

LIDA ZHANG

(224)-616-4394
lidazhang@tamu.edu

363 L.F. Peterson Building
College Station, TX, 77843

EDUCATION

Texas A&M University, College Station, TX Aug. 2019 - Now
Ph.D student in Computer Science (GPA: 4.0/4.0)
Advisor: Dr. Bobak Mortazavi

Northwestern University, Evanston, IL Sept. 2016 - Dec. 2017
Master of Science in Computer Science (GPA: 3.87/4.0)

Shandong University, Jinan, China Sept. 2012 - Jun. 2016
Bachelor of Engineering in Software Engineering (GPA: 89.01/100, rank: 1.2%)

RESEARCH INTERESTS

Machine Learning, Deep Learning, Artificial Intelligence

- Automated machine learning on time-series data.
- Data heterogeneity and feature space changing on medical data.
- Transfer learning for various data resource and limited training data.

WORK EXPERIENCE

Google Internship (Google Local Search Ads) May. 2022 – Now

- Improving local ads coverage by expanding keywords using Salient Terms and a deep-learning-based query clustering method. Filter close-variant keywords by normalization.
- Bad keyword filtering in the smart campaign using a multimodal language DL model.
- A semi-supervised model to address the problem of limited labels for sequence breakpoint prediction. A simulation experiment for YouTube ads positioning prediction.
- Build pipelines for new data generation, production health, and productionisation.

Teaching Assistant at Texas A&M University (CSCE 482: Senior Capstone Design) Jan. 2022 – May. 2022

Teaching Assistant at Texas A&M University (CSCE 421: Machine Learning) Aug. 2021 – Dec. 2021

Google Internship (Google Ads Forecasting) May. 2021 - Aug. 2021

- Implemented Prediction Intervals (PI) model in Tensorflow Extended (TFX) to predict expected advertising intervals (e.g., the interval of clicks on an advertisement).
- Addressed the consistency problem of advertisement forecasting by using Tensorflow Lattice (e.g., when the cost of each advertisement increases, the max and min of the expected clicks should both not decrease).

Research Assistant at Texas A&M University Aug. 2019 - May. 2021

Research Assistant at Northwestern University Jun. 2017 - Jun. 2019

PUBLICATION

- **Zhang, Lida**, Xiaohan Chen, Tianlong Chen, Zhangyang Wang, and Bobak J. Mortazavi. "Risk Prediction in Heterogeneous Electronic Health Records with Semi-supervised Meta-learning." Submitted to NeruIPS 2022 (under review).
- Wang, Xingyu, **Lida Zhang**, and Diego Klabjan. "Keyword-based topic modeling and keyword selection." In 2021 IEEE International Conference on Big Data (Big Data), pp. 1148-1154. IEEE, 2021.

- **Zhang, Lida**, Xiaohan Chen, Tianlong Chen, Zhangyang Wang, and Bobak J. Mortazavi. "DynEHR: Dynamic adaptation of models with data heterogeneity in electronic health records." In 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), pp. 1-4. IEEE, 2021.
- Huo, Zepeng, **Lida Zhang**, Rohan Khera, Shuai Huang, Xiaoning Qian, Zhangyang Wang, and Bobak J. Mortazavi. "Sparse Gated Mixture-of-Experts to Separate and Interpret Patient Heterogeneity in EHR data." In 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), pp. 1-4. IEEE, 2021.
- **Zhang, Lida**, Abdolghani Ebrahimi, and Diego Klabjan. "Layer Flexible Adaptive Computation Time." In 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1-9. IEEE, 2021.
- **Zhang, Lida**, Nathan C. Hurley, Bassem Ibrahim, Erica Spatz, Harlan M. Krumholz, Roozbeh Jafari, and Bobak J. Mortazavi. "Developing Personalized Models of Blood Pressure Estimation from Wearable Sensors Data Using Minimally-trained Domain Adversarial Neural Networks." Proceedings of the Machine Learning for Healthcare Conference, MLHC 2020.
- King, Zachary D., Judith Moskowitz, Begum Egilmez, Shibo Zhang, **Lida Zhang**, Michael Bass, John Rogers, Roozbeh Ghaffari, Laurie Wakschlag, and Nabil Alshurafa. "Micro-stress EMA: A passive sensing framework for detecting in-the-wild stress in pregnant mothers." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3, no. 3 (2019): 1-22.
- **Zhang, Lida**, Zachary King, Begum Egilmez, Jonathan Reeder, Roozbeh Ghaffari, John Rogers, Kristen Rosen et al. "Measuring fine-grained heart-rate using a flexible wearable sensor in the presence of noise." In 2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN), pp. 160-164. IEEE, 2018.

RESEARCH EXPERIENCE

Automated Machine Learning in Medicine

Mentor: Dr. Bobak Mortazavi, Dr. Zhangyang Wang

Jan. 2020 - now

- Proposed LSTM-based meta-learning for data various ICU durations on EHR data.
- Analyzing various ICU measurement frequencies by clustering.
- Proposed semi-supervised learning method and hierarchical structure for modeling heterogeneous ICU measurement frequencies.

Cuffless Blood Pressure Regression Modeling

Mentor: Dr. Bobak Mortazavi

Dec. 2019 - now

- Built multitask learning-based blood pressure regression models from bio-impedance data.
- Proposed adversarial training approach for reduced training data.
- Proposed meta-learning based adversarial training method to address the personalization barrier in wearable sensing modeling.

Layer Flexible Adaptive Computational Time for Recurrent Neural Networks

Mentor: Dr. Diego Klabjan

May. - Sept. 2018

- Proposed and implemented a flexible layer RNN model with adaptive computational time, targeting different difficulties of work in prediction tasks (e.g., unbalanced prediction).
- Proposed a new teacher-forcing strategy for seq2seq model and improved F-measure 2% to 10%.
- Proposed and implemented a well-performed seq2seq model with flexible layers and teacher-forcing algorithm on financial dataset (stock), and applied the model on Hutter prize Wikipedia dataset.

Keywords Latent Dirichlet Allocation: a keyword selection algorithm for information expanding

Mentor: Dr. Diego Klabjan

Aug. 2018 - now

- Captured keyword candidates based on viral tweets prediction and POS tagging algorithm.
- Built a keyword topic joint model based on probabilistic graphical model simulating Twitter data collection process, and proposed approximate inference techniques based on variational methods and an EM algorithm for parameter estimation.

Social Media Analysis & PR Robot for Marketing, partnership with large global bank

Mentor: Dr. Diego Klabjan

Mar. - Jul. 2018

- Monitored bank related tweets from Twitter API using multi-processing.
- Used regular expression to preprocess monitored data and created word cloud for unigram and bigram using stemming.
- Implemented Spanish spell check algorithm, trained word2vec model, and built an unsupervised product classification algorithm to detect potential tweets related to bank products (credit card).
- Identified potential new keywords from detected viral tweets applying POS tagging algorithm.
- Built sentiment analysis model specialized for Spanish Tweets applying Random Forest.
- Identified weekly topics using Latent Dirichlet Allocation (LDA), and named the topics automatically.

HABits Lab, Northwestern University, Chicago, IL, mentor: Dr. Nabil Alshurafa Apr. 2017 – Mar.2018

- **Biosensor Data Noise Detection:** a reliable heart rate calculation process with noise position detection.
 - Developed the most effective known algorithm for ECG noise detection using band-pass filter, neural network, and SVM. Successfully applied the algorithm on professional baseball players' dataset. Algorithm was adopted by the HABits Lab at Northwestern and is still in current use.
 - Participated in HABits Lab python library PASDAC development (noise cleaning section).
 - Designed and led a wearable biosensor test targeting noise data collection, and tested the noise detection algorithm on the data collected from college students.
- **Stress Detection and Prediction:** Designed and developed a stress detection model
 - Extracted and selected features from on physiological signals (ECG & GSR) collected by biosensors, proposed a novel stress detection model based on Random Forest.
 - Designed in-door and in-the-field stress detection experiments, tested experiments on college students and pregnant women. Proposed a method to capture reliable subjective stress level.
 - Co-led stress detection project: managed experiment finance, ordered devices and equipment, trained testers, scheduled participants, and took charge of experiments.

MIMA Lab, Shandong University, Jinan China

Sept. - Dec. 2014

- **Semi-supervised Learning in Text Analysis**
 - Led weekly project group meeting and machine learning discussion group.
 - Applied Multinomial Naïve Bayes and MNB-FM on Multi-Domain Sentiment Dataset, and proposed an optimizing method by combining propagation methods.

PROFESSIONAL EXPERIENCE

HEART SPACE – Startup Focusing on Critical Thinking and Social Emotional Education

Aug. 2016 - Feb. 2017

- Analyzed Chinese education market and built startup blueprint as one of the original founders.
- Built company website and ran a Wechat online communication social group.
- Won ¥50,000 from the Sany Foundation, and negotiated the acquisition of ¥ 20,000 in lab funding.
- Interviewed prospective course lecturers and online writers.
- Designed and led family parent-child workshops.

Intelligent Anti-Theft Tracking System

Jul. - Aug. 2016

- A teamwork program for Android mobile system
- Planned the project, and was personally responsible for telephone function, message function, and software testing.

Summer Internship at Shaanxi Amsinpul Data Communication Co., LTD. (XIPU Net), Shaanxi, China

- **Clothing Style Prediction System** Jul - Aug. 2015
 - A shopping mall business recommendation system for Nanchang Wanda Plaza: segmented people from background images and detected age by opencv libraries, classified clothing attributes from identified body region and extracted features such as texture, and built a model to predict the relation between age and clothing information.
- **Web Design** Jul - Aug. 2014
 - Participated in company website development with HTML, CSS, Java Script, and PHP.

TECHNICAL AND LANGUAGES SKILLS

Professional experience in Machine Learning, Deep Learning, NLP, data mining and device testing.

Python (PyTorch, Tensorflow, etc), Java, Matlab, C++, Database (SQL), Docker, Parquet

Web development abilities in HTML, CSS, Javascript, PHP.

Android application development ability.

Familiarity with Linux, Mac OS and Windows operating systems.

English - full professional proficiency