Han ZHOU

hanzhou925@gmail.com | (+44) 07536403853 | Homepage: https://hanzhou925.github.io/

Education

MSc. in Language Science in Neuroscience, Language & Communication, University College London Sept.2022

B.A. in German, Shanghai Jiao Tong University (Grade: 88/100) | Minor: AI Sept.2018-Jun. 2022

Courses: Natural Language Processing (90); Probability and Statistics (88); Introduction to Calculus (97); User-Centred Data Visualization; German Linguistics (94); Frontiers in Linguistics (98); AI Programming Framework; Introduction to AI

Awards: SJTU Scholarships (Top 20%); Third Prize of National English Competition for College Students (Top 5%); IBM Certificate of Data Science

Work Experience

NLP Product Operation Intern, AI-LAB, ByteDance Technology Co., LTD Feb. 2022-Jun. 2022

- Led the TikTok Cross-Language Assessment (CLA) project in German, cooperated with other language teams, and analysed data from over 1000 problematic captioned videos, thus determining factors affecting users' satisfaction with the live subtitle function on TikTok
- Organised interviews with 5 groups of German speakers, researched their habits of using subtitles watching videos in foreign languages on TikTok, locating users' pain points and priorities to improve TikTok's subtitle function
- Analysed over 1500 corpus items corrected by Feishu's auto-correct function from 5 dimensions, and compared the statistical results visually with Grammarly, thus producing structured reports on the existing problems and providing feasible suggestions in communication with R&D to upgrade Feishu

TTS Linguistic Data Intern, Ximalaya Technology Co., LTD (top Internet company in AI voice with over 2000 employees) Jun. 2021-Sept. 2021

- Led the high-prioritized program of sentiment annotation in the data team, analysed sentimental polarities and 8 categories of emotions in dialogues, thus establishing the sentiment dataset with over 24,000 data to develop Text-To-Speech (TTS) within a deadline of 10 days
- Proofread over 70,000 polyphone/homograph Mandarin data, and added over 3000 missing data from BCC corpus after analysing the polyphone coverage with Python, thus expanding the coverage of the dataset by 2% and improving the accuracy of machine learning through cooperation with R&D
- Evaluated and fixed over 1000 bad cases of AI pronunciation in Mandarin based on prosody, phonetics and sentiment, regularised and cleaned the newly added Mandarin data with Python, enhancing the performance of speech synthesis

Projects

Data Science & Analytics Virtual Intern Program, BCG Jul. 2022-Aug. 2022

- Applied CRISP-DM Model to diagnose the source of SME customers for PowerCo., and put forward the hypothesis that churn is driven by customers' price sensitivity, locating the goal of modelling
- Performed exploratory data analysis based on over 15,000 data from clients, so that assisted the client PowerCo. in analysing how different factors like price sensitivity affected the churn rate of their customers
- Engineered the features after cleaning the datasets and built a Random Forest Classification model to predict which customer subscribed at PowerCo. is more likely to churn, reaching an accuracy of 90%

Image Restoration based on Linear-regression Analysis, SJTU

- Generated damaged image by adding noise masks to the original image, segmented the image into regions to apply Linear Regression Model in Sklearn and restore the image with a Cosine similarity of 0.998 with the original image

Feature Face Recognition (Eigenface) based on PCA, SJTU

- Processed face images from ORL database, applied covariance matrix to quantitatively describe "face difference" and create eigenface matrix, trained data with eigenface and average face and reached an accuracy of over 88% in prediction
- Transformed data into the eigenface space and inversely transformed it into the pixel space to rebuild human face

Research Experience

The Regulation of Relative Clause Types on Causality Expectation in Real-time Discourse Processing

Advisor: Professor Fuyun Wu May. 2021-Jul. 2022

- Compared and analysed over 10 recent papers, thus finding out the current research gap, thus expanding the experiment to Chinese to explore the effect of three relative clause (RC) types on prediction of causality in discourse
- Designed 30 sets of Chinese stimuli and 60 fillers, each including a matrix sentence with a neutral/causal/concessive relative clause, a connective clause, and programmed with Experiment Builder for the eye-tracking experiment
- Gathered 50 participants' data of eye movements, extracted the targeted indexes and cleaned data with R, built the Mixed-linear Regression Model based on 6000 datapoints to find significance and calculate the impact of RC types on the sentence processing during online reading

Other Information

Leader, Public Relations Department, Student International Organization Association Sept. 2018-Sept. 2020

- Contacted and invited several IO officials, organizing the Training Camp for IO Talents with an attendance of over 50
- Organised university-wide visiting programs to offices of IOs in Bangkok and Beijing and expanded the visit to Tokyo

Tech Skills: SQL; Python(sklearn/pandas/numpy); RStudio(lme4/lmerTest/ggplot2); GCP(BigQuery/Virtual Machine) Tableau

Language Skills: Chinese (Native); English (Fluent); German (Fluent); Sanskrit (basic)

Soft Skills: Interpersonal Communication; Adaption to multicultural atmosphere; Teamworking; Learning capacity; Multitasking