YONGSHEN HE

12132901@mail.sustech.edu.cn | +86 177 2470 2474 | 1088 Xueyuan Avenue, Shenzhen, 518055 | megasunny
Research interest: text mining, scientific mobility, computational social science, applied statistics.

EDUCATION

M. Sc., Statistics & Data Science, **Southern University of Science and Technology(SUST)** — **Shenzhen, China 2021.9 - ongoing**
GPA: 3.45/4.5

B. Sc. - Statistics, **Shenzhen University** — **Shenzhen, China 2017.9 - 2021.6**
GPA: **4.15**/4.5, ranking **1**/45.

PUBLICATIONS

Yongshen He, Yurui Huang, Chaolin Tian, Shibing Xiang, Yifang Ma*. Neural Embeddings of Scientific Mobility Reveal the Stratification of Institutions in China. Information Processing & Management. — **under review**

Yurui Huang, Jialong Guo, Chaolin Tian, Shibing Xiang, **Yongshen He**, Yifang Ma*. Domestic Brain Circulation in China. Scientometrics. — **under review**

RESEARCH EXPERIENCE

Semantic-Based Study on the Mobility of Researchers and the Shift of Research Interests — **Shenzhen, China 2022.7 - on going**
Master thesis, supervised by assistant professor Yifang Ma

- Implemented topic classification task combining document embedding techniques of TF-IDF, word2vec and SPECTER with classifier of logistics regression, random forest and supporting vector machine.
- Embedded institutions into a low dimensional space combining word2vec and UMAP based on the scientists' moving trajectories.
- Conducted an empirical analysis of the effect of scientific mobility on the shift in research interests via propensity score weighting.

Short Text Clustering Based on Word2vec Model — **Shenzhen, China 2020.11 - 2022.5**
Bachelor thesis, Supervised by associate professor Songchen Li

- Compared different document representation methods of word2vec embeddings clustering methods in clustering titles of articles.

Health Information Recommendation Decision Based on Machine Learning — **Shenzhen, China 2018.11 - 2020.5**
Leader of a funding project, Supervised by associate professor Yingjue Fang

- Designed neural networks of Fasttext and TextCNN to distinguish the health news from the news on other topics.

COURSE PROJECT

Statistical Graph Models on Community Detection — **Shenzhen, China 2022,2 - 2022.6**
Instructor: associate professor Qingyan Hu

- Detected communities in the binary network with the probability-base model.

High Dimensional Statistics on Square-root LASSO — **Shenzhen, China 2022,2 - 2022.6**
Instructor: associate professor Zeng Li

- Derive Conic Programming and CCD algorithm to solve the square-root LASSO.
- Implement simulate study to compare the rates of precise and recall of both LASSO and square-root LASSO.

Categorical Data Analysis on American National Election Survey — **Shenzhen, China**
Instructor: associate professor Xiyun Jiao — **2022,2 - 2022.6**

- construct binomial regressions of differnet link functions to estimate the effect on the vote in the presidential election.
- utilize AIC criteria to select of the vital factors of the vote.

HONOR

Outstanding Graduate in Shenzhen University — **2021**
First Prize Scholarship in Shenzhen University — **2018-2021**

TEACHING EXPERIENCE

Network Science and Computing, Teaching Assistant, SUST. — **2022.2 - 2022.6**
Bayesian Statistics, Teaching Assistant, SUST. — **2022.2 - 2022.6**

RESEARCH INTEREST, LANGUAGE AND SKILL

Languages: CET-6 532, Chinese (Mandarin, Cantonese).
Skills: Python, R, Latex.