HUIYU CAI

Peking University ⋄ 5th Yiheyuan Rd., Beijing, China 100871
(+86) 13269719900 ⋄ hy_cai@pku.edu.cn

EDUCATION

Peking University (PKU) *Sep. 2017 - Present*
School of Electronic Engineering and Computer Science (EECS) Overall GPA: 3.59/4
Ranking: 1/62 in Department of Machine Intelligence, 2019–2020
Core Courses (Grades): Signal and Systems (99), Machine Perception Lab (98), Mathematical Foundations for the Information Age (96), Deep Learning (93), Introduction to Pattern Recognition (93), Machine Learning (92), Algorithm Design and Analysis (87), Computer Net And Web Technology (87), Data Structure and Algorithms (87), Introduction to Parallel and Distributed Computing (86.5), Introduction to Computer Systems (85), Algorithms in Game AI (85), Information Theory

Peking University (PKU) *Sep. 2016 - Jun. 2017*
College of Chemistry and Molecular Engineering (CCME) Overall GPA: 3.68/4
Ranking: Top 5% in CCME
Core Courses (Grades): General Chemistry (93), Organic Chemistry (95.5), Quantitative Chemical Analysis (95), General Chemistry Lab (89), Comprehensive Analytical Chemistry Lab (89)

TECHNICAL STRENGTHS

Computer Languages	C/C++ (gdb), Assembly Language, Python (pdb), Java, R, JavaScript, Bash
Computing Frameworks	Spark, MapReduce, OpenMP, MPI
Python Libraries	Re, NumPy, Sci-kit Learn, Pandas, Matplotlib, Seaborn, Sympy, TensorFlow, Pytorch, BeautifulSoup, Selenium, RDKit, ScanPy
English Fluency	TOEFL iBT 114 (Reading 30, Listening 30, Speaking 27, Writing 27) GRE Verbal 157, Quantitative 170, Analytical Writing 3.5
Mathematics	Mathematical Analysis, Set and Graph Theory, Probability Theory and Statistics, Linear Algebra, Stochastic Processes, Convex Analysis and Optimization Methods, Machine Learning Theory

RESEARCH INTERESTS

Graph Neural Networks, Representation Learning, Bioinformatics, Drug discovery, Natural Language Processing, Music Processing/Generation

RESEARCH EXPERIENCE

Unsupervised scRNA-seq Data Mining via Embedded Topic Model Jan. 2020 – Oct. 2020
Submitted to RECOMB 2020 (First Author)

- Designed, tuned and benchmarked several models for representation learning on scRNA-seq data, including a joint graph-community learning model and an embedded topic model with batch correction.
- Practiced various scRNA-seq preprocessing and visualization techniques.
- Formed a comprehensive understanding of scRNA-seq data mining.

Robustness Study of Neural Code Comment Generation Systems Jun. 2019 – Aug. 2019
Submitted to ICSE 2020 (Second Author)

- Designed and implemented multiple ways of attacking a neural code comment generation model while fully preserving the syntactic correctness and semantic integrity of the code blocks.

· Conducted human evaluation study on adversarial-trained and vanilla models, proving adversarial training can significantly improve the robustness of the seq2seq model.

Multimodal Sarcasm Detection in Twitter Apr. 2018 – Oct. 2018
Accepted by ACL 2019 (Second Author)

· Co-developed and implemented a hierarchical fusion model for multimodal (image + text + image attributes generated from image) sarcasm detection. This is the first approach of deep representation fusion for Twitter sarcasm detection.
· Labeled and cleaned noisy Twitter data. This dataset is now publicly available.[1]

TEMPO and Its Derivatives in Organic Redox-Flow Batteries Apr. 2017 – Oct. 2017
Accepted by University Chemistry, 2017 (Third Author)

· Read, categorized and summarized previous literature on inorganic redox-flow batteries (RFBs).
· Compared inorganic RFBs to organic ones, highlighting the advantages of the latter.

PROJECTS

Music Source Separation: Theory and Applications Apr. 2020 – Jun. 2020
Class project, Machine Perception Lab (Individual project)

· Reviewed and benchmarked six influential algorithms for music source separation.
· Used the SOTA method, Demucs, to remix two songs with similar chord progressions, by extracting the voice of one song and pasting it to the extracted background music of the other song.

Raiden Game Implementation in Java[2] Jan. 2020 – Jun. 2020
Class project, Java Programming (First Author)

· Designed and implemented (with graceful object-oriented programming) a vertical shooting game similar to Raiden in Java from scratch, with delicate visual & sound effects.
· Led and coordinated my group via GitHub.

Fine-grained Face Manipulation via DLGAN[3] Oct. 2019 – Jan. 2020
Class project, Introduction to Artificial Intelligence (Second Author)

· Co-built a GUI program which takes a photo of the user, then manipulates its face using Disentangled Label-specific GAN (DLGAN).
· Implemented network modules between server (hosts DLGAN model) and client (requests DLGAN service).

Bird Sound Classification Mar. 2019 – Jun. 2019
Class project, Algorithm Design and Analysis (First Author)

· Led a three-member group to build an end-to-end model for bird sound classification.
· Proposed and tuned a convolution network specifically for bird sound spectrogram classification on a subset of *Bird-CLEF 2016*. Implemented multiple data augmentation techniques to boost the robustness of the model.

TEACHING EXPERIENCE

Teaching Assistant, Introduction to Computer Systems Sep. 2019 - Jan. 2020

[1] https://github.com/headacheboy/data-of-multimodal-sarcasm-detection
[2] https://github.com/hui2000ji/RaidenGame
[3] https://github.com/sunyaofeng8/AI-Intro

- Organized a 2-hour seminar per week where students present class reviews which I then comment and supplement to deepen their understanding of course materials.
- Answered student questions and guided them in completing lab assignments.
- Graded assignments and exams.

AWARDS AND HONORS

Leo Koguan Scholarship (1 out of 150), 2017
Leo Koguan Scholarship (1 out of 62), 2020
Founder Scholarship (2 out of 61), 2019
Second Prize, ACM-ICPC PKU campus competition, 2018
First Prize, 29th Chinese Chemistry Olympiad (Preliminary), 2015
First Prize, 28th Chinese Chemistry Olympiad (Preliminary), 2014

EXTRA-CURRICULAR ACTIVITIES

Student Acappella Club, Peking University Oct 2016 - present
President of Club Council
Deputy director of Department of Music Research and Training
Music director of Paca Vocal Group

Standing Committee of EECS Student Representative Assembly Sep 2017 - present
Member