

XIANGWEN WANG

wxwdata@gmail.com ◇ github.com/XiangwenWang ◇ linkedin.com/in/wxw

COMPUTING LANGUAGES & TOOLS

Computing Languages Python, C, SQL, R, Bash, Java
Tools Numpy, Scipy, Pandas, Scikit-learn, Tensorflow, Keras, Xgboost, Gensim, Statsmodels, NLTK, Selenium, Flask, Hadoop, Git, SVN, Docker, AWS

EDUCATION

Virginia Polytechnic Institute and State University (Virginia Tech)

Ph.D. in Physics *March 2019 (expected)*

M.S. in Statistics *May 2019 (expected)*

M.S. in Computer Science *May 2018*

Graduate Certificate in Data Analytics *May 2017*

University of Science and Technology of China (USTC)

B.S. in Physics *July 2011*

EXPERIENCES

Graduate Research Student February 2017 - June 2018

Department of Computer Science, Virginia Tech (Advisor: Dr. Gang Wang) *Blacksburg, VA*

- ◇ Photo-based vendor re-identification on Darknet marketplaces using Deep Neural Networks
 - Utilized machine learning to solve a cyber-security problem: re-identification of anonymous darknet sellers
 - Captured photography styles from TB-level product photos with Neural Networks (VGG, ResNet, etc.)
 - Performed writing-style analysis on product descriptions using lexicon and Part of Speech (POS) features
 - Identified coordinated activities such as price manipulation, buyer scam, and product reselling

Data Science Summer Intern June 2016 - August 2016

Rule14, LLC (Supervisor: Kai McDonald) *Santa Monica, CA*

- ◇ NLP for key sentences (issues, holdings, facts) extraction from legal documents
 - Performed POS and Named Entity Recognition (NER) tagging during feature engineering
 - Adopted Conditional Random Fields (CRF) for sentence sequence classification
- ◇ Churn analysis of commercial utility usage with Logistic Regression and Random Forest
- ◇ Medical Form Categorization with Optical Character Recognition and Support Vector Machine (SVM)

Graduate Research Assistant March 2013 - Present

Stochastic Processes Laboratory, Virginia Tech (Advisor: Dr. Michel Pleimling) *Blacksburg, VA*

- ◇ Foraging patterns in online searches
 - Quantified human information foraging strategies on search engines under random walk framework
 - Analyzed a hundred million empirical click-through logs collected from major search engines
 - Estimated model parameters and hyperparameters via Maximum Likelihood Estimation and K-S Statistics
 - Performed model selection with Akaike Information Criterion (AIC)
 - Utilized diffusion theory to show the similarity between human online searching and animal food foraging
- ◇ Behavioral analysis of virtual item gambling
 - Collected gambling logs and wealth information of 100K gamblers
 - Discovered that the random walk pattern of player payoffs is close to a truncated Lévy Flight
 - Simulated the dynamics of payoffs with Markov Chain Monte Carlo (MCMC) Method
 - Provided an analytic solution to explain the crossover from super-diffusion to diffusion
- ◇ Scaling properties and correlations in human mobility
 - Quantified human mobility patterns based on large datasets of human GPS trajectories
 - Revealed the long-term memory effect and cascade-walk patterns in human movements
 - Raised a new mechanism to explain the non-Poisson distribution in human mobility

- Proposed an agent-based stochastic model to explain human mobility patterns
- ◇ Human mobility patterns in different contexts
 - Built an 8-node Hadoop & Spark cluster for data processing
 - Analyzed TB-level GPS logs, and identified movements in urban, suburban and rural areas
 - Evaluated the similarities and differences in human movements under different contexts
- ◇ Modeling the wager distribution and risk attitude in online gambling of pure chance
 - Scraped 8 TB gambling logs from online gambling websites
 - Performed data encryption before storing them in SQL databases
 - Revealed that the wager distribution in online gambling is close to log-normal
 - Discovered scaling properties in the risk management strategies of online gamblers

Graduate Teaching Assistant

August 2012 - December 2018

Department of Physics, Virginia Tech (Instructor: Dr. John Simonetti)

Blacksburg, VA

- ◇ Instructed labs and recitations for 1200+ students to help them better understand lecture contents
- ◇ Provided services for students with disabilities to ensure they have equal educational opportunities
- ◇ Held office hours to explain physics concepts to students from different backgrounds

SELECTED COURSE AND SIDE PROJECTS

- ◇ Mediumlarity Prophet: a webapp to predict and improve the popularity of Medium articles January 2019
- ◇ Time series prediction of flight delays with recurrent neural networks April 2017
- ◇ Predicting taxi fare with stochastic gradient boosting November 2016
- ◇ Noise reducing module for Twitter information retrieval engine (based on Apache Solr) April 2015
- ◇ Parallel simulation of relaxation processes in a system with logarithmic growth November 2014

SELECTED PUBLICATIONS

- Behavior Analysis of Virtual Item Gambling** July 2018
Physical Review E, Xiangwen Wang, Michel Pleimling
- You Are Your Photographs: Detecting Multiple Identities of Vendors in the Darknet Marketplaces** June 2018
ASIACCS'18 (acceptance rate: 17%), Xiangwen Wang, Chun Wang, Peng Peng, Gang Wang
- Foraging Patterns in Online Searches** March 2017
Physical Review E, Xiangwen Wang, Michel Pleimling
- Correlations and Scaling Laws in Human Mobility** January 2014
PLOS ONE, Xiangwen Wang, Xiaopu Han, Binghong Wang

SELECTED HONORS AND AWARDS

- Mu Sigma Rho Membership (Virginia Tech) April 2017
- Ray F. Tipsword Graduate Scholarship (Virginia Tech) April 2015
- Outstanding Student Scholarship, Grade 1 (USTC) October 2007

RELATED COURSES

- Machine Learning** Data Analytics I, Data Analytics II, Information Storage and Retrieval, Applied Machine Learning in Security, Information Visualization
- Statistics** Inference Fundamentals, Probability and Distribution Theory, Statistical Inference, Experimental Design, Regression and ANOVA, Linear Models Theory, Bayesian Statistics, Survival Analysis, Statistical Programming Packages, Communication in Statistical Collaborations, Internship in Statistics, Time Series Analysis
- Computer Science** Theory of Algorithm, Database Management System, Graph Theoretic Methods, Computational Physics, Data Structure and Database, Fundamentals of Java Software Development, Computational Methods, C Language Programming