

# Weizhe Yuan

weizhey@andrew.cmu.edu | (412)-721-5313 | <https://yyy-apple.github.io/>

## EDUCATION

---

### Carnegie Mellon University

*Master of Computational Data Science*

Pittsburgh, PA

*Aug. 2019 – May 2021*

### Wuhan University

*Bachelor of Engineering in Computer Science*

Wuhan, China

*Feb. 2017 – Jun. 2019*

### Wuhan University

*Bachelor of Engineering in Civil Engineering (Rank: 1/168)*

Wuhan, China

*Sep. 2015 – Jun. 2019*

## PRE-PRINTED

---

Can We Automate Scientific Reviewing? Weizhe Yuan, Pengfei Liu, Graham Neubig, arXiv 2020

## RELEVANT COURSES

---

**Machine Learning:** Machine Learning, Deep Reinforcement Learning and Control, Convex Optimization

**Natural Language Processing:** Algorithms for NLP, Neural Networks for NLP

**Computer Science:** Data Structure, Algorithms, Computer Systems, Cloud Computing, Operating Systems

## RESEARCH EXPERIENCE

---

### Capstone Project: Can We Automate Scientific Reviewing?

Fall 2020

*Advisors: Dr. Pengfei Liu, Prof. Graham Neubig*

*Carnegie Mellon University*

- **Dataset:** (1) Collected a dataset consisting of ICLR papers and NIPS papers together with their aligned reviews. (2) Defined an aspect typology that contains eight aspects for review structuralization. (3) Set up an interactive online annotation platform for crowd workers to annotate aspect information in reviews. (4) Used partial annotated reviews to train a sequence labeling model and annotated the whole dataset.
- **Evaluation:** (1) Summarized five most frequently mentioned desiderata for a good review. (2) Took the first step towards review evaluation by defining seven quantifiable metrics that can measure how well a review realizes those desiderata.
- **Model:** (1) Decomposed review generation into extractive and abstractive stages. (2) Investigated three different unsupervised summarization strategies in the extractive stage and two paradigms in the abstractive stage. (3) Extensively evaluated generated reviews from six systems as well as reference reviews and interpreted their relative merits based on our fine-grained evaluation metrics.
- **Fairness:** Proposed two ways to quantify bias in reviews, one measures absolute bias in a single system, the other measures relative bias between two systems.
- **Online Demo:** Designed an online system for automatic review generation: <http://review.nlpedia.ai/>.

## COURSE PROJECTS

---

### Deep RL and Control *Lecturer: Katerina Fragkiadaki*

Carnegie Mellon University | Fall 2020

- Implemented Twin Delayed Deep Deterministic Policy Gradients (TD3) and used Model-Based Policy Optimization with probabilistic ensembles to help faster the learning process where interacting with the environment is expensive
- (1) Investigated four exploration methods for TD3 in a sparse-reward situation, which are action noise, time-correlated action noise, parameter space noise, random network distillation. (2) Analyzed their relative advantages and disadvantages.

### Neural Networks for NLP *Lecturer: Graham Neubig*

Carnegie Mellon University | Spring 2020

- Implemented a text classifier using Convolutional Neural Network.
- Re-implemented the state-of-the-art abstractive summarization system on CNN/DM.
- Explored the task of aspect-based abstractive summarization and designed a novel way of injecting aspect information into pre-trained sequence to sequence model: retrieve all neighbor entities of a given aspect word on knowledge graph, and give external reward when model generates those words.

### Algorithms for NLP *Lecturer: Yulia Tsvetkov*

Carnegie Mellon University | Fall 2019

- Implemented a trigram language model using Kneser-Ney smoothing and achieved efficient storage using bit-packing trick to fully use the 64 bit size in a machine.
- Built an array-based CKY parser for English and improved its performance by doing complex structural annotation for the training data set.
- Implemented a word alignment model for English-French translation using Hidden Markov Chain and Expectation Maximization.

#### Cloud Computing *Lecturer: Majd F.Sakr*

Carnegie Mellon University | Spring 2020

- Processed a large Wikipedia dataset using MapReduce programming model and analyzed topic trends.
- Designed a suitable policy to maintain the Quality of Service (QoS) of a web service and realized it through configuring and deploying Elastic Load Balancer and AutoScaling groups on AWS.
- Implemented strong and eventual consistency models for a distributed key-value store.
- Processed and analyzed a huge Twitter social graph with the PageRank algorithm using Spark.

### COMPETITION EXPERIENCE

---

#### Interdisciplinary Contest in Modeling: Climate's Impact on Fragility

Winter 2018

*Core member, Advisor: Prof. Xinqi Hu*

*Wuhan University*

- Implemented 4 models based on study demands, including Polynomial Regression Model, Wavelet Neural Network Model, Analytic Hierarchy Process Model, Probabilistic Risk Assessment Model of Flooding.
- Applied these models into examples to identify the relationship between climate and fragility of a country, and evaluated interventions.
- Replaced Polynomial Regression Model with Distributed Lag Nonlinear Model based on Robustness Analysis, and conducted sensitivity analysis.
- Proved the models' value in predicting the trend of climate change, foretelling a country's fragility, thus helping government reduce potential loss.
- Won Honorable Mention.

#### SAS Data Mining Champion of HSBC Cup: Promotional Activity Design

Fall 2016

*Core member, Advisor: Prof. Ning He*

*Wuhan University*

- Comprehensively pre-processed customers' deposit and withdrawal information, including programming for analyzing mass of data and transforming the original data into desired forms.
- Chose some applicable SAS templates, such as BASE and STAT, to carry out statistical analysis based on the processed data, while taking some factors, such as noisy data, into consideration.
- Proved that the best way to promote bank's new product was telemarketing; determined the optimal amount of credit for customers through analyzing deposit rate, borrowing rate and customers' information with time series and methods of machine learning.
- Mastered SAS; won the National Second Prize, ranked top 7%.

### HONORS & AWARDS

---

Jun. 2019	Outstanding Graduate
Jan. 2018, Dec. 2016	Title of Merit Student
Nov. 2017, Nov. 2016	National Scholarship, Top 1%

### TECHNICAL SKILLS

---

**Programming Languages:** Python, JAVA, C/C++, MATLAB, Scala, Javascript, Bash  
**Software and Frameworks:** Git, Maven, PyTorch, TensorFlow  
**Operating Systems:** Linux, Windows, MacOS  
**Languages:** Chinese (Native Speaker), English

### TEST SCORES

---

**TOEFL:** 110 (L: 30; R: 30; S: 23; W: 27)  
**GRE:** 330 (V: 162-91%; Q: 168-94%; A/W: 4.0-59%)