# Pengfei Murphy **He**

*4700 Brooklyn Ave NE, Seattle, WA*

425-772-7623 | hepengfe@uw.edu | feipenghe | hepengfe


## Education

**University of Washington** *Seattle, WA*

B.S. in Applied Computational & Mathematical Sciences, Data Science Track. / Minor Linguistics *June. 2018 - Expected June. 2021*

## Skills

**Programming** Python, Java, PyTorch, Numpy, Spark, MapReduce, C, SQL, Linux,
**Mathematics** Optimization, Linear Programming, Probability and Statistics, Numerical Analysis
**Languages** English, Chinese

## Project

**Sentence Corruption Classifier/Generator** *Seattle, WA*

Personal Project *Summer 2020*

- Reduced the text noise by using BPE to transform text into tokens
- Implemented specialized collate function transforming sentences to sequences with uniform batch size for generator model
- Built a pipeline of RNN/LSTM/GRU to classify corrupted sentence and seq2seq model to generate corrupted sentence using PyTorch
- Wrote Linux scripts and implemented DataParallel to run experiments on multiple GPUs

**Movie Recommendation System** *Seattle, WA*

CSE546 Project *Spring 2020*

- Implemented baseline model, average movie rating model and SVD model
- Derived the given non-convex optimization function into an implementable algorithm
- Designed and developed alternation minimization algorithm based on functions derived from the optimization problem

**Dealing with Intra-class Imbalance Using Constructive Samples Based on X-means** *Seattle, WA*

CSE547 Project *Spring 2020*

- Applied X-means algorithm on the representation from last layer of classifier model to compute the significant number of clusters and cluster centers
- Implemented the efficient utility methods, squared norm distances computing in Numpy to measure the closeness to the cluster centers of data per label
- Added a closeness threshold argument to measure the significance of data per label and used its value to reweight the training speed

**Survey of Emergent Compositional and Discrete Signal in Autoencoder Settings** *Seattle, WA*

ACMS Honor thesis *Spring 2020*

- Implemented additional cross entropy loss function for existing experiments
- Adjusted the model output and the target to comply the loss function
- Analyzed clustering results of the intermediate layer under the new training conditions and discovered symmetry phenomenon for min/max functions.

## Courses

**Optimization** MATH407 Linear Optimization, MATH164 Optimization(UCLA)
**Machine Learning** STAT548 Machine Learning for Big Data, CSE546 Machine Learning
**Deep Learning** CSE599G1 Deep Learning, CSE599I Generative Model
**Natural Language Processing** CSE517 Natural Language Processing, CSE599D1 Multilingual NLP Seminar, LING572 Statistical NLP
**Algorithm** CSE542 Reinforcement Learning, CSE573 Artificial Intelligence, CSE521 Advanced Algorithms
**Numerical Analysis** AMATH352 Linear Algorithm&Numerical Analysis, AMATH383 Mathematical Modeling

## Extracurricular Activity

**NeurIPS Conference** *Vancouver, BC*

Workshop Volunteer *Dec. 2019*

- Helped organize workshop and maintain the order of meeting
- Reported rising problems and discussed possible solutions with the manager effectively

## Honors & Awards

2018~ **Dean's List**, Undergraduate academic scholarship over six quarters *Seattle, WA*
2020 **ACMS Honor Student**, Departmental Honor *Seattle, WA*