Sijia Ge

E-mail: sijiage007@gmail.com
Mobile: (720)5307370

Smiley Court, 1300 30th St.
Boulder, Colorado, 80303

EDUCATION

University of Colorado-Boulder, CO, USA Aug.2021-Now

M.S Computational Linguistics(CLASIC) 4.0/4.0 (Expected graduation:May.2023)

Nanjing Normal University, Jiangsu, China Sep.2016-Jun.2019

M.A Linguistics&Applied Linguistics (with Concentration on Computational Linguistics) 86.3/100(overall)

Shanxi University, Shanxi, China Sep.2012-Jul.2016

B.A Chinese Language and Literature 85.8/100(overall)

RELATED SKILLS

Technical: Python, Keras, Pytorch, Java, NLTK, Scikit-learn, Transformers, Tensorflow, SQL, Django, Flask, HTML, Axure RP, JavaScript, JQuery, Linux, Shell, SPSS, LaTex, Git, PlantUML, Neo4j, Tableau
Language: Mandarin Chinese(Native), English(Fluent)

PROFESSIONAL EXPERIENCE

Student Research Assistant
CLEAR lab, CU-Boulder *May.2022-present*
1.compute the similarity of pair of events which mapped to the wikidata use the rule-based and machine learning algorithms (a small track of AIDA project funded by DARPA)
2.update the online annotation tool with more efficient and better user experience interface with front-end skills and Flask.

Beijing Lingosail Tech Co., Ltd.
Product Manager Intern, Beijing Mar.2021-June.2021
Mainly focus on a term extraction tool whose target users are the students and teachers who use CAT software.
Design the prototype of the new products including basic UI, function and interaction;Schedule the overall progress of development as well as promotion model;Test the function of new products with black box testing

COURSE&RESEARCH PROJECTS

Viz-Wiz Visual Question& Answer Challenge: Answer Visual Questions from People Who Are Blind
Course project,The University of Colorado-Boulder CSCI 5922 Deep Learning
- Adopted the feature map on the next to last layer of VGG-16 as the image features
- Adopted the the hidden state of the last layer of BERT as question features
- classify the concatenated feature into first 50000 frequent answers and got the score of 0.47

E-commerce online shopping website
Course group final project,The University of Colorado-Boulder CSCI 5448 Object-oriented design&analysis
- Simulate the fundamental function for a online shopping website
- Built on Django 2.0, adopted JQuery&JavaScript to interact with the users at the front end
- Adopted SQLite as the model and built-in admin module as the backstage manage system

Music Store Simulation
Course group project, The University of Colorado-Boulder CSCI 5448 Object-oriented design&analysis
- Simulate the operation for a music store, including order new items, sell items, buy items and so on
- Apply design pattern into the project, such as strategy, decorator, observer, singleton and etc.
- Based on Java and utilize Junit for unit test

Patronizing and Condescending Language Detection
Course group project, The University of Colorado-Boulder CSCI 5832&SemEval 2022 shared task 4(practice phase)
- Applied XLNet as a pre-trained model for the binary classification task and reached a f-score of 0.53, ranked #7 for the practice phase, which was higher than the Roberta baseline.
- Also applied Glove embeddings as features for the support machine vector model to reach a f-score of 0.38

Named Entity Recognition for Gene text

Course group project, The University of Colorado-Boulder CSCI 5832

- Implemented gene named entity recognition with crf++,crfsuite, Bi-LSTM-CRF and pre-trained language model(BioBERT), and reached entity based F-score of 62%, 62%, 77%, 85% respectively.
- utilized comet_ml as a tool for training visualization

A Joint Model of Automatic Sentence Segmentation and Lexical Analysis for Ancient Chinese

Research Assistant, Nanjing Normal University Sep.2018-Mar.2019

- Established a seq2seq model for sentence segmentation and lexical analysis based on Bi-LSTM-CRF and TensorFlow
- Transferred two separated tasks into one task, the F1-score of sentence segmentation, word segmentation, and POS reached 78.95,85.73% and 72.65% separately, with an average increase of 3.5%, 0.18%, and 0.35% respectively

Chinese Abstract Meaning Representation Corpus

Team Member, Nanjing Normal University Jul.2017-Mar.2018

- Counted the distribution of semantic roles for the predicate in 5000 sentences automatically
- Compared the data with semantic roles labeling based on Chinese PropBank to verify the efficiency of AMR for solving the conflicts between core and non-core roles, representation of multi-functional roles, and solution to dropped roles

The Platform for monitoring the popularity of Mandarin Chinese

Team Member, The Education Department of Jiangsu Province Dec.2016-Dec.2017

- Developed a web platform collaboratively based on spring framework for recording the nationwide result of the survey for Mandarin Chinese popularity
- Implemented the import and export data of Excel and audio files

SELECTED PUBLICATION

Paper

Ning Cheng, Bin Li, Liming Xiao, Changwei Xu, **Sijia Ge**, Xingyue Hao, Minxuan Feng. Integration of Automatic Sentence Segmentation and Lexical Analysis of Ancient Chinese based on BiLSTM-CRF Model. *Proceedings of LT4HALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages*. Marseille, France,2020:52-58.

Li Song, Yuan Wen, **Sijia Ge**, Bin Li, Junsheng Zhou, Weiguang Qu and Nianwen Xue. An Easier and Efficient Framework to Annotate Semantic Roles: Evidence from the Chinese AMR Corpus.*The 13th Workshop on Asian Language Resources on LREC 2018*. Miyazaki, Japan, May 07, 2018:29-35.

Patent

A Method and System of Automatic Lexical Analysis for Ancient Chinese. (the 3rd applicant).2019. No.CN201910085019.3

VOLUNTEER EXPERIENCE

Volunteer for the China National Conference on Computational Linguistics 2017(CCL 2017) **Oct.2017**

Volunteer for the Museum of Shanxi Province **Nov.2012-May.2013**

Incoming international peer mentor at CU-Boulder **2022-2023 academic year**

Incoming graduate school peer mentor at CU-Boulder **2022-2023 academic year**