JINGYI HE

Tel: (+1)514-661-0316 ⋄ Email: jingyi.he@mail.mcgill.ca
Website: kylie-box.github.io ⋄ github: github.com/kylie-box
Address: Apt.1210 195 Rue Peel, Montreal, QC, Canada H3C 0T1

EDUCATION

BSc, Major Statistics and Computer Science *Sept 2017 - May 2020(expected)*
McGill University, Montreal, QC, Canada *CGPA: 3.86/4.00*
School of Computer Science

BSc, Major Dietetics *Sept 2015 - June 2017(transfer)*
McGill University, Ste-Anne-de-Bellevue, QC, Canada
School of Human Nutrition

RESEARCH EXPERIENCE

Attention on Meta-embedding *Sept 2019 - Present*
Undergraduate Research Assistant *Supervisor: Jackie C.K. Cheung*

- Aim at improving the interpretability of contextualized embedding models by integrating diverse static pre-trained word embedding systems and combine them into a single representation depending on tasks and context via attention mechanism.

Sampling Methods for Training Word Embeddings *May 2019 - Aug 2019*
Undergraduate Research Assistant *Supervisor: Jackie C.K. Cheung*

- Investigated, derived and implemented three novel sampling based machine learning algorithms to learn word representations that capture linguistic features of words on 5.4 million tokens corpus using Pytorch.
- Substantially improved the efficiency, scalability and stability of existing word embedding training algorithms.

INDUSTRY EXPERIENCE

Data Analyst Intern *May 2018 - Aug 2018*
China United Network Communications Group

- Predicted the loss of customers up to 97% of accuracy on held out test sets by training SVM, XGBoost and Neural Network etc. models using Scikit-learn library.
- Applied statistic and machine learning knowledge to efficiently analyze and select user usage features of high correlation among more than 300 raw features using Pandas library.
- Enhanced data visualization using Seaborn and Matplotlib libraries.

PROJECTS

Semi-supervised learning of Tweets Sentiment Analysis During Californian Campfire
Natual Language Processing Course Project *Fall 2018*

- Scraped over 5,000 raw tweets data related to the 2018 Californian Campfire using Twitter API.
- Well-documented and published the clean dataset for the community interested in low resource natural disaster corpus research.
- Implemented semi-supervised learning algorithms that overcome the low resource caveat to analyze the sentiment of the population towards 2018 Californian Campfire over time.

Google Quick! Draw! Hand-drawn picture recognition
Applied Machine Learning Project *Fall 2018*

- Implemented CNN models to classify hand-drawn pictures of 31 classes.
- Achieved 78.2% of accuracy comparing to baseline of 3% accuracy.
- Ranked top 15 among 50 groups in class.

Smart City IoT: Pothole Detection
McHack 24hrs Hackathon, Montreal, QC, Canada *February 2018*

- Awarded the Telus IoT Prize
- Built a web app prototype to improve the efficiency of collecting road construction information in Montreal by providing simulating crowd source data aggregation from sensors in individual cars.

TECHNICAL STRENGTHS

Programming Languages	Python(proficient), Java, Matlab, R, Bash
Utilities & Tools	LaTeX, Git
Software	Pytorch, scikit-learn, matplotlib, pandas
Natural Languages	Mandarin, Cantonese, English

SELECTED COURSE WORK

Computer Science	**Mathematics and Statistics**
Algorithms and Data Structures	Probability & Statistics
Algorithm Design	Mathematics in Machine Learning
Numerical Computing	Regression and analysis of variance
Database Systems	Sampling Theory
Applied Machine Learning	Linear algebra
Natural Language Processing	Abstract Algebra
Probabilistic Graphical Models	Analysis

AWARDS

- Jacqueline Johnson Desoer Science Undergraduate Research Award *May 2019*
- Eliza M Jones in course scholarship *2017*
- Dean's Honour List *2017*