(781) 539-3142
Meford, Massachusetts
jiaxing0218@gmail.com

Jiaxing He

linkedin.com/in/jiaxing-he-43b91523b/

Education

Master of Data Science GPA: 3.73/4.0 Sep. 2021 -now
Tufts University

Bachelor of Economic Statistic GPA: 3.43/4.0 Sep. 2016 -Jun. 2020
Guangdong University of Foreign Study

Skills

Tools and Languages	Python(Numpy,Pandas,Scipy), Spark, Pytorch, Java, Shell, Sk-learn, Git, LaTeX,
Quantitative Research	Machine Learning(cv2/pytorch-geometric/graph neural network/natural language processing/computer vision/traditional machine learning) , MySQL,Google Cloud Platform,SPSS
Financial Analysis skill	CFA level Ipassed,(straight A pass)

Publication

1. Chen, X., **He, J.**, Han, X., Liu, L. Efficient and Degree-Guided Graph Generation via Discrete Diffusion Modeling, International Conference on Machine Learning (ICML) 2023

Experience

Student Research Assistant (Pytorch, Variational Auto Encoder, Latent Variable Distribution). **Sep.2022 — Feb.2023**
Tufts University *Medford, Massachusetts*

- Generating graph (eg. Molecule, Network) by conducting Diffusion Model.Application: drug discovery, knowledge graph completion, etc.
- Writing code for the forward process in a discrete diffusion model. Including the sampling from the latent categorical distribution and defining the convergent state of the diffusion model.
- Running experiment in HPC to evaluate model performance in the standard dataset including caveman, grid, and ego, etc.(python/diffusion model/shell)
- Our model outperformed existing benchmarks in terms of both molecule legality and formation speed.(already accepted in ICML2023)

Data Analyst **May.2020 – Sep.2020**
China International Capital Corporation Limited *Guangzhou, China*

- Feature extraction, by using text embedding from the Roberta model. and then Take advantage of that feature and concatenate the feature and original numerical feature of stock (like 15 days average price, and buy-in amount)to get a vector to represent the stock feature. and then utilize LSTM(long-short-term memory) to do multi-target prediction.
- Data preprocessing by using NLTK and Pyspark to get the stem word and get rid of the useless word and punctuation.

Data Analyst/Intern of Financial Leasing Department **May.2019 – Sep.2019**
China Bank *Guangzhou, China*

- Customer credit rating classification by using decision tree and SVM.(technique:SVM/python/sk-learn/PCA/SQL)
- Using A/B test to testify the performance between different models.(technique:z-test/t-test/pandas/python)
- Assisted in leasing business plan negotiation, data collection, data visualization, etc.(excel/power BI/python)

Projects

Computer Vision project/ image generation model(GAN) **Jan.2022 — May.2022**
Tufts University *Medford, Massachusetts*

- The project's objective is to convert landscape images into artistic styles inspired by renowned artists like Van Gogh and Monet. Additionally, the project aims to explore how Generative Adversarial Network (GAN) models capture data patterns and modify the model to improve its performance.
- participating in developing code to modify both the discriminator and generator components of the GAN model.(relative technique: Pytorch/ neural network/ sampler)

Computer Vision project/ image classification **Sep.2022 — Dec.2022**
Tufts University *Medford, Massachusetts*

- Image data preprocessing by using Gaussian filter/median filter, and image resize,etc.(technique: python/cv2/sk-learn)
- By using ensemble learning to do image classification on the FMNIST dataset. and check the performance and do visualization. (Technique:cv2, python,scikit-learn, SVM, decision tree, Neural network)

(781) 539-3142
Meford, Massachusetts
jiaxing0218@gmail.com

Jiaxing He

linkedin.com/in/jiaxing-he-43b91523b/

Natural Language Process course project **Sep.2021 — Dec.2021**
Guangdong University of Foreign Study *Medford, state*

- The primary goal of this project is to categorize customer reviews into distinct types and perform semantic analysis on the collected data.(technique:re/python/pytorch/multi-object classification)
- Writing code about Text Data processing, LSTM/GRU Modeling, and Model tuning.(Latex, recurrent neural network, text data preprocessing, SQL, shell, Google Cloud Platform)

Related Course

Machine Learning and Data Mining; Algorithm; Statistical Pattern Recognition; Bayesian Deep Learning; Deep Learning; Statistic; Big Data; C ; Multiple Variable Statistic; Database System;