XUHUI ZHOU

zhouxuhui2018@gmail.com| https://xuhuizhou.github.io/

EDUCATION

Nanjing University Sep 2015 – Jun 2019
Majoring in Statistics
GPA:4.30/5.00

- Exceled in courses related to C++, SQL, Discrete math, Numerical computing and analysis.
- People's Scholarship for Research and Innovation (2 out of 132 students in Department of Mathematics).
- University-level outstanding students for 2018.

University of California Berkeley (International student) Aug 2017 – May 2018
GPA:3.55/4.00

- Exceled in Stat135: Concepts of Statistics (Prof. Yun S. Song), CS61: Structure and Interpretation of Computer Programs (Prof. John DeNero), and CS189: Intro to machine learning (Prof. Anant Sahai).
- Won the third place in Hog contest (Out of 1,000 students).

RESEARCH EXPERIENCE

Knowledge-based Dependency Parsing with Sentence-state LSTM June 2018 – Aug 2018
Research Assistant to Prof. Yue Zhang. Singapore University of Technology and Design

- Fully funded internship. Worked on building dependency parser that can naturally incorporate common sense knowledge with modifying sentence-state LSTM to an encoder. Written in Python with Pytorch frame.
- Built a dependency information database by parsing Gigaword (15GB) with Stanford Dependency Parser.
- Presented poster *Sentence-state LSTM* on 2018 Singapore Symposium of Natural Language Processing.
- Current model reached state of the art in the newswire domain and showed promising generalization to other domains such as emails, reviews and web blogs.

Analysis of Voting Systems Using Monte Carlo Simulation Sep 2017 – Dec 2017
Member of analytic committee of Data Science Society at Berkeley. University of California, Berkeley

- Examined the effectiveness of various voting systems via Monte Carlo simulation. Implemented in R and optimized the simulation part with careful design of the algorithms.
- Proposed a metric to quantify the satisfaction of voters. Identified strategic voting as an important factor.
- Found that the score voting system is the most effective, although plurality is least vulnerable to strategic manipulation. Ruled out Condorcet because of its inconsistency in producing a winner.

SELECTED PROJECTS

2018–2019

DNN speech recognizer

- Trained and tested models with LibriSpeech. (a corpus of 1,000 hours speech in English)
- Preprocessed the raw audio to Spectrograms and MFCCs. Experimented with with various neural network architectures for acoustic modeling.
- Built an end-to-end automatic speech recognition pipeline that accepts raw audio and returns transcription.

Machine translation: English to French

- Trained and evaluated models with datasets from Workshop on Statistical Machine Translation.
- Experimented with various neural network architectures and used an Encoder-Decoder RNN as the final model.
- Built a pipeline accept English text as input and return the French translation.

2017–2018

Mathematical handwritten symbols recognition

- Trained and tested the models with the Kaggle Handwritten math symbols dataset (Over 100,000 samples).
- Invoked novel CNN structures to automatically recognize 81 mathematical symbols with 82% accuracy.
- Introduced a platform to digitalize handwritten symbols to more robust Latex form.

Training a Smartcab to drive

- Worked towards constructing an optimized Q-Learning driving agent that will navigate a Smartcab through its environment towards a goal.

Finding donors for *CharityML*

- Employed several supervised algorithms to accurately model individuals' income using data collected from the 1994 U.S. Census, thus, helping a non-profit organization better understand how large a donation to request.

2016–2017

Titanic survival visualization and prediction

- Explored the Titanic disaster dataset from various aspects such as gender, age and social class.
- Implemented a naïve decision tree to predict survival chance.
- Built an interactive web visualization with CSS, Html5 and JavaScript.

Wrangling OpenStreetMap data

- Chose Nanjing, China in OpenStreetMap and used data munging techniques, such as assessing the quality of the data for validity, accuracy, completeness, consistency and uniformity, to clean the OpenStreetMap data.
- Proposed practical suggestions for OpenStreetMap to improve its data quality.

TEACHING EXPERIENCE

Teaching Assistant**, Machine Learning Nanodegree of Udacity** Sep 2018 – present

- Taught students from all over the world with different background basic machine learning knowledge ending with a project predicting cuisine from ingredients with logistic regression or other techniques.

Head Teaching Assistant, **Data Science Nanodegree of Udacity, China Region** Dec 2016 – June 2017

- Managed WeChat learning group and gave supplementary lectures for over 200 students with various backgrounds.

Volunteer Teacher, **AIESEC: Dare To Dream Program, Ningxia, China** Dec 2016– Jan 2017

- Taught Chinese high school students foreign cultures with excellent people from all over the world.

Community Teacher, **Summer School for Long Pao, Nanjing, China** July 2016 – Aug 2016

- Taught math to local students and assisted with courses in English, Geography and Literature.

NON-DEGREE STUDIES

Coursera June 2018 – Dec 2018

University of California San Diego & National Research University Higher School of Economics

- Algorithmic Toolbox, Data Structures, Algorithms on Graphs, Algorithms on Strings.

Udacity Mar 2016 – June 2018

- Data Analyst Nanodegree, Machine Learning Engineer Nanodegree, NLP Engineer Nanodegree.

Machine Learning Summer Camp, Nanjing University June 2017 – August 2017

SKILLS

Technical skills

- *Languages:* —Excellent: Python, C++. —Familiar: Java, SQL, Matlab, R, Javascript, CSS.
- *Libraries*: TensorFlow, Keras, PyTorch, scikit-learn, Pandas, NumPy, Matplotlib.
- *Other*: Git and GitHub, Bash, Linux (Fedora).

Languages

- Chinese: native speaker
- English: proficient (108 TOEFL iBT)
- Japanese and Korean: Entry level

INTERESTS

- Writing fictions, Taekwondo, calligraphy, violin, and Arena Of Valor.