

XUEQIN PANG

Relocating to Fort Washington, PA | Do **NOT** need work authorization sponsorship | 608-628-0319 | xpang3@wisc.edu

PROFILE

Dedicated professional with a M.S. in Computer Science and a Ph.D. in Pharmaceutical Science paired with five years of experience in programming, machine learning, data mining, and computer modeling. Solid knowledge of supervised/unsupervised machine learning, natural language processing, statistical analysis, data storage, and database system structure. Proven successes in solving problems and applying experience and knowledge to tasks.

EDUCATION

UNIVERSITY OF WISCONSIN-MADISON

Ph.D. in Pharmaceutical Science, GPA: 3.9/4.0

Anticipated: May 2017

Thesis: "Fast and Automated: Employing Machine Learning Algorithms to Analyze Neuropeptide Sequence and Structure in Mass Spectrometry"

- 2014 Paul and Kathryn Pisarzewicz Scholarship

M.S. in Computer Science, GPA: 3.8/4.0

Dec. 2016

Relevant Graduate Coursework: Machine Learning, Artificial Intelligence, Database Management, Advanced Database Management, Big Data, Operating Systems, Distributed Systems, Network, Data Structure, Algorithms

- 2016 officer in the Association for Computing Machinery (SACM) student chapter.
-

COMPUTER SKILLS

Programming:	Python/iPython, SQL, Java, Scala, R, C/C++, MATLAB, HTML, JavaScript
Platform:	MySQL, ScikitLearn, Weka, Eclipse, Github, Linux, LaTeX
Cloud Computing:	MapReduce, Hadoop, GraphX, SPARK, Tez, Hive, MLLib, Storm
Modeling/Simulation:	Molecular Dynamics, Monte Carlo, Docking, homology modeling, distributed computing

PROJECT EXPERIENCE

Performed Natural Language Processing on Twitter *via* Unsupervised Learning (M.S.) | 2016

Environment: Java, Scala, JSON, Hadoop, Spark & Storm Data Analytics Stacks

- Deployed Hadoop, Spark and Storm on a self-built virtual machine cluster
- Streamed 5 million tweets and summarized top 10 common words in a high-throughput manner for every 30 seconds
- Built TF-IDF vectors and clustered with k-means clustering algorithm to identify trending topics in Twitter
- Identified most popular words and frequency of appearance using GraphX

Created a Simplified Operating System within Systems Development Life Cycle Framework (M.S.) | 2014-2016

Environment: C & C++

- Built command line interpreter that parses/executes user commands
- Constructed dynamic memory allocator for user-level processes
- Created multi-threaded concurrency control *via* spinlock
- Developed HTTP based multi-threaded web server with FIFO, SFF, and SFF-BS scheduling policies
- Built an AFS-like distributed file system with crash recovery

Designed Data Mining Pipeline to Automate Product Matching in Amazon and Wal-Mart (M.S.) | 2016

Environment: Python, SQL, JSON, Natural Language Toolkit, and Scikit-learn

- Performed natural language processing on product information, transformed and stored in JSON format
- Made predictions on 200K pairs of unlabeled products based on 20K pairs of labeled product data
- Predicted entity matching with classifiers: decision tree, random forest, logistic regression, SVM, and naïve Bayes
- Achieved 96% precision and 90% recall by applying ensemble method and heuristic rules

XUEQIN PANG

Relocating to Fort Washington, PA | Do **NOT** need work authorization sponsorship | 608-628-0319 | xpang3@wisc.edu

Predicted Protein Mass Spectrometry Signal-Structure Relationship *via* Supervised Learning (Ph.D.) | 2016

Environment: Java, R, and Weka

- Crawled, indexed, and unified protein structural data from online protein databank
- Generated feature vectors from protein structures and calculated protein collision cross-section
- Predicted protein mobility signal out of reduced structural features *via* LASSO regression
- Improved prediction accuracy *via* support vector regression algorithm

Developed Statistical Analysis Tools for Ion Mobility Measurements (Ph.D.) | 2014-2016

Environment: Java

- Designed tools to automatically identify target signal distribution from a noisy and wide spectrum
- Identified structure-mobility relation *via* nonlinear regression
- Modeled dynamic motion of ions and predicted their conformation distribution
- Clustered ion structures *via* K-Nearest Neighbors algorithm and derived their collision cross-section distributions

Diagnosed Lymphatic Diseases from Medical Images by Tree Augmented Naïve Bayes (M.S.) | 2014-2015

Environment: Java

- Implemented expectation-maximization (EM) algorithm to impute missing data
- Constructed a Bayes network, mapping symptoms to disease *via* Chow-Liu algorithm and Prim's algorithm
- Determined lymphatic disease probability with trained Bayes network of symptoms

Improved Performance of a Medical Database Management System (M.S.) | 2013

Environment: SQL and C++

- Implemented clock algorithm for buffer management
- Built file manager for heap files and created B+ tree index to organize medical data
- Developed front-end utilities for SQL, DDL, and DML

SELECTED PUBLICATIONS

- Pang, X.;** Jia, C.; Chen, Z.; Li, L. Structural characterization of monomers and oligomers of D-amino acid containing peptides using T-wave ion mobility separations, *Journal of the American Society for Mass Spectrometry*. 2016, in press
- Lietz, C.; Chen, Z.; Son, C.; **Pang, X.;** Cui, Q.; Li, L. Intrinsic structural preferences of the L7P Neuropeptide Y signal peptide revealed by ion mobility-mass spectrometry, *Analyst*. 2016, 141, 4863
- Pang, X.;** Han, K.; Cui., Q. A simple but effective modeling strategy for structural properties of non-heme Fe(II) sites in proteins: Test of force field models and application to proteins in the AlkB family, *Journal of Computational Chemistry*. 2013, 34, 1620 (cover article)
- Pang, X.;** Yang, M.; Han, K. Antagonist Binding and Antagonism Induced Conformational Dynamics of GPCR A2AAR, *Proteins*. 2013, 81, 1399-410

SELECTED CONFERENCE TALKS

- Pang, X.** "Structural Characterization of Monomers and Oligomers of D-amino Acid Containing Peptides (DAACPs) Using T-Wave and Nonlinear Ion Mobility Separations" Presented at American Society for Mass Spectrometry National Conference, San Antonio, TX. June 2016
- Pang, X.** "Structural analysis of monomeric and dimeric neuropeptide Y (NPY) with IM-MS, MD simulations and HDX-MS" Presented at American Society for Mass Spectrometry National Conference, St Louis, MO. June 2015