

School of Information Science and  
Technology, Southwest Jiaotong  
University, Chengdu 610031, China  
& Department of Computer Science,  
Georgia State University, Atlanta,  
GA 30303, USA

+86-18657495668

✉ JunboZhang86@163.com,  
jbzhang@cs.gsu.edu

🌐 cs.gsu.edu/jbzhang



# Curriculum Vitae

*Junbo Zhang Ph.D candidate*

## Education

Dates	Sep. 2009 – 2013(Expected)
Degree pursuing	Ph.D.
Subject	Computer Applied Technology
Name of educational organization	Southwest Jiaotong University, China
Dates	Feb. 2012 – Feb. 2013
	Visiting PhD Student
Subject	Computer Science
Name of educational organization	Georgia State University, USA
Dates	Sep. 2005 – Jun. 2009
Degree awarded	B.Eng.
Subject	Communication Engineering
Name of educational organization	Southwest Jiaotong University, China

## Work/Intern Experience

Dates	Feb. 2012 – Feb. 2013
Position	Research Associate
Organization	Georgia State University, USA
Dates	Aug. 2011 – Sep. 2011
Position	Intern
Organization	SCK•CEN (Belgian Nuclear Research Centre), Belgium

## Research Topics

Junbo Zhang's research focus on the design, analysis, and implementation of algorithms in the fields of cloud computing, data mining, rough sets, granular computing. The research topics include large-scale knowledge discovery & feature selection (attribute reduction), performance analysis in different MapReduce runtime systems, translators

between programming languages in cloud computing, rough sets based efficient algorithms in dynamic information systems and etc.

### 1. **Large-scale Knowledge Discovery & Feature Selection (Attribute Reduction)**

- PLAR: Parallel Large-scale Attribute Reduction on Cloud Systems, submitted to **IEEE TKDE**.  
Attribute reduction for very large data is viewed as an important preprocessing step in the areas of pattern recognition, machine learning and data mining. In this paper, a novel parallel method based on MapReduce for large-scale attribute reduction is proposed. By using this method, several representative heuristic attribute reduction algorithms in rough set theory have been parallelized. Further, each of the improved parallel algorithms can select the same attribute reduct as its sequential version, therefore, owns the same classification accuracy.
- A Parallel Matrix-based Method for Computing Approximations in Incomplete Information Systems, submitted to **IEEE TKDE**.  
Large-scale data mining and knowledge discovery present a tremendous challenge with the data volume growing at an unprecedented rate. Rough set theory has been successfully applied in pattern recognition, machine learning and data mining. The lower and upper approximations are basic concepts in rough set theory. The efficient computation of approximations is vital to improve the performance of data mining or other related tasks. The recently introduced cloud computing model, MapReduce, has gained a lot of attentions from the scientific community for its applicability in large-scale data analysis. In previous research, we proposed a MapReduce based method for computing approximations in parallel, which is efficient to process complete data, but failed to process incomplete data (missing data). To overcome this shortcoming, in this paper, parallel matrix-based methods are introduced to process large-scale incomplete data and four different parallel strategies are presented.
- Parallel Approaches to Neighborhood Rough Sets: Classification and Feature Selection, submitted to **IJ-CAI2013**.  
Classification is one of the most common methods in data mining. It has been widely used in many fields. There are many growing number of techniques which have been used to help improve classification algorithms. Rough set theory is a novel soft computing tool and has been applied in not only classification algorithms but also other data mining algorithms. The rough set based neighborhood classifier has higher accuracies in some data sets than the  $K$ -NN algorithm. However, nowadays, the rapidly growing volume of data requires that the data mining algorithms have not only high accuracy but also high performance. We proposed two parallel rough neighborhood algorithms based on MapReduce, namely, parallel neighborhood classifier and parallel neighborhood feature selection, to process large-scale numerical data.

### 2. **Performance Analysis in different MapReduce Runtime Systems**

- Comparisons of Parallel Large-scale Knowledge Acquisition using Rough Set Theory on Different MapReduce Runtime Systems, submitted to **International Journal of Approximate Reasoning**  
Nowadays, with the volume of data growing at an unprecedented rate, large-scale data mining and knowledge discovery have become a new challenge. Rough set theory for knowledge acquisition has been successfully applied in data mining. The recently introduced MapReduce technique has received much attention from both scientific community and industry for its applicability in big data analysis. To mine knowledge from big data, we present parallel large-scale rough set based methods for knowledge acquisition using MapReduce in this paper. And, we implement it on several representative MapReduce runtime systems: Hadoop, Phoenix and Twister. We present performance comparisons on these runtime systems. The experimental results show that (1) The computational time is mostly minimum in Twister while employing the same cores; (2) Hadoop has the most excellent speedup in the larger data sets; (3) Phoenix has the most excellent speedup in the smaller data sets. The excellent speedup also demonstrates that the proposed parallel methods can effectively process very large data on different runtime systems.

### 3. **Translators between Programming Languages in Cloud Computing**

- M2M: A simple Matlab-to-MapReduce Translator for Cloud Computing, **Tsinghua Science and Technology**.  
MapReduce is a very popular parallel programming model for cloud computing platforms, and has become an effective method for processing massive data by using a cluster of computers. X-to-MapReduce (X is a programming language) translator is a possible solution to help traditional programmers easily deploy an application to cloud systems through translating sequential codes to MapReduce codes. Recently, some SQL-to-MapReduce translators emerge to translate SQL-like queries to MapReduce codes and have good performance in cloud systems. However, SQL-to-MapReduce translators mainly focus on SQL-like queries,

but not on numerical computation. Matlab is a high-level language and interactive environment for numerical computation, visualization, and programming, which is very popular in engineering. We proposed and developed a simple Matlab-to-MapReduce translator for cloud computing, called M2M, for basic numerical computations.

- H2T: A simple Hadoop-to-Twister translator for Cloud Computing, Submitted to **ISBAST13**.

There are several platforms for users to develop their applications based on MapReduce framework such as Hadoop and Twister. Hadoop is one of the most popular runtimes for MapReduce applications and supported by various organizations, however, the original design for Hadoop did not propose an iterative feature efficiently which is required for many scientific applications. Twister, another system for Iterative MapReduce is introduced by Indiana University, designed to facilitate iterative applications based on MapReduce framework. It has shown that Twister has the better performance than Hadoop on many iterative applications. Automatic translations between two program languages in cloud platforms can help developers move their applications from one cloud to another cloud without changing code. We proposed a simple Hadoop-to-Twister translator named H2T which is designed for converting simple Hadoop applications into Twister applications.

#### 4. Rough Sets based Efficient Algorithms in Dynamic Information Systems

- Composite rough sets for dynamic data mining, submitted to **Information Sciences**

As a soft computing tool, rough set theory has become a popular mathematical framework for pattern recognition, data mining and knowledge discovery. It can only deal with one kind of data in the information system by using a specific binary relation. However, there may be multiple kinds of data in information systems in real-life applications, e.g., categorical data, numerical data, set-valued data, interval-valued data and missing data. Such information systems are called as composite information systems in this paper. A composite relation is proposed to process multiple kinds of data simultaneously in composite information systems. Then, an extended rough set model, called as composite rough sets, is presented. Moreover, combined with the incremental learning technique, a novel matrix-based method for fast updating approximations is proposed in dynamic composite information systems.

#### 5. Others

- Heterogeneous Parallel Programming: A CUDA Implementation of Boolean Matrix-based Algorithms in Composite Rough Sets
- Large-scale feature selection in SNPs data on Cloud Systems

## Papers

### Journal

- [12] **Junbo Zhang**, Tianrui Li, Yi Pan, PLAR: Parallel Large-scale Attribute Reduction on Cloud Systems, Submitted to **IEEE Transactions on Knowledge and Data Engineering**.
- [11] **Junbo Zhang**, Jian-Syuan Wong, Yi Pan, Tianrui Li, A parallel matrix-based method for computing approximations in incomplete information systems, Submitted to **IEEE Transactions on Knowledge and Data Engineering**.
- [10] **Junbo Zhang**, Tianrui Li, Hongmei Chen, Composite rough sets for dynamic data mining, Submitted to **Information Sciences**.
- [9] **Junbo Zhang**, Jian-Syuan Wong, Tianrui Li, Yi Pan, Comparisons of Parallel Large-scale Knowledge Acquisition using Rough Set Theory on Different MapReduce Runtime Systems, Submitted to **International Journal of Approximate Reasoning**.
- [8] **Junbo Zhang**, Tianrui Li, Yi Pan, Chuan Luo, Hongmei Chen, Research on Parallel Incremental Algorithm for Updating Knowledge Based on Rough Sets in Cloud Computing Environments, Submitted to **Chinese Journal of Computers** (in Chinese).
- [7] **Junbo Zhang**, Tianrui Li, Da Ruan, Zizhe Gao, Chengbing Zhao, A Parallel Method for Computing Rough Set Approximations, **Information Sciences**, vol. 194C, pp. 209-223, 2012. (SCI, EI)
- [6] **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. (SCI, EI)

- [5] **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. (SCI, EI)
- [4] **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
- [3] 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.
- [2] 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, pp. 325-344, 2011. DOI: 10.1007/s10898-010-9607-8. (SCI, EI)
- [1] Zhonggang Wang, Tianrui Li, **Junbo Zhang**, Chengbing Zhao, Zizhe Gao, Parallel Method for Preprocessing High-Speed Railway Noise Data based on MapReduce, *Tongxin Xuebao/Journal on Communications*, vol. 32, no. 9A, pp. 263-269, 2011. (in Chinese) (EI)

### Conference

- [19] **Junbo Zhang**, Chizheng Wang, Yi Pan, Tianrui Li, Parallel Approaches to Neighborhood Rough Sets: Classification and Feature Selection, Submitted to *IJCAI13*.
- [18] **Junbo Zhang**, Jian-Syuan Wong, Yi Pan, Tianrui Li, H2T: A simple Hadoop-to-Twister translator for Cloud Computing, Submitted to *ISBAST13*.
- [17] **Junbo Zhang**, Tianrui Li, Hongmei Chen, Composite Rough Sets. *AICI2012*, Chengdu, China, pp. 150-159, 2012. (EI, ISTP)
- [16] **Junbo Zhang**, Tianrui Li, Yi Pan, Parallel Rough Set based Knowledge Acquisition using MapReduce from Big Data. *ACM SIGKDD12 Big Data Mining (BigMine'12) Workshop*, Beijing, China, pp. 20-27, 2012. (EI)
- [15] **Junbo Zhang**, Tianrui Li, Da Ruan, Rough sets based incremental rule acquisition in set-valued information systems. *Autonomous Systems: Developments and Trends*, (edited by H. Unger, K. Kyamaky, and J. Kacprzyk), Springer Berlin / Heidelberg, vol. 391, pp. 135-146, 2012.
- [14] **Junbo Zhang**, Tianrui Li, Yan Yang, Wei Wang, Neighborhood Rough Sets based Matrix Approach for Calculation of the Approximations. In: Proceedings of the 6th International Conference on Rough Sets and Knowledge Technology (*RSKT2011*), Canada, pp. 166-171, 2011. (EI)
- [13] **Junbo Zhang**, Tianrui Li, Da Ruan, Dun Liu, Neighborhood Rough Sets for Dynamic Data Mining. In: Proceedings of 2011 World Conference on Soft Computing (*WConSC2011*), San Francisco, USA, 2011.
- [12] **Junbo Zhang**, Tianrui Li, Lei Wang, A New Method for Calculation of the Approximations under the Probabilistic Rough Sets. In: Proceedings of 2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering (*ISKE2010*), Hangzhou, China, pp. 73-76, 2010. (EI, ISTP)
- [11] **Junbo Zhang**, Tianrui Li, Dun Liu, An Approach for Incremental Updating Approximations in Variable Precision Rough Sets while Attribute Generalized. In: Proceedings of 2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering (*ISKE2010*), Hangzhou, China, pp. 77-81, 2010. (EI, ISTP)
- [10] **Junbo Zhang**, Tianrui Li, Muhammad Zia ur Rehman, Shaoyong Li, An Approach for Selection of the Proper Lever of Granularity in Granular Computing. In: Proceedings of the 9th International Conference on Foundations and Applications of Computational Intelligence (*FLINS2010*), Chengdu, China, pp. 764-769, 2010. (ISTP)

- [9] Hongmei Chen, Tianrui Li, **Junbo Zhang** and Luo Chuan, Probabilistic Composite Rough Set and Attribute Reduction, *ISKE2012*, Beijing, China.
- [8] Anping Zeng, Tianrui Li, **Junbo Zhang**, Dun Liu: An Incremental Approach for Updating Approximations of Rough Fuzzy Sets under the Variation of the Object Set. *RSCTC2012*, pp. 36-45, 2012
- [7] Zizhe Gao, Tianrui Li, **Junbo Zhang**, Chengbing Zhao, Zhonggang Wang, A Parallel Method for Unpacking Original High Speed Rail Data Based on MapReduce. Practical Applications of Intelligent Systems, (edited by Y. Wang and T. Li), Springer Berlin / Heidelberg, vol. 124, pp. 59-68, 2011.
- [6] Dun Liu, **Junbo Zhang**, Tianrui Li, A Probabilistic Rough Set Approach for Incremental Learning Knowledge on the Change of Attributes. In: Proceedings of the 9th International Conference on Foundations and Applications of Computational Intelligence (*FLINS2010*), Chengdu, China, pp. 722-727, 2010. (**ISTP**)
- [5] Hongmei Chen, Tianrui Li, **Junbo Zhang**, A Method for Incremental Updating Approximations when Objects and Attributes Vary with Time. In: Proceedings of 2010 IEEE International Conference on Granular Computing, California (*GrC2010*), USA, pp. 90-95, 2010. (**EI, ISTP**)
- [4] Hongmei Chen, Tianrui Li, **Junbo Zhang**, A Method for Incremental Updating Approximations Based on Variable Precision Set-Valued Ordered Information Systems. In: Proceedings of 2010 IEEE International Conference on Granular Computing (*GrC2010*), California, USA, pp. 96-101, 2010. (**EI, ISTP**)
- [3] Shaoyong Li, Tianrui Li, **Junbo Zhang**, Approach for Dynamically Updating Set Approximations based Rough Sets while Condition Attributes Value Coarsening and Refining. In: Proceedings of 2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering (*ISKE2010*), Hangzhou, China, pp. 27-32, 2010. (**EI**)
- [2] Shaoyong Li, Tianrui Li, **Junbo Zhang**, Updating Approximations Dynamically in Dominance-based Rough Sets. In: Proceedings of the 2010 International Conference on Electronic-Business Intelligence (*ICEBI2010*), Kunming, China, pp. 534-540, 2010.
- [1] Dun Liu, Tianrui Li, **Junbo Zhang**. An Incremental Approach for Rule Induction under Coarsening and Refining of Attribute Values in E-Business Systems. In: Proceedings of the 2010 International Conference on Electronic-Business Intelligence (*ICEBI2010*), Kunming, China, pp. 541-547, 2010.

## Project Experience

- |                         |  |
|-------------------------|--|
| Nov. 2012–<br>Present   | <p><b>Research on dynamic knowledge discovery techniques and efficient algorithms under granular computing</b>, the <i>Fostering Foundation for the Excellent Ph.D. Dissertation of Southwest Jiaotong University, China</i>.</p> <p>I am leader and principal investigator in this project, which is the follow-up project of the Doctoral Innovation Funding Project of Southwest Jiaotong University, and supports my Ph.D. dissertation.</p> |
| Otc. 2012–<br>Present   | <p><b>Dynamic knowledge discovery system based on rough sets in cloud computing environments</b>, the <i>Science and Technology Planning Project of Sichuan Province, China</i>.</p> <p>I am leader and principal investigator in this project, which is the follow-up project of the Young Software Innovation Foundation of Sichuan Province, and is one part of my Ph.D. dissertation.</p>  |
| Otc. 2011–<br>Sep. 2012 | <p><b>Research on dynamic knowledge discovery techniques and efficient algorithms under granular computing and probabilistic rough sets</b>, the <i>Doctoral Innovation Funding Project of Southwest Jiaotong University</i>.</p> <p>I was leader and principal investigator in this project, which was excellently completed and helpful to apply for the follow-up project.</p>  |
| Otc. 2011–<br>Sep. 2012 | <p><b>Research on dynamic knowledge discovery system based on cloud computing and rough sets</b>, the <i>Young Software Innovation Foundation of Sichuan Province</i>.</p> <p>I was leader and principal investigator in this project, which was excellently completed and helpful to apply for the follow-up project.</p>   |

- May. 2012–  
Jul. 2012 **8-week NSF REU Undergraduate Summer Research Program hosted by the Department of Computer Science, Georgia State University, the National Science Foundation.**  
I was participated as a mentor, and provided guidance about high performance data mining to the students from University of Illinois at Urbana-Champaign, University of Maryland, Georgia State University, Morehouse College, North Carolina Agricultural and Technical State University, Southern Polytechnic State University, University of Tennessee, Virginia Commonwealth University and et al.
- Jan. 2012–  
Present **Research on dynamic updating knowledge theories and algorithms based on granular computing, the National Natural Science Foundation of China.**  
My main research work was the design, analysis and implementation of algorithms.
- Jan. 2012–  
Present **Research on composite rough set models and algorithms of knowledge discovery, the National Natural Science Foundation of China.**  
My main research work was the design, analysis and implementation of algorithms.
- Jan. 2009–  
Dec. 2011 **Research on incremental learning theories and methods based on granular computing, the National Natural Science Foundation of China.**  
My main research work was the design, analysis and implementation of algorithms. I have developed the software of dynamic knowledge discovery based rough sets.

## Honors and Awards

- 2012 "Si Shi Yang Hua" Medal in Southwest Jiaotong University (Less than 1/1000, the students' top honor of Southwest Jiaotong University)
- 2012 First Class Prize in the 9<sup>th</sup> "Huawei" National Postgraduate Mathematical contest in Modeling
- 2010 Second Class Prize in the 2<sup>nd</sup> "Huawei" Innovation Programming Contest
- 2009–2013 Top Class Scholarship of Southwest Jiaotong University
- 2011 Second Class Prize in the 8<sup>th</sup> National Postgraduate Mathematical contest in Modeling
- 2009 Second Class Prize in the 6<sup>th</sup> National Postgraduate Mathematical contest in Modeling
- 2008 Second Class Prize in the 2<sup>nd</sup> Sichuan Provincial Programming Contest
- 2008 Honorable Mention in the 33<sup>th</sup> ACM Asia Programming Contest
- 2007 Second Class Prize in the 1<sup>st</sup> Sichuan Provincial Programming Contest
- 2007 Second Class Prize (the 12<sup>th</sup> Place in person) in the TopCoder Sichuan Provincial Contest
- 2007 Third Class Prize in National Electrical Engineering Association Mathematical contest in Modeling
- 2006,2007 School of Information Science and Technology President Scholarship of Southwest Jiaotong University
- 2005-2009 Undergraduate Comprehensive Scholarship of Southwest Jiaotong University

## Skills

- Languages Chinese(Native), English
- Operating Systems Linux-based, Windows
- Programming Languages C/C++, Java, Matlab, Python, CUDA C
- MapReduce Runtime Systems Apache Hadoop, Phoenix (by Stanford Univ.), Twister & Twister4Azure (by Indiana Univ.), Mars (by HKUST), Granules (by Colorado State Univ. & Indiana Univ.)
- Software LaTeX, Git, Weka, MS Office (Word, Excel, PowerPoint, Visio)

---

## Professional Activities

- 2012 Reviewer of International Journal of Cloud Computing (IJCC)
- 2012 Reviewer of International Journal of Bioinformatics Research and Applications (IJBRA)
- 2011 Reviewer of World Wide Web Journal (WWWJ)
- 2010,2011 Reviewer of International Journal of Computational Intelligence Systems (IJCIS)

---

## Extracurricular Activities

- Apr. 2012 IBM Cloud Academy CON, NC, Research Triangle Park (RTP), North Carolina, USA
- Dec. 2010 2010 Workshop on Frontiers of Data Management, Suzhou, China
- Nov. 2010 2010 Workshop on Massive Data Mining and Knowledge Discovery, Chengdu, China
- Nov. 2010 2010 China Computer Federation Advanced Disciplines Lectures (the 11<sup>th</sup> issue) – Massive Data Mining and Knowledge Discovery, Chengdu, China
- Jul. 2010 2010 National Graduate Summer School – Data Intensive Computing and Unstructured Data Management, Beijing, China
- Dec. 2009 2009 Workshop on Massive Data Mining and Knowledge Discovery, Chengdu, China