Zach (Zhi ZHANG) GID：903203669

Contact #: 859-699-9829 Email: zzhang608@gatech.edu Boston/Atlanta/Beijing

STATEMENTS AND STRENGTHS

Have intensive and extensive practice in industry, self-motivated and funded with high commitment and research aptitude – curious, collaborative, persistent, assertive, and candid, easy-going and love to teach and share. Love math, algorithm and programming, a multi-disciplined generalist, have profound faith and favor in machine learning and AI.

EDUCATION

- Georgia Institute of Technology Master of Computer Science (GPA:4.0) Attending Since Jan2016
- University of Kentucky Master of Animal and Food Science (GPA: 3.8) Jan2011 - May2012
- Peking University School of Medicine Pre-Med Studying Sep2009 - Aug2010
- China Agricultural University Bachelor of Food and Nutrition Engineering, with Honor (China college entrance examination score 644/750, GPA: 3.7, finished 4 year program in 3 years) Sep2006 - Jul 2009

Some favorite and related courses (100 as full score, A: 90-100, G: Graduate, U: Undergraduate): Machine Learning (G: A), Reinforcement Learning, AI for Robotics (G: A), Computer Vision (G: A), Knowledge-based AI (G: A), Machine Learning for Trading (G: A), Software Development Process (G: A), Statistical Analysis (G: A), Bio-Statistics (U: A), Advanced Mathematics I-II (U: 99), Probability Theory & Mathematical Statistics (U: 96), C Programming Language (U: 88), Electrotechnics & Electronics (U: 93), Engineering Mechanics (U: 95), Economics Theory (U: 91)

HONHONORS

- National Scholarship Ministry of Education of China 2007
- First Class Scholarship for Academic Excellence China Agricultural University 2007, 2008
- China National Cereals, Oils and Foodstuffs CorporationScholarship 2008
- Hua Luogeng Golden Cup Youth's Invitational Competition (Math) 2rd 2002
- Chinese Physics Olympiad 3rd in Province 2003

SOCIAL ACTIVITY

- Personal Trainer, muscle building and diet balance, Wills Gym, Beijing, Oct 2016- Feb 2017
- Active Chinese Registered Dietitian License, since 2008
- Yellowstone National Park Service volunteer, April – Jun, 2015
- Promoting Animal Welfare and Services volunteer, University of Kentucky, 2011, 2012
- JiuZhai Valley national park reconstruction volunteer after Sichuan Earthquake, 2009
- 29th Summer Olympics Volunteer, Red Cross volunteer service, Beijing, Jul - Aug 2008
- Engineer Internship in China National Cereals, Oils and Foodstuffs, Beijing, Jan – Feb, 2008
- Journalists' Association, Student Community of China Agricultural University, 2006, 2007

EMPLOYMENT

Selux Diagnostics Inc. (medical device and biotechnology company) Charlestown, MA

Data Scientist Sep 2017 – Current

- Worked with two Yale PhD funders and R&D team, collaborated with MGH, U. Michigan, to develop new generation of patients' antibiotics susceptibility test device, helped with successful 50 million fund raise.
- Responsible for the algorithm development of the company's product: to predict the patients' antibiotic susceptibility testing index. Tested different ways to represent the problem structure, such as multilabel classification, sequence classification, applied and modified different machine learning algorithms to improve the prediction accuracy: Kernel Bayes, XGB, hmm, Bayes nets with Chow-Liu algorithm, conditional random fields, and deep learning LSTM, GRU
- Defined and implemented own scoring function considering more conditions to be optimized in model selection cross validation and grid search parameter tuning. Tried different optimization algorithms including randomized optimized algorithms and compare their performance to solve the learning objective.
- Evaluation: compared different models performance, learning curves and training time over sample ticks
- Set up the server environment for the algorithm everyday training, back testing, communicating with database
- Embedded the algorithms into the system of device to do seamlessly device running and prediction by building a multi-functional software: which transmits Json data through pipe message between the platform of robotic device and Python/Tensorflow/Matlab through different threads, and de/re-serializes Json files with C# or Python/Matlab data objects, and maps database entity framework to data objects and backfills the training and

testing models into database, which also makes the algorithm training and testing results visualized through the web, then lunched the application for internal and integrated it with the whole cloud system in customer sites.

Brigham and Women's Hospital & Harvard Medical School (Division of Rheumatology, Immunology and Allergy, Section of Clinical Research) Boston, MA
Advanced Data Analyst Jun 2015 – Aug 2017

- Evaluated different predictive models to improve computational efficiency and reduce false-positive edges, and applied them in large-scale inference of disease risks on individual patients
- Identified different groups of pharmacy adherence that differentiate patients regarding clinical resources, pharmacy characteristics and health outcomes by using group-based trajectory model and marginal structural model (PI: Dr. Candace Feldman)
- Applied mix effected models to predict cost of pharmacy and hospitalization using features such as diagnosis and disease severity (PI: Dr. Daniel H. Solomon)
- Applied unsupervised learning clustering algorithms to characterize the patients' pain-related profiles, used the elbow methods to select the number of clusters. Applied the PCA to reduce the dimension of patients' features before clustering. Identified the relationship between patient's features and patients' clustering groups. (PI: Dr. Yvonne Lee)
- Applied genome-wide association study (GWAS) analysis and PCA to identify the potential risk SNPs for gout patients with hyperuricemia using genomic sequencing data, applied the classification models to evaluate the association of risk SNPs with the presence of patient disease outcome. (PI: Dr. Daniel H. Solomon and Jing Cui)
- Studied NLP methods to perform segmentation, lexical analysis, natural language parsing, and conceptual indexing to extract accurate medical concepts from narrative doctor notes to identify and classify patients with specific diseases and status with a high positive predictive value. (PI: Dr. Katherine Liao)
- Studied and compared methodologies and algorithms for solving high-dimensional data prediction, such as dimensionality reduction, tensor factorization, and regularized regression models: elastic net, sparse modeling (e.g. Lasso), Ridge, and application in patients' phenotypes predication
- Helped maintain the scalability, robustness, effectiveness, security of the healthcare data environment to enable scalable storage and parallel processing of data
- Conducted data ETL, interacted with source system to understand the data and layout the specification to bring the data into the landing area. Enforced best practices and guiding principles around data acquisition and data quality.
- Linked, sorted, annotated, and augmented data such as massively disparate, potentially incomplete and noisy big insurance claims, electronic medical records, Nex-gen sequencing, clinical procedures
- Developed tools and applications for scientists and doctors to easily store, processing, analyze and visualize data.

Teletech Insights & FocusKPI Boston, MA
Consultant Feb 2013 – May 2015

- Worked on multiple big data SaaS projects related to customer behavior, based on structured and unstructured data, performed hypothesis testing, algorithm application, prediction, and data visualization, built the algorithm training, testing and reporting software.
- Performed content based and collaborative filtering for item recommendation, applied vector space model and performed latent semantic analysis to match the advertisement with webpage for accurate online advertising, and predicted CTR for client
- Scraped Harvard Business Review online contents and NLP contents to quantify words and topics; applied logistic regression, XGboost, and Bayes classifier to predict online subscription by features such as article page view counts and article major attributes including topic, word count, images, bullets, and insets; evaluated the performance of different algorithms through cross-validation
- Used Web Spider to pull information from hundreds of career web pages; retrieved data and information by using pattern recognition, clustering, and feature selection to explore patterns and predict candidate fits to support job seekers-company match for client
- Applied multiple predictive models to solve business questions related to marketing engagement and finance:
- Fitted ARIMAX models using marketing channel spend and revenue data to forecast channel level revenue for different prediction duration and skip months' scenarios; applied an ADL model and calculated the Adstock effect to make recommendations on advertisement spending allocation; applied various models such as Markov chain and logistic regression to predict customer life time value

- Applied k-means clustering segmentation framework to cluster customer, discover factors most strongly affecting customer's response to a promotion and to identify the best customers, marginal buyers and non-buyers for targeted marketing

University of Kentucky UK bioinformatics microarray core facility Lexington, KY
Research Assistant Jan 2011 – May 2012

- Used Microarray generated big genomic data to model the gene, RNA, protein regulatory pathway networks
- Applied PCA to do dimension reduction for expression data
- Investigated the relationship between the gene expression networks and phenotype by statistical learning models

Fudan University School of Public Health Shanghai, China
Research Assistant Aug 2010 – Dec 2010

- Helped on projects to study the relationship between the food systems, environmental sustainability, dietary and nutrition and their interactions with genetic factors in the cause and prevention of obesity and diabetes (PI: Dr. Jing Cui)
- Used the nurses' 10 years longitudinal data and genomic data to generalize association and inference by PCA and Proportional hazards model analysis.
- Assisted in research project design and progress; presented results at the meetings and seminars

Peking University School of Basic Medicine Beijing China
Research Assistant Sep 2009 – Jul 2010

- Conducted computational analysis of gene regulation, genetic variation, and the computational algorithm based inference of genome-transcriptome-phenome association of patients' neural stem cell (PI: Dr. JinHua Wen)
- Computational modeling the whole physiological pathway regulated by targeted genes and proteins of stem cell
- Performed classification, association and cluster rule mining, extracting useful information out of enormous complex genomic datasets

SKILLS

Applied Statistics/Machine Learning (Understand the Basic theory and able to Implement): Time Series, Cox Model, Supervised Learning (Regression and Classification, Decision Trees, Neural Networks, Support Vector Machines, Bayesian Learning and Inference), Unsupervised Learning (Feature Selection, Tensor Factorization), Reinforcement Learning (Markov Decision Processes, Game Theory, Q-Learning), Randomized Optimization, Bayes Network, Probabilistic Graphical Models.
Programming: Python/Matlab/R, C#/C/C++, Java
Database: SQL, Hadoop (HDFS, MapReduce, Hive, Storm, Hbase), Spark (Core, Streaming, MLlib, SQL, GraphX)
Tools: OpenCV, TensorFlow, Keras, Scikit-learn***,*** Numpy, Pandas, Scipy***,*** Weka, PBNT, dynet, BURLAP, Gym
Vision Analysis: registration, alignment, matching, scene categorization, 2D/3D analysis, MHI
Text Analytics/NLP: Part-of-speech tagging, Bag-of-Words, N-Gram, TF-IDF, Word2Vec, NLTK
Cloud Platform used: AWS (EC2, RDS, DynamoDB)

SCHOOL PROJECTS

Georgia Institute of Technology 2016 – 2018
College of Computing

- Class: Machine learning for Trading, 2016
  In course project, I developed a simple decision support learner by evaluating different machine learning algorithms, linear regression, KNN regression, bagging, ensemble learning and Q-learning to predict stock price and decide real time long/short strategy to allocate investments resource and maximize portfolio value
- Class: Knowledge-Based Artificial Intelligence, 2016
  Through the course, I tested approaches to find a human-level, human-like intelligent agent that can answer human intelligence tests such as visual analogy problems based on both verbal and visual representations
- Class: Artificial Intelligence for Robotics, 2017
  Simulated Intelligent Robot Tracking Agent: in course project, I and two team members developed a naive intelligent agent to predict the future trajectory of a Nano robot's dynamic moving position; evaluated multiple training algorithms in Bayesian probabilistic model, linear-Gaussian model (Kalman Filters), sequential Monte Carlo simulation (particle filters), residual learning model; reduced video data dimensionality by PCA; tuned residual neural network hyperparameters and applied bootstrap aggregation with multiple residual neural networks
- Class: Computer Vision, 2017

Activity classification using MHI: In class project, I classified different human movements behavior by training video data containing multiple human movements, using the motion history image stack to represent the video data, performed background subtraction, motion history images moments calculation, and applied different classifiers to train the image moments, then predicated the human behavior from real-world video.

- Class: Reinforcement Learning, 2018
  I created an agent by implementation of Double Deep Q-learning with experience replay trained with 5000 epoch, that can guide a space vehicle to land autonomously in the environment without crashing, by setting up an environment using the OpenAI's Lunar Lander problem, which has an 8-dimensional state space and 4-dimensional action space.

CO-AUTHORED PUBLICATIONS OR MANUSCRIPTS OR ABSTRACTS:

Abstracts and papers

- Feldman Candace, Zhi Zhang, Daniel Solomon, Association between Hydroxychloroquine Nonadherence and Adverse Outcomes Among Patients with Systemic Lupus Erythematosus: A marginal structural model approach. 2017 ACR/ARHP Annual Meeting
- Jing Cui, Zhi Zhang, Elizabeth Karlson, Daniel Solomon, Whole genome association study of Gout in Patients with Hyperuricemia Identified Many Possible New Candidate Risk Alleles. 2017 ACR/ARHP Annual Meeting
- Alyssa Wohlfahrt[1], Zhi Zhang[1], Bing Lu, Clifton O. Bingham III, Marcy B. Bolster, Wendy Marder, Larry W. Moreland, Kristine Phillips, Tuhina Neogi and Yvonne C. Lee, Identification of Clinically Relevant Pain Profiles in Individuals with Active RA. 2017 ACR/ARHP Annual Meeting (1: co-first author)
- Katheeln Vinny, Zhi Zhang, Danial Soloman, Prevalence of Cytopenias Associated with Low-Dose Methotrexate and Folic Acid Among Patients with RA: A Systematic Review and Meta-Analysis. 2017 ACR/ARHP Annual Meeting
- Feldman Candace, Collins Jamie, Zhang Zhi, Subramanian S.V., Solomon Daniel, Kawachi Ichiro, Costenbader Karen. 2017. Dynamic Patterns and Predictors of Hydroxychloroquine Nonadherence among Medicaid Beneficiaries with Systemic Lupus Erythematosus Arthritis & Rheumatology
- Lee YC, Bingham CO, Edwards RR, Marder W, Phillips K, Bolster MB, Clauw DJ, Moreland LW, Lu B, Wohlfahrt A, Zhang Zhi, Neogi T. 2016. Pain sensitization is associated with disease activity in rheumatoid arthritis patients: A cross-sectional study. Arthritis Care Res. 2017
- David Kreps, Florencia Halperin, Sonali Desai, Zhi Zhang, Elena Losina, Amber Olson, Elizabeth Karlson, Bonnie Bermas, Jeffrey Sparks. 2017. Association of weight loss with improved disease activity in patients with rheumatoid arthritis Clinical Rheumatology
- Sara K. Tedeschi, Michelle Frits, Jing Cui, Zhi Zhang, Taysir Mahmoud, Christine Iannaconne, Tzu-Chieh Lin, Kazuki Yoshida, Michael E. Weinblatt, Nancy A. Shadick, Daniel H. Solomon. 2016. Diet and Rheumatoid Arthritis Symptoms: Survey Results From 217 Patients. Arthritis Care & Research

RECOMMENDERS:

Daniel H. Solomon, M.D., M.P.H.,
Chief, Section of Clinical Sciences, Professor of Medicine, Harvard Medical School
Matthew H. Liang Distinguished Chair in Arthritis and Population Health
75 Francis Street Boston, Massachusetts 02115 Tel: 617 732-5356, Fax: 617 732-5505
Email: dsolomon@bwh.harvard.edu

De Liu
Association Professor, University of Minnesota, 3M Fellow in Business Analytics
PhD Coordinator, Information and Decision Sciences
Carlson School of Management
Office phone: 612-626-4480
Email: deliu@umn.edu

Yvonne C Lee, MD
Associate Professor of Medicine, Northwestern University
McGaw Pavilion Suite M-300 240 E Huron Chicago IL 60611
Phone: 312/503-8003 Email: yvonne.lee@northwestern.edu

Candace H. Feldman, MD, ScD
Assistant Professor of Medicine (HMS), Associate Physician (BWH)
60 Fenwood Road, Office 6016P
Phone: 617-525-1035, Fax: 617-264-3019
Email: cfeldman@bwh.harvard.edu

Kuey Chu Chen
Research Associate Professor, Director of UK Microarray Core Facility
Health Science Research Building Rm#154
Phone: (859) 323-6241
Email: kueyc@uky.edu