

Jinghe Zhang

CONTACT INFORMATION	Olsson Hall 227C, 151 Engineer's Way, Charlottesville, VA 22903 Homepages: https://www.linkedin.com/in/jinghezhang https://github.com/JingheZ Phone: 607-651-8906 Email: jz4kg@virginia.edu
RESEARCH INTERESTS	Machine Learning , e.g., representation learning, metric learning; Machine Learning Applications , e.g., data mining, predictive modeling, recommender systems, health and medical informatics
EDUCATION	University of Virginia <i>08/2013 - Present</i> Ph.D., Systems & Information Engineering GPA: 3.91/4.0 Binghamton University, State University of New York <i>08/2011 - 05/2013</i> Master of Science, Industrial & Systems Engineering GPA: 3.76/4.0 Hebei University of Technology <i>09/2007 - 06/2011</i> Bachelor of Science, Industrial Engineering GPA: 3.65/4.0 <i>Charlottesville, VA</i> <i>Binghamton, NY</i> <i>Tianjin, China</i>
ACADEMIC POSITIONS	Graduate Research Assistant <i>08/2013 - Present</i> Department of Systems & Information Engineering, University of Virginia Graduate Teaching Assistant <i>01/2014 - 05/2014</i> SYS6016 Machine Learning, University of Virginia Graduate Research Associate <i>08/2011 - 05/2013</i> Watson Institute for Systems Excellence, Binghamton University Continuous Process Improvement in Healthcare, Quality Management Department, United Health Services (UHS)
PROJECT EXPERIENCE	High-risk Patients Retrieval using Electronic Health Data <i>09/2014 - Present</i> <ul style="list-style-type: none">Performed representation learning on patients' medical histories to construct feature vectors with reduced information loss than using common methodsConducted metric learning on the representation of patients' medical data and discovered informative features for clinical useTrained well-established classifiers, such as support vector machine, linear discriminant analysis, random forest, k nearest neighbors, etc., to discriminate patients with and without anxiety/depressionEvaluated the proposed framework by the early detection of anxiety/depression on new patients and achieved a 1%–4.5% higher accuracy compared to baseline methods Predictive Modeling on Hyperlactatemia Sepsis <i>05/2015 - Present</i> <ul style="list-style-type: none">Analyzed a variety of characteristics of different ICU patient populations extracted from the MIMIC-II database, such as mortality rates, the usage of antibiotics and intravenous fluids, and timing of severe sepsis

- Examined different definitions of severe sepsis and its correlation with ICU, hospital, and 30-day mortalities
- To build predictive models on a patient's likelihood of developing severe sepsis using statistical approaches (e.g., Cox proportional hazards model and logistic regression) and machine learning algorithms (e.g., support vector machine and random forest)
- To predict the real-time probability of lactate clearance considering a patient's demographic and physiological characteristics and clinical interventions

Care Variation among Congestive Heart Failure Patients *09/2013 - 08/2014*

- Applied vector space models to measure the similarities between congestive heart failure patients based on their medication and procedure orders
- Utilized clustering (K-means and co-clustering using bipartite graph) and topic modeling methods to detect patient subgroups within the congestive heart failure population according to medication orders, diagnostic and demographic information
- Implemented the above models using Spark on Hadoop clusters with significantly decreased computation time

Text Mining on Restaurant Reviews from Yelp *03/2015 - 04/2015*

- Preprocessed 1 million restaurant reviews crawled from Yelp with tokenization, stemming, stop word removal, and normalization, to construct N-gram vector space representation for text documents and computed similarity among different documents
- Implemented statistical language models with maximum likelihood estimation and smoothing; generated text documents from language models and evaluated the constructed N-gram language models
- Developed a text categorization system, including feature selection, Naive Bayes and KNN classifier, to distinguish positive and negative restaurant reviews
- Evaluated the text categorization system with 10-fold cross-validation and performed parameter tuning to explore the best configuration of KNN with brute force and random vector projection

Spam Filtering Using Machine Learning Approaches *09/2013 - 11/2013*

- Constructed static models using generalized linear regression based on principle component analysis and log transformation on the explanatory variables
- Modeled the trend, seasonality, and random fluctuation for spam and ham using time series data
- Combined the static and time series models using Bayes rule to further improve the performance of the spam filter

Web Crawler and Document Analysis & Retrieval *09/2014 - 11/2014*

- Programmed a web crawler for a medical forum in Java and extracted all posts in every threaded discussion about some particular topics, such as Diet, Hypertension, etc.
- Generated a document analyzer to tokenize, normalize, and stem the crawled documents, identified the most frequent words to compare with the standard stop words, and validated the Zipf's law
- Built an information retrieval system to retrieve similar documents of the input queries based on vector space model and popular language models, such as BM25

Creative Scheduling in Cardiac Catheterization Lab

01/2013 - 05/2013

- Analyzed case-level data and identified root causes for low utilization and volume variation of the catheterization lab
- Identified patient arrival patterns and predicted resources required using queuing theory

Predictive Modeling on Hospital Readmission

04/2012 - 05/2013

- Collected clinical and demographic data related to patient characteristics and diagnosis and identified significant risk factors in readmission using statistical analysis
- Accomplished prediction models using Nave Bayes, particle swarm intelligence-based support vector machine (SVM), and ensemble neural networks, and improved average classification accuracy by 22% to assist clinicians in identifying high-risk patients

PUBLICATIONS

1. **Jinghe Zhang**, Haoyi Xiong, Yu Huang, Hao Wu, Kevin Leach, and Laura E. Barnes. *M-SEQ: Early Detection of Anxiety and Depression via Temporal Orders of Diagnoses in Electronic Health Data*, 2015. *IEEE International Conference on Big Data (IEEE BigData 2015)*, Santa Clara, CA, submitted.
2. Bichen Zheng, **Jinghe Zhang**, Sang Won Yoon, Sarah S. Lam, Mohammad Khasawneh, and Srikanth Poranki. Predictive modeling of hospital readmissions using metaheuristics and data mining, 2015. *Expert Systems With Applications vol.42, no.20*, pp 7110–7120.
3. **Jinghe Zhang**, Sarah S. Lam, and Srikanth Poranki. A Classification Model for Hospital Readmission Using Combined Neural Networks, 2013. *Proceedings of Industrial and Systems Engineering Research Conference (ISERC). May 18-22. San Juan, PR.*
4. **Jinghe Zhang**, Sang Won Yoon, Mohammad Khasawneh, Srikanth Poranki, and Krishnaswami Srihari. A Readmission Prediction Model Using Swarm Intelligence-based Support Vector Machine, 2013. *Proceedings of Industrial and Systems Engineering Research Conference (ISERC). May 18-22. San Juan, PR.*
5. Chanchal Saha, **Jinghe Zhang**, Sang Won Yoon, Mohammad Khasawneh, and Krishnaswami Srihari. Selection and Matching of Kidney Donor and Recipient Using Fuzzy Techniques and Analytic Hierarchy Process, 2012. *Proceedings of Industrial and Systems Engineering Research Conference (ISERC). May 18-22. Orlando, FL.*

HONORS & AWARDS

Graduate Research Scholarship, University of Virginia, 2014-2016
Commonwealth Fellowship, University of Virginia, 2013-2014
Graduate Research Scholarship, Binghamton University, 2011-2013
National Aspiration Scholarship, China's Ministry of Education, 2009

SKILLS

Programming: Python, Java, R, Matlab, SQL, HTML, CSS, JavaScript, Hadoop, Spark
Operating Systems: Windows, Unix/Linux
Tools: Latex, Git, SVN, Vim, Emacs, Arena, Simio, ExpertFit, Minitab, NetLogo
Certificates: Lean Six Sigma Green and Black Belt from Dartmouth College
Languages: English and Chinese