DONG
Junjie

📞 (+86)13989565788   ✉ *jd445@qq.com* 🖥 *https://junjie.102514.xyz/*

🎓 EDUCATION

Dalian University of Technology **985**	**Software of Engineering**	**MPhil**

- **Average Grade:** 91.1   • **Research Direction:** Data Mining, Machine Learning, AI4SCI
- **Honors:** First-class Graduate Scholarship (twice)

Dalian University of Technology **985** & *University of Leicester*	**Chemistry**	**Bachelor**

- **Average Grade:** 91.9   • **Ranking:** Top 3%   • **Interdisciplinary Electives:** Java, Data Structures, Data Mining, etc.
- **Honors:** Liaoning Province Outstanding Graduate, Academic Excellence Scholarship (three times), Outstanding Merit Student (twice), Outstanding League Cadre, Outstanding League Member

👥 INTERNSHIP EXPERIENCE

City University of Hong Kong   📍 Hong Kong   **2023.9 ~ 2024.9**

Computational Department   *Research Assistant*   *Research Topic: Tracking the Evolution of Academic Research through Language Models*

- Data Collection: Python scraping of ArXiv papers from 2008 to 2022;
- Data Cleaning: Cleaning LaTeX and PDFs to obtain high-quality, clean academic corpora;
- Model Construction: Pre-training from scratch and fine-tuning to obtain a series of BERT models on a yearly basis;
- Model Analysis: Constructing different proxy and probe tasks (Link Prediction, etc.) using PyTorch to explore the semantic changes of language models over time.

Zhejiang NHU Co., Ltd.   📍 Zhejiang   **2020.7 ~ 2020.8**

Laboratory   *Researcher*   *Research Project: Purification of Acetone through Quaternary Simulated Moving Bed*

- Platform Setup: Building pilot and intermediate scale experimental platforms and devices;
- Experiment Execution: Testing various columns for their separation capabilities of the target substance;
- Analysis and Recording: Using HPLC to test the purity of the target substance and completing experimental records.

University of Leicester   📍 UK   **2019.7 ~ 2019.8**

Summer School   *Participant*   *Programme: Chemistry Experiments*

- Conducted hands-on chemistry experiments under the guidance of experienced instructors, enhancing practical laboratory skills;
- Created an academic poster to summarize and present experimental findings effectively;
- Delivered a presentation on the poster and was awarded the **Best Presentation Award**.

Harbin Institute of Technology   📍 China   **2020.7**

Summer School   *Participant*   *Programme: Electrochemistry and the Era of New Energy*

Jinkai Life Technology Co., Ltd.   📍 China   **2021.11 ~ 2021.12**

Training Programme   *Participant*   *Focus: Modern Biotechnological Techniques*

⚗ RESEARCH

Interpretable Sequence Clustering

Published in Information Sciences (CAS Zone 1 Top, CCF B Class)   *First Author*

- Preliminary Investigation: Literature collection and reading, surveying the latest achievements in sequence clustering and interpretable clustering;
- Experimental Verification: Proposing for the first time an interpretable discrete sequence clustering structure, constructing an interpretable clustering tree using relative risk index and Boost method, achieving a leading level in sequence clustering algorithms under interpretable conditions;
- Paper Writing: Completing data collection and paper writing.

This project has spawned a National Natural Science Foundation of China General Project (funded 500,000 RMB).

Hamming Encoder: Mining Discriminative k-mers for Discrete Sequence Classification.

Data Mining and Knowledge Discovery (Major Revision)   *First Author*

- Preliminary Investigation: Literature collection and reading, surveying binary interpretable neural networks;
- Experimental Verification: Implementing interpretable binary functions and training code using PyTorch, using sklearn to validate pattern discriminative power with classification algorithms, results show significant lead in pattern quality;

- **Paper Writing:** Proving the equivalence between CNN kernels and substrings, constructing a substring and feature distance function based on Hamming distance, completing data collection and paper writing.

General Ultra-Long Read Nucleic Acid Sequence Cutting Algorithm
Collaboration with Wang Pengfei's Research Group at Shanghai Jiao Tong University

- **Feasibility Verification:** Investigating the cutting feasibility of disordered linked gene sequences;
- **Preliminary Experiments:** Using greedy algorithms to find the best reference sequence combination does not yield accurate results;
- **Later Optimization:** Re-formulating the problem as a word segmentation issue in NLP, constructing an LSTM-CRF gene cutting model, and for the first time achieving short segment cutting of third-generation long sequence sequencing results.

♜ PROFESSIONAL SKILLS

- **Languages:** English, Cantonese
- **Skills:** C-1 Driving License
- **Programming Languages:** Proficient in Python, familiar with C++, Java, and other common languages.
- **Data Processing:** Familiar with NLTK, spaCy, and other natural language data cleaning tools.
- **Algorithm Tools:** Familiar with common classification (decision trees, random forests, neural networks, etc.) and clustering (k-means, HDBSCAN, etc.) algorithms, as well as the use of sklearn, understanding of common dimensionality reduction and visualization techniques (PCA, t-SNE, etc.), and understanding of HuggingFace's series of toolkits such as transformers for pre-trained natural language models.
- **Operating Systems:** Proficient in Linux and Windows operating systems.

▁▃▅ COMMUNITY SERVICES

President of Career Planning and Employment Association　　　　　**2020.9 ~2021.7**
Organized major events such as the "Northeast Five Universities Joint Career Fair" and the Resume Competition.

Secretary of the Youth League Branch, Class of Applied Chemistry 1801　　　　**2019.9 ~2022.7**
Organized "Three Meetings and One Lesson" activities and led the branch to achieve the title of "Outstanding Youth League Branch."

Deputy Team Leader, Social Practice Team　　　　　**2020.11 ~2021.1**
Organized activities of "Interconnected Past and Future," earning the team the "First Prize in Social Practice Team of Dalian University of Technology."

Team Leader, Social Practice Team　　　　　**2019.11 ~2020.1**
Organized and led a field social practice project "Zhejiang Helpers: In Search of Water," focusing on community water resources.

Deputy Head of Women's Affairs Department, Student Association　　　　**2019.1 ~2020.1**
Organized large-scale events such as the Freshman Orientation and the Chinese-Western Cultural Arts Gala.

Youth League Construction Instructor, Youth League Guidance Team　　　　**2020.9 ~2022.6**
Guided branches to achieve titles such as "Five-Star Smart Youth League Branch" and "Outstanding Youth League Branch."

🏆 HONORS AND AWARDS

- China Mathematical Modeling Competition **(First Prize)** 2018
- Mathematical Contest in Modeling (MCM) **(Second Prize)** 2018
- **National Scholarship** for Undergraduates 2019
- National College Students Digital Media Technology and Creativity Competition **(Second Prize)** 2020
- **National Scholarship** for Undergraduates 2020
- **National Scholarship** for Postgraduate 2022