JIAYU ZHOU

email Jiayu.Zhou@asu.edu
phone +1(480)334-5283
website jiayuzhou.github.io/

EDUCATION

2009 - 2014 — *Ph.D. in Computer Science*, Arizona State University, USA
Advisor: Prof. Jieping Ye

2004 - 2008 — *B.Eng. in Computer Science*, Beijing Jiaotong University, China

RESEARCH INTERESTS

Machine Learning — Multi-task learning, matrix completion, convex optimization, non-convex optimization and metric learning.

Bioinformatics — Disease progression modeling, biomarker identification, predictive modeling.

Healthcare Analytics — Data processing, feature extraction, feature construction and predictive modeling from electronic medical records (EMR).

HONORS & AWARDS

2013 — Invention 1st File Patent Award, IBM Research

2013 — Student Travel Award, The 13th SIAM International Conference on Data Mining (SDM 2013)

2012 — Best Video Award, The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD 2012)

2012 — Best Poster Award, The 12th SIAM International Conference on Data Mining (SDM 2012)

2012 — Student Travel Award, The 12th SIAM International Conference on Data Mining (SDM 2012)

2011 — Student Travel Award, The 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD 2011)

2011 — Student Travel Award, The 11th SIAM International Conference on Data Mining (SDM 2011)

2009 — University Graduate Fellowship, Arizona State University

2007 — National Scholarship, Ministry of Education, P.R.China

RESEARCH EXPERIENCES

Jan 14 - Present — Senior Research Scientist, Samsung Research America, CA.
· Design algorithms and sytems for recommending TV programs using large-scale TV watch log data.
· Develop models for targeted advertising.

June 10 - Jan 14 — Research Assistant, Arizona State University, AZ.
Supervisor: Prof. Jieping Ye
· Built general disease progression models using multi-task learning techniques. The disease progression models were successfully applied to the prediction of Alzheimer's disease and cancer progression.
· Designed formulations and algorithms for low-rank modeling.
· Developed efficient optimization algorithms for solving various multi-task learning formulations.

May 13 - Aug 13 — Research Intern, IBM Thomas J. Watson Research Center, NY
Mentor: Dr. Fei Wang
· Designed machine learning models for effective and efficient electronic medical record (EMR) imputation for large scale healthcare problems, leveraging the temporal nature of the patients and the shared information among the patients. The models significantly improved the predictive performance of healthcare systems.

May 12 - Aug 12 — Research Intern, IBM Thomas J. Watson Research Center, NY

Mentor: Dr. Jimeng Sun
· Designed and implemented a sparse predictive modeling pipeline for large-scale data analysis in PYTHON that takes advantage of the sparsity in the data matrices and has greatly improved the scalability of predictive modeling.
· Designed experiments to build effective hospitalization admission models and identify important clinical features for congested heart failure patients from their health records.

Aug 09 - Dec 10 — Research Assistant, Arizona State University, AZ
Developed a semantic translation engine in JAVA that translates English to first order logic.

Aug 08 - Feb 09 — Software Engineer, Moka LLC, Beijing
Developed an English-Chinese machine translation engine in JAVA for online translation services.

Dec 07 - Feb 08 — Research Intern, Microsoft Research Asia, Beijing
Mentor: Dr. Jingfang Xu
Developed a system in C# for analyzing large-scale log data of Microsoft Live search engine and modeling user behavior patterns to improve search experience.

Oct 06 - Jan 08 — Research Intern, Chinese Academy of Sciences, Beijing
Designed and developed a system in JAVA for concept verification and hyponymy retrieval from large-scale corpus.

JOURNAL PUBLICATIONS

NeuroImage — Rashmi Dubey, **Jiayu Zhou**, Yalin Wang, Paul M. Thompson, and Jieping Ye, "Analysis of Sampling Techniques for Imbalanced Data: An N=648 ADNI Study", *NeuroImage* (5-Year Impact Factor: 7.063), Vol 87, 2014

NeuroImage — **Jiayu Zhou**, Jun Liu, Vaibhav A. Narayan and Jieping Ye, "Modeling Disease Progression via Multi-task Learning ", *NeuroImage* (5-Year Impact Factor: 7.063), Vol 78, 2013

IJCINI — **Jiayu Zhou**, Shi Wang and Cungen Cao, "Learning Hierarchical Lexical Hyponymy", *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 2010

REFEREED CONFERENCE PUBLICATIONS

KDD' 14 — **Jiayu Zhou**, Fei Wang, Jianying Hu and Jieping Ye, "From Micro to Macro: Data Driven Phenotyping by Densification of Longitudinal Electronic Medical Records", *The 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining* (KDD)

KDD' 14 — Pinghua Gong, **Jiayu Zhou** and Jieping Ye, "Efficient Multi-Task Feature Learning with Calibration", *The 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining* (KDD)

ICPR' 14 — Fei Wang, **Jiayu Zhou** and Jianying Hu, "Density Transfer: A Data Driven Approach for Imputing Electronic Health Records", *The 22nd Int'l Conf. on Pattern Recognition* (ICPR)

SPIE' 14 — Sinchai Tsao, **Jiayu Zhou**, Jie Shi, Jieping Ye, Yalin Wang, Natasha Lepore, "Evaluating the predictive power of multivariate tensor-based morphometry in Alzheimer's disease progression via convex fused sparse group Lasso", *SPIE Medical Imaging*

PSB' 13 — Yashu Liu, Zhi Nie, **Jiayu Zhou** and Jieping Ye, "Sparse Generalized Functional Model for Predicting Remission Status of Depression Patients", *Pacific Symposium on Biocomputing* (PSB)

NIPS' 13 — Jie Wang, **Jiayu Zhou**, Peter Wonka and Jieping Ye, "Lasso Screening Rules via Dual Polytope Projection", *The 27th Annual Conf. on Neural Information Processing Systems* (NIPS) Spotlight

ICDM' 13 — Shayok Chakraborty, **Jiayu Zhou**, Vineeth Balasubramanian, Sethuraman Panchanathan, Ian Davidson, and Jieping Ye, "Active Matrix Completion" *IEEE International Conference on Data Mining* (ICDM)

KDD' 13 **Jiayu Zhou**, Zhaosong Lu, Jimeng Sun, Lei Yuan, Fei Wang and Jieping Ye, "FeaFiner: Biomarker Identification from Medical Data through Feature Generalization and Selection", *The 19th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining* (KDD)

SDM' 13 **Jiayu Zhou**, Jimeng Sun, Yashu Liu, Jianying Hu and Jieping Ye, "Patient Risk Prediction Model via Top-*k* Stability Selection", *The 13th SIAM Int'l Conf. on Data Mining* (SDM)

DMMH' 13 **Jiayu Zhou**, Jimeng Sun, Fei Wang, Jianying Hu, Shahram Ebadollahi and Jieping Ye, "Discover Temporal Dynamics of Biomarkers in Predictive Modeling with Longitudinal Data", *2nd Workshop on Data Mining for Medicine and Healthcare, held in conjunction with the 13th SIAM Int'l Conf. on Data Mining* (SDM-DMMH)

KDD' 12 **Jiayu Zhou**, Jun Liu, Vaibhav A. Narayan and Jieping Ye, "Modeling Disease Progression via Fused Sparse Group Lasso", *The 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining* (KDD) Best Video Award

NIPS' 11 **Jiayu Zhou**, Jianhui Chen and Jieping Ye, "Clustered Multi-Task Learning via Alternating Structure Optimization", *The 25th Annual Conf. on Neural Information Processing Systems* (NIPS)

KDD' 11 Jianhui Chen, **Jiayu Zhou** and Jieping Ye, "Integrating Low-Rank and Group-Sparse Structures for Robust Multi-Task Learning", *The 17th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining* (KDD)

KDD' 11 **Jiayu Zhou**, Lei Yuan, Jun Liu and Jieping Ye, "A Multi-Task Learning Formulation for Predicting Disease Progression", *The 17th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining* (KDD)

IWCS' 11 Chitta Baral, Juraj Dzifcak, Marcos Alvarez Gonzalez and **Jiayu Zhou**, "Using Inverse Lambda and Generalization to Translate English to Formal Language", *The 9th Int'l Conf. on Computational Semantics* (IWCS)

AAAI' 08 **Jiayu Zhou**, Youfang Lin and Xi Wang "Visualization of Large-Scale Weighted Clustered Graph: An Evolutionary Approach", *The 23rd National Conference on Artificial Intelligence* (AAAI)

ICCI' 08 **Jiayu Zhou**, Youfang Lin , Shi Wang and Cungen Cao "Discover Hierarchical Lexical Hyponymy Relation from Large-Scale Concept Set", *The 7th IEEE International Conference on Cognitive Informatics* (ICCI)

KSEM' 07 **Jiayu Zhou**, Shi Wang and Cungen Cao "A Statistical Acquisition Model of Chinese Lexical Concept based on Google", *The 2nd International Conference on Knowledge Engineering and Management* (KSEM)

BOOK CHAPTERS

2010 **Jiayu Zhou** and Shi Wang "Concept Mining and Inner Relationship Discovery from Text", *New Advances in Machine Learning*, 2010, ISBN 978-953-307-034-6

PATENTS

2013 (with Fei Wang and Jianying Hu) "A Method for Densification of Longitudinal EMR for EMR Driven Phenotyping", 2013, *IBM Rate-1*, Submitted

2013 (with Fei Wang, Hanghang Tong and Yinglong Xia) "System and Method for Patient Disease Status Prediction", 2013, *IBM*, Submitted

2012 (with Jimeng Sun, Fei Wang, Jianying Hu and Shahram Ebadollahi) "A Method for Biomarker Identification Leveraging Temporal Dynamics", 2012, Submitted

SELECTED RESEARCH PROJECTS

2010 - present Large Scale Multi-Task Learning

Machine Learning Methodology

Many real world machine learning problems involve the learning of related tasks, which commonly share certain knowledge. The multi-task learning (MTL) aims at simultaneously learn the related tasks and improve the generalization performance by leveraging the shared knowledge. In this project, my goal is to design novel MTL formulations and develop efficient optimization algorithms for large scale learning problems. Besides, I have

developed an open source machine learning package called Multi-tAsk Learning via StructurAl Regularization (MALSAR), which is the largest MTL software and widely used in the community. The package is available at www.MALSAR.org. Related Papers [NIPS' 11, KDD' 11, KDD' 12, KDD' 13]

2010 - present **Modeling Disease Progression**

Machine Learning Applications

The development of many progressive diseases such as Alzheimer's involves complex pathological manifestations. Longitudinal studies on such diseases have provided important information about the progression of the disease and also many candidate biomarkers from medical images to plasma panels. In this project I have develop novel machine learning models to build effective predictive models for such diseases, leveraging the temporal information in the longitudinal data. The models are also capable of identifying important biomarkers that signal the progression and their temporal dynamics. Related Papers [NeuroImage, KDD' 12, KDD' 11, DMMH' 13]

TALKS AND POSTER PRESENTATIONS

IBM' 13 "Densification of Longitudinal Electronic Medical Records: A Data Driven Approach", IBM Research, KDD PIC Presentation, Yorktown Heights, NY, USA. Aug 2013

KDD' 13 "FeaFiner: Biomarker Identification from Medical Data through Feature Generalization and Selection", The 19th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD), Chicago, IL, Aug 2013

SDM' 13 "Patient Risk Prediction Model via Top-k Stability Selection", SIAM Int'l Conf. on Data Mining (SDM), Austin, TX, May 2013

DMMH' 13 "Discover Temporal Dynamics of Biomarkers in Predictive Modeling with Longitudinal Data", 2nd Workshop on Data Mining for Medicine and Healthcare (DMMH), SIAM Int'l Conf. on Data Mining (SDM), Austin, TX, May 2013

SDM' 13 (Poster)"MALSAR: Multi-Task Learning via Structural Regularization", Doctoral Forum, SIAM Int'l Conf. on Data Mining (SDM), Austin, TX, May 2013

IBM' 12 "Think Sparsely: Predictive Modeling System on Sparse Medical Data", IBM Research, Healthcare and Life Sciences Webminar Series, Hawthorne, NY, USA. Aug 2012

SDM' 12 Tutorial (with Prof. Jieping Ye), "Multi-Task Learning: Theory, Algorithms, and Applications", SIAM Int'l Conf. on Data Mining (SDM), Anaheim, CA, April 2012

SDM' 12 (Poster) "Modeling Disease Progression via Multi-Task Learning Formulations", Doctoral Forum, SIAM Int'l Conf. on Data Mining (SDM), Anaheim, CA, April 2012. Best Poster Award

NIPS' 11 "Clustered Multi-Task Learning via Alternating Structure Optimization", Advances in Neural Information Processing Systems, Granada, Spain, Dec 2011

KDD' 11 (Poster) "Integrating Low-Rank and Group-Sparse Structures for Robust Multi-Task Learning", KDD, San Diego, CA, August 2011

KDD' 11 (Poster) "Modeling Disease Progression: A Multi-Task Learning Formulation", KDD, San Diego, CA, August 2011

SDM' 11 (Poster) "Modeling Disease Progression: A Multi-Task Learning Formulation", Doctoral Forum, SIAM Int'l Conf. on Data Mining (SDM), Mesa, AZ, April 2011

ICCI' 08 "Learning Hierarchical Lexical Hyponymy", ICCI, Stanford, CA, Aug 2008

KSEM' 07 "A Statistical Acquisition Model of Chinese Lexical Concept based on Google", KSEM, Melbourne, Victoria Australia, Nov 2007

TEACHING AND TUTORIAL EXPERIENCE

2012 Instructor, "Multi-Task Learning: Theory, Algorithms, and Applications", Tutorial at SIAM Int'l Conf. on Data Mining (SDM)

Fall 2012 Guest Instructor, two lectures on multi-task learning to the class Numerical

	Linear Algebra at Arizona State University
Fall 2011	Volunteer Speaker, one lecture on Linear Regression, Machine Learning Seminar, Arizona State University
Spring 2011	Volunteer Speaker, four lectures on Convex Functions, Machine Learning Seminar, Arizona State University
Fall 2010	Volunteer Speaker, two lectures on Predicting Cognitive Scores from MRI, Machine Learning Seminar, Arizona State University
Fall 2010	Teaching Assistant, CSE205: Java Programming Language, Arizona State University
Summer 2010	Volunteer Speaker, two lectures on Posterior Regularization Framework, Machine Learning Seminar, Arizona State University
Spring 2010	Teaching Assistant, CSE230: Assembly Language and Computer Architecture, Arizona State University
Spring 2010	Guest Instructor, three lectures on Log-linear Model and Statistical Machine Translation to the class CSE576: Natural Language Processing at Arizona State University
Fall 2009	Teaching Assistant, CSE205: Java Programming Language, Arizona State University

PROFESSIONAL ACTIVITIES

2005 - Present	Graduate Student Member, IEEE
2010 - Present	Graduate Student Member, SIAM
2011 - 2012	Grant Reviewer, GPSA, ASU
PC Member	· 2014 Neural Information Processing Systems (NIPS) · 2014 Intl. Conf. on Data Mining (ICDM) · 2014 ICDM Workshop on Scalable Data Analytics: Theory and Applications (SDA-ICDM) · 2014 Intl. Workshop on Data Mining for Brain Science (BrainKDD) · 2014 Intl. Conf. on Data Science and Advanced Analytics (DSAA) · 2014 IEEE/CIC ICCC Symposium on Social Networks and Big Data (SNBD) · 2014 6th Asian Conference on Machine Learning (ACML) · 2014 The 3rd ASE Intl. Conference on Big Data Science and Computing (BigDataScience), Program Vice Chair · 2014 IEEE Intl. Conference on Big Data (IEEE BigData) · 2014 WSDM Workshop on Diffusion Networks and Cascade Analytics · 2014 PAKDD Workshop on Scalable Data Analytics · 2014 The 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) · 2014 SDM Workshop on Data Mining for Medicine and Healthcare · 2014 SDM Workshop on Heterogeneous Learning · 2014 SIAM Intl. Conference on Data Mining (SDM) · 2013 ICDM Workshop Mining and Understanding from Big Data · 2013 5th Asian Conference on Machine Learning (ACML)
Reviewer	· ACM Transactions on Knowledge Discovery from Data (TKDD) · IEEE Transactions on Knowledge and Data Engineering (TKDE) · IEEE Transactions on Pattern Recognition and Machine Intelligence (TPAMI) · Data Mining and Knowledge Discovery (DMKD) · Pattern Recognition Letters (PRLETTERS) · Knowledge and Information Systems (KAIS) · Computational Statistics and Data Analysis (CSDA) · BMC Bioinformatics · Neurocomputation · IBM Journal of Research and Development · Intl. Journal on Artificial Intelligence Tools (IJAIT)

PROGRAMMING SKILLS

Languages	Skilled in MATLAB, SCALA, JAVA (Sun Certified Java Programmer since 2004), C#, PYTHON and BASH. Familiar with C/C++
OS	Familiar with LINUX and WINDOWS programming
Analytics	Skilled in predictive modeling, collaborative filtering, large scale optimization, HADOOP, MAPREDUCE and SPARK
Algorithms	Skilled in data structures and algorithms.

July 2, 2014