approximations.
   Information Sciences, vol. 194, pp. 209-223, 2012.

8. **Junbo Zhang**, Tianrui Li, Da Ruan, Dun Liu.
   Rough Sets Based Matrix Approaches with Dynamic Attribute Variation in Set-valued Information Systems.
   International Journal of Approximate Reasoning, vol. 53, no. 4, pp. 620-635, 2012.

9. **Junbo Zhang**, Tianrui Li, Da Ruan, Dun Liu.
   Neighborhood Rough Sets for Dynamic Data Mining.
   International Journal of Intelligent Systems, vol. 27, no. 4, pp. 317-342, 2012.

10. Dun Liu, Tianrui Li, **Junbo Zhang**.
    Incremental Updating Approximations in Probabilistic Rough Sets Under the Variation of Attributes
    Knowledge-Based Systems, vol. 73, pp. 81–96, 2015.

11. Anping Zeng, Tianrui Li, Dun Liu, **Junbo Zhang**, Hongmei Chen.
    A Fuzzy Rough Set Approach for Incremental Feature Selection on Hybrid Information Systems.
    Fuzzy Sets and Systems, vol. 258, pp. 39–60, 2015.

12. Dun Liu, Tianrui Li, **Junbo Zhang**.
    A Rough Set-based Incremental Approach for Learning Knowledge in Dynamic Incomplete Information Systems.
    International Journal of Approximate Reasoning, vol 55, no. 8, pp. 1764–1786, 2014.

13. Anping Zeng, Tianrui Li, **Junbo Zhang**, Hongmei Chen.
    Incremental Maintenance of Rough Fuzzy Set Approximations under the Variation of Object Set.
    Fundamenta Informaticae, vol. 132, no. 3, pp. 401-422, 2014.

14. Yi Pan, **Junbo Zhang**.
    Parallel Programming on Cloud Computing Platforms: Challenges and Solutions.
    KITCS/FTRA Journal of Convergence, vol. 3, no. 4, pp. 23-28,2012.

15. Dun Liu, Tianrui Li, Da Ruan, **Junbo Zhang**.
    Incremental learning optimization on knowledge discovery in dynamic business intelligent systems.
    Journal of Global Optimization, vol. 51, no. 27: pp. 325-344, 2011.

Book Chapter

1. **Junbo Zhang**, Tianrui Li, Da Ruan.
   Rough Sets Based Incremental Rule Acquisition in Set-Valued Information Systems.
   Autonomous Systems: Developments and Trends, pp. 135-146, 2012.

Awards & Honors

2012, 2013 National Scholarship, China.

2012 "Si Shi Yang Hua 埃实扬华" Medal in Southwest Jiaotong University, 2012. (Top 1/1000, the students' top honor of Southwest Jiaotong University).

2009-2014 Special Grade Scholarship for PhD Students, Southwest Jiaotong University.

2012 First Prize in the 9th National Postgraduate Mathematical Contest in Modeling, China.

2006, 2007 The President Scholarship, School of Information Science and Technology, Southwest Jiaotong University, China.

2005-2009 Undergraduate Scholarship of Southwest Jiaotong University, China.

Projects

11/2012 - present Research on dynamic knowledge discovery techniques and efficient algorithms under granular computing, the Fostering Foundation for the Excellent Ph.D. Dissertation of Southwest Jiaotong University, China. Leader and principal investigator.

10/2012 - 10/2013 Dynamic knowledge discovery system based on rough sets in cloud computing environments, the Science and Technology Planning Project of Sichuan Province, China. Leader and principal investigator.

10/2011 - 09/2012 Research on dynamic knowledge discovery techniques under granular computing and probabilistic rough sets, and its fast algorithms based on cloud computing, the Doctoral Innovation Funding Project of Southwest Jiaotong University, China. Leader and principal investigator.

10/2011 - 09/2012 Research on dynamic knowledge discovery system based on cloud computing and rough sets, the Young Software Innovation Foundation of Sichuan Province, China. Leader and principal investigator.

05/2012 - 07/2012 8-week NSF REU Undergraduate Summer Research Program hosted by the Department of Computer Science, Georgia State University, the National Science Foundation, USA. Participated as a mentor.

01/2012 - present Research on dynamic updating knowledge theories and algorithms based on granular computing, the National Natural Science Foundation of China.

01/2012 - present Research on composite rough set models and algorithms of knowledge discovery, the National Natural Science Foundation of China.

01/2009 - 12/2011 Research on incremental learning theories and methods based on granular computing, the National Natural Science Foundation of China.

Professional Activities

Review for Journal

- ACM Transactions on Knowledge Discovery from Data (TKDD)
- IEEE Transactions on Computers (TOC)
- IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)
- IEEE Transactions on NanoBioscience
- World Wide Web Journal (WWWJ)
- International Journal of Cloud Computing (IJCC)
- International Journal of Bioinformatics Research and Applications (IJBRA)
- International Journal of Computational Intelligence Systems (IJCIS)

Extracurricular Activities

04/2012 IBM Cloud Academy CON, NC, Research Triangle Park (RTP), North Carolina, USA.

12/2010 Workshop on Frontiers of Data Management, Soochow University, Suzhou, China.

11/2010 Workshop on Massive Data Mining and Knowledge Discovery, Southwest Jiaotong University, Chengdu, China.

11/2010 China Computer Federation Advanced Disciplines Lectures (the 11th issue) - Massive Data Mining and Knowledge Discovery, Southwest Jiaotong University, Chengdu, China.

08/2010 Intel Software College - Multi-core Programming for Academia, Fudan University, Shanghai, China.

07/2010 National Graduate Summer School - Data Intensive Computing and Unstructured Data Management, Renmin University of China, Beijing, China.

12/2009 Workshop on Massive Data Mining and Knowledge Discovery, Southwest Jiaotong University, Chengdu, China.

08/2009 National Graduate Summer School (Dragon Star Plan) - Data Mining, Zhejiang University, Hangzhou, China.

Skills

Operating System Linux-based OS (e.g., Cent OS, Ubuntu), Mac OS, Windows XP, Windows 7/8

Programming C/C++, CUDA C (for NVIDIA GPUs), Python, Java, Shell and various languages with practical experiences.

Others Hadoop/Spark/Phoenix++/Twister (MapReduce-like Softwares), GraphChi (Graph Computing System), Theano/Pylearn2 (GPU-based Deep Learning in Python), Scikit-learn (Machine Learning in Python), Weka(Data Mining Software in Java), Matlab, LATEX, MS Office (e.g., Visio, Word, Excel, PowerPoint)

Languages

Chinese (native), English (fluent)