Jianpeng Xu[1]

Email: jianpeng.xu@gmail.com/xujianpe@msu.edu

EDUCATION

Ph.d Candidate, Computer Science and Engineering, Michigan State University (MSU)
Fall 2011 - Summer 2016 (expected) **Advisor**: Pang-Ning Tan
M.S Degree, Computer Science, Harbin Institute of Technology (HIT), China
Fall 2007 - Spring 2010
B.S Degree, Computer Science, Shandong University (SDU), China
Fall 2003 - Spring 2007

EXPERIENCE

Intern Data Scientist — Samsung Research America, San Jose, CA — Spring, Summer 2015

Design recommender system algorithms for Samsung Smart TV. Build TV Ads system, especially for the large-scale data streaming and database support. (Techniques include Hadoop, Spark, Pig, Java, Scala, Python, bash, etc.)

Research Assistant — MSU, East Lansing, MI — Fall 2013-Present

Multi-task Learning/Ensemble forecasting/Downscaling/Personalized medical modeling. (Techniques include matrix factorization, regression, online learning, etc.)

Intern Data Scientist — Samsung Research America, San Jose, CA — Summer 2014

Analysis on user behavior and user demographics using smart TV watching history.

Teaching Assistant — MSU, East Lansing, MI — Spring 2013

Teaching Assistant for the course Computational Techniques for Large-Scale Data Analysis (CSE491/891), covering techniques including Hadoop, Pig, Hive, Mahout, as well as AWS.

Intern Software Engineer — Narus Inc, Sunnyvale, CA — Summers 2012

Develop methods to detect malicious endpoints using data mining techniques, including label propagation.

Research Assistant — MSU, East Lansing, MI — Fall 2011-Fall 2012

Land cover and land use change detection using MODIS data.

PUBLICATIONS

- **Jianpeng Xu**, Kaixiang Lin, Pang-Ning Tan and Jiayu Zhou. Synergies that Matter: Efficient Interaction Selection via Sparse Factorization Machine. To appear in SDM, 2016
- **Jianpeng Xu**, Pang-Ning Tan, Lifeng Luo and Jiayu Zhou. GSpartan: a Geospatio-Temporal Multi-task Learning Framework for Multi-location Prediction. To appear in SDM, 2016
- **Jianpeng Xu**, Jiayu Zhou and Pang-Ning Tan. Formula: FactORized MUlti-task LeArning for task discovery in personalized medical models. SDM 2015, pages 496-504
- **Jianpeng Xu**, Pang-Ning Tan and Lifeng Luo. ORION: Online Regularized multI-task regressiON and its application to ensemble forecasting. ICDM 2014, pages 1061-1066, Shenzhen, China, December 14-17, 2014

[1] CV udpated on February 25, 2016.

- **Jianpeng Xu** and Shufan Ji. HDminer: Efficient Mining of High Dimensional Frequent Closed Patterns from Dense Data and Its Application. ICDM workshop (ICDMW 2014) on Scalable Data Analytics: Theory and Applications, pages 1061-1067, Shenzhen, China, December 14-17, 2014
- Lei Liu, Sabyasachi Saha, Ruben Torres, **Jianpeng Xu**, Pang-Ning Tan, Antonio Nucci, Marco Mellia, Detecting Malicious Clients in ISP Networks Using HTTP Connectivity Graph and Flow Information, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM2014), August 17 August 20, 2014, Beijing, China.

PATENT

- Sabyasachi Saha, Lei Liu, Ruben Torres, **Jianpeng Xu**, Antonio Nucci, Detecting Malicious Endpoints Using Network Connectivity and Flow Information, US8813236 B1, Aug 19, 2014

HONORS AND AWARDS

Student Travel Award, SDM 2016
Student Travel Award, SDM 2015
Student Travel Award, ICDM 2014
Student Travel Award, MSU Graduate School, 2014
Graduate Assistant Support from Michigan State University, Fall 2011-Present
Scholarships for outstanding students, HIT, Fall 2007-Spring 2010
Scholarships for outstanding students, SDU, Fall 2004, 2006
Honorable mentioned in Mathematical Contest in Modeling (MCM), 2006

SERVICES

- **PC Member** for MLRec2016, BIBM2015, IJCAI2015, MLRec2015(in conjunction with SDM2015)
- **Conference External Reviewer** for IJCAI2016, AAAI2016, SDM2016, SDM2015, WSDM2015, DSAA2015, CIKM2014, DSAA2014, ICDM2014
- **Invited Journal Reviewer** for Neurocomputing, BMC Bioinformatics, TNNLS

MACHINE LEARNING SKILLS

Liner regression, Logistic regression, SVM, PCA, SVD, Matrix factorization, Tensor decomposition, Online learning, (stochastic) Gradient descent, Probabilistic modeling, Gaussian process, Clustering, etc.

PROGRAMMING SKILLS

Software Developments: C/C++, Java, C#, Python, Matlab, SQL
Web Developments: Perl, PHP, AJAX, Javascript, HTML, XML
Big Data Analysis: Hadoop, Hive, Pig, Mahout, Spark

REFERENCESAREAVAILABLE UPON REQUEST