LIN, KUO

(+86) 159 8900 3052 ◇ 1155152472@link.cuhk.edu.hk

EDUCATION

The Chinese University of Hong Kong *2020.9 - 2021.11*
M.Sc. in Computer Science *GPA 3.625/4*

Sun Yat-sen University *2016.9 - 2020.6*
B.Eng. in Computer Science *GPA 3.7/4*

RESEARCH INTERESTS

Computational Social Science, Public Opinions, Social Networks, Natural Language Processing

RESEARCH EXPERIENCE (SOCIAL SCIENCE RELATED)

Topic modelling: Based on People's Proposal to the government
Research Assistant, Supervised by Prof. XIA Ying *2021.11-2021.12*

- Responsible for data cleaning and analysis. Processed People's Proposal data for the period 2009-2019.
- Use LDA to build a topic model and extract the themes of the text by year and month respectively. Make predictions about the theme of each text and statistical analysis of individual topics.

Correlation Analysis: Based on News and Proposals from members of the Legislative Council of Hong Kong
Research Assistant, Supervised by Prof. XIA Ying *2022.1-2021.3*

- Responsible for data transformation, analysis and modeling.
- Data cleaning and tokenizing. Processed the data of News and Proposals in the period of 15 years.
- Build Language model using Word2vec and extract the keywords of the news.
- Analyze the word frequency of keywords in the news on a monthly basis, do correlation analysis with legislative council proposals in the following months. According to the party affiliation of the member, analyze the correlation between the member's proposal and the news event.

Analysis of corporate social accounts
Research Assistant, Supervised by Prof. XIA Ying *2022.7 – Present*

- Responsible for data collection, collation, cleaning and analysis
- Analyze the vocabulary of social accounts of large enterprises. Analyze the differences in vocabulary before and after significant political events.

RESEARCH EXPERIENCE (OTHERS)

Design and Implementation of K-Nearest Neighbor Algorithm Based on Homomorphic Encryption
Undergraduate Research Project *2019.11 – 2020.5*

- Based on the existing K-nearest neighbor algorithm using traditional encryption methods, use the method of homomorphic encryption to optimize the existing algorithm.
- Complete and evaluate the homomorphic encryption-based K-nearest neighbor algorithm with the help of the homomorphic encryption library HElib.

Data mining based on Covid-19 related scientific literature
Research Project of Master's course *2020.11 – 2020.12*

- Use NLP algorithms such as vectorization and PCA for data preprocessing, based on the literature dataset Cord-19 provided by the US government.
- Extract document features and use clustering methods such as K-means++ to cluster documents with similar topics.
- Use T-sne algorithm for dimensionality reduction and data visualization

WORK EXPERIENCE

Miaozhen Information Technology Co., Ltd. *2021.4 - Present*
Algorithm researcher of Natural Language Processing

MiningLamp Technology *2022.11 - 2021.4*
Algorithm engineer of Natural Language Processing

MiningLamp Technology *2021.7 - 2021.11*
Intern

PROJECT AND ENGINEERING EXPERIENCE (IN WORK)

New-word discovery and Phrase Mining
Main developer *2021.11 – 2022.1*

- Divide the text into segments of different lengths, and use statistical information to calculate the richness of information on both sides of the segment and the tightness of the interior of the segment, and determine whether the segment can form a new word or phrase.
- It is used to optimize the recall results of text retrieval and improve the relevance. The patent has been written and passed the preliminary examination
- It is also used for mining social media data to discover current new products, new trends and new topics.

Post-Search Recommender System in Smart Search
Main developer *2022.3 – 2022.4*

- Perform entity recognition on a large-scale text database, use the results of entity recognition to segment words, and train a word embedding model.
- According to the user's query request, word segmentation and calculation of similar entities and phrases based on language model are used for recommendation. The process has been written as a patent and passed the preliminary examination

Text Classification based on TextCNN and BERT
Main developer *2022.5 – 2022.6*

- Extract text features using pre-trained models, and train text classification models using convolutional neural networks (TextCNN) and BERT models.
- Models are trained separately according to different social media platforms. A model is trained to distinguish marketing ads from normal text.
- Mainly used for text classification and label extraction of social media data.

Named Entity Recognization in different area
Developer *2022.6 – Present*

- Train entity recognition models on corpora in different fields, including automobiles, food and beverages, beauty, etc.
- Deploy models on the K8S platform and make predictions on large-scale data. The scale of data is tens of millions.

SKILLS

Language	Mandarin (Native), English(TOFEL 104, GRE 321), Cantonese (Simple usage)
Coding	Python, C/C++, R, SQL, Shell, LaTeX
Platform	Pytorch, Tensorflow, Scikit-Learn, Kubernetes, Elastic-Search, Docker

AWARDS AND RECOGNITIONS

Dean List Scholarship 2020-21

Distinguished Academic Performance Scholarship 2020-21

Honorable Mentioned in The Mathematical Contest in Modeling (MCM/ICM) of 2019

Second Prize in 2019 Sun Yat-sen University Computer Programming Competition (top 10%)

Second Prize in 2018 Sun Yat-sen University Computer Programming Competition (top 10%)

Second Prize in 2017 Sun Yat-sen University Computer Novice Programming Competition (top 12%)