Shanshan Zhang

Contact Information

Department of Computer and Information Science
Temple University
Science Education and Research Center 303
1925 N. 12th St, Philadelphia, PA, USA 19122

Cell: +1-216-533-3431
zhang.shanshan@temple.edu

Research Interests

Data Mining Applications (Remote Sensing, Behavioral Data Analysis, Text Mining, Computational Advertising, Bioinformatics).
Machine Learning (Deep Learning, Structural Regression, Sequential Pattern Mining, Representation Learning).
Data Visualization, **Optimization**, **Graph Algorithms**.

Education

Temple University, Philadelphia, PA, USA

Ph.D. student, Computer and Information Science — Aug. 2013 - Present
Center for Data Analytics and Biomedical Informatics Department of Computer and Information Science. GPA 3.95/4.

- Advisors: Dr. Slobodan Vucetic
- Committees: Dr. Slobodan Vucetic Dr. Krishna Kant and Dr. Eduard C. Dragut

Case Western Reserve University, Cleveland, OH, USA

Teaching Assistant, Computer Science — Aug. 2012 - May 2013
Department of Electrical Engineering and Computer Science. GPA: 4/4

Wuhan University, Wuhan, Hubei, China

M.S., Computer Science — Sept. 2010 - Jun. 2012
State Key Lab of Software Engineering. GPA: 3.85/4

B.Eng., Software Engineering — Sept. 2006 - Jun. 2010
International School of Software. GPA: 3.54/4

Research Experience

Leader — Feb. 2016 – Present
BDD: Dynamic Evolution of Smart-Phone Based Emergency Communications Network, NSF.

- Crawled data through Twitter API; Developed semi-supervised approach to discover emergency related tweets.
- Analysed human mobility and disturbance during emergency with geo-enabled Tweets.
- Developing **CNN-based sentence classification** algorithm.

Leader — May 2015 – Present
III Small: A Discriminative Modeling Framework for Mining of Spatio-Temporal Data in Remote Sensing, NSF

- Developed and implemented a semi-supervised **Gaussian Random Fields** that can predict AOD values with missing features and missing labels.
- Developing **Fully Convolutional Networks** with remote sensing data.

Leader — May 2014 – Present
Knowledge fusion with millions of Linkedin profiles

- Crawled around 67 million Linkedin profiles; Organized data with MongoDB.
- Analysed sampling bias on Linkedin; Applied **entity linking** and **word embedding** techniques to normalize noisy names, titles of user profiles.

Core member of a 2-member team — Sept. 2015 – Jan. 2016
SBIR I: Using Data Mining to Optimally Customize Therapy for Individuals with Autism, NSF

- Visualized trial data and helped one autism child to customize the therapy.
- Extracted **sequential patterns** and developed models to predict trials.

Core member of a 2-member team Feb. 2014 – Feb. 2015
Computational Advertising, *Yahoo! Labs*
- Extracted Yahoo! commercial email data using **MapReduce** platform;
- Represented email data and developed models with efficient feature selection to increase the conversion rate of multiple types of customers.

Core member of a 2-member team Nov. 2013 – Jan. 2014
Critical Assessment of Function Annotations 3
- Developed a multi-source k-nn algorithm to predict Gene Ontology annotations for hundreds of thousands proteins.
- Performance was among the top 10 teams (out of 60 teams).

Refereed Publications

1. Jiang, Y., *et al.*, Kansakar L., **Zhang, S.**, Vucetic, S., *et al.*, Friedberg I., Radivojac P., *An expanded evaluation of protein function prediction methods shows an improvement in accuracy,* Genome Biology, 2016. (*in press*)
2. **Zhang, S.**, Vucetic, S., *Semi-supervised Discovery of Informative Tweets During the Emerging Disasters,* The 4th International Workshop on Social Web for Disaster Management, CIKM, Indianapolis, IN, Oct 2016.
3. **Zhang, S.**, Vucetic, S., *Sampling Bias in LinkedIn: A Case Study.* The International World Wide Web Conference (WWW), Montreal, Canada, 2016. (*short*)
4. Han, C., **Zhang, S.**, Ghalwash, M., Vucetic, S, Obradovic, Z., *Joint Learning of Representation and Structure for Sparse Regression on Graphs,* The 16th SIAM Intl Conf. Data Mining (SDM), Miami, FL, May 2016..

Unpublished Work

1. **Zhang, S.***, Han, C.*, Vucetic, S., Obradovic, Z., *Aerosol Retrieval Using Gaussian Conditional Random Fields with Constraints.* (* for equal contribution)
2. Kansakar, L., **Zhang S.**, Vucetic, S., Djuric N., Bhamidipati, N., Radosavljevic, V., Grbovic, M., *How Valuable is an Inbox: Exploring Machine Learning Approaches for Purchase Prediction using E-mail Data.*

Internship Experience

Research & Development Intern Sept. 2011 – Dec. 2011
Alibaba Cloud Computing of Alibaba.com Corporation, Beijing

Professional Service

External Conference Reviewer: KDD, AAAI, ICDM, BIBM, IJCAI, ICML, NIPS, AISTATS.

Skills

Programming Languages
Python, Matlab, Java, C++, MongoDB, SQL, HTML, PHP, Javascript, MapReduce with practical experiences.
Deep Learning Models with Caffe and Theano
NN, CNN, RNN, LSTM, FCN

Awards

- SDM'14 Travel Award 2016
- SDM'16 Travel Award 2014
- Outstanding Undergraduate Award, Wuhan University (10% awarded) 2010
- National Scholarship from Ministry of Education, China (8 out of 485) 2009