XINYI ZHENG (CAROL ZHENG)

zxycarol@umich.edu (1)734-882-7055
www.linkedin.com/in/carolxinyizheng

EDUCATION

University of Michigan, Ann Arbor, MI Sept. 2017 - Apr. 2020
Bachelor of Science in Computer Science and Mathematical Science (Algorithm Track) Overall: 3.88 / 4.00 Major: 3.97 / 4.00
Coursework: Natural Language Processing(grad), Operating Systems, Machine Learning, Optimization Methods (grad), Computer Organization, Database Management Systems, Data Mining, Data Structure and Algorithms, Numerical Methods

Shanghai Jiao Tong University, Shanghai Sept. 2015 - Aug. 2020
Bachelor of Science in Electrical and Computer Engineering

SKILLS

Programming Languages: Java, Python, C/C++, Matlab, Html/CSS, MySQL
Framework: Tensorflow, Flask, CoreNLP, Mallet, ElasticSearch, NLTK
Tool: Git, Screen, LaTeX, Virtuoso

PROFESSIONAL EXPERIENCE

Research Assistant, Michigan Database Research Group, advised by Prof. H.V. Jagadish **Ann Arbor, MI**
Hybrid Question Answering System Apr. 2018 - Present

- Got 95% parsing accuracy on *ComplexQuestions* (Bao et al., 2016) by designing and implementing a novel heuristic algorithm for query graph based multi-constraint semantic parser
- Transferred natural language question into sub-queries for Open Knowledge Base with ElasticSearch, built SPARQL for RDF knowledge graph (Freebase)
- Designed and Implemented a heuristic algorithm for knowledge graph retrieval and search space pruning based on meta data of knowledge base
- Design a knowledge graph and text joint path reasoning algorithm for joint inference on Entity Linking, Relation Extraction and Information Retrieval for

Research Assistant, Language and Information Technologies Group, advised by Prof. Rada Mihalcea **Ann Arbor, MI**
Improving Active Interporetation of Disparate Alternatives through knowledge base construction (AIDA) Sept. 2018 - Present

- Implemented off-the-shelf methods to generate structured representation of assertions of documents, and integrated the system with computer vision team, having multimodal representations for hypothesis
- Enhance Knowledge Base by designing robust event extraction algorithm

Analytics for Learners as People (LEAP) https://midas.umich.edu/research/leap/ Mar. 2018 – Sept. 2018

- Collected 100K text and image data of student volunteers by constructing a twitter crawler and a stream for social media activities
- Conducted Sentiment analysis on text social media, and mine relationship between students' texts and mental state
- Used LDA topic model and LIWC lexicon to analyze social media data based on different criterions, build an ensemble classifier and get ~70% accuracy on depression state prediction based on social media information
- Used survey, location and academic information to build classifiers for students' mental health, got ~75% precision
- Designed a command-line Instagram crawler tool that downloads all photos relying on API calls and user authentication

Sinolink Securities, Shanghai Dec. 2016 - Feb. 2017
Software Engineer Intern

- Did data cleaning and wrangling with MySQL, including filtering useless data and inferring semi-structured data
- Constructed an automatically fake news detection and classification module in Public Opinion System by segmenting news texts, capturing keywords and analyzing sentiment in news
- Selected data and reviewed selected data for project manager based on clients' requirement

PROJECTS

School of Information (UMSI), Ann Arbor, MI Feb. 2018 – Jul. 2018
Understanding How Social Relationships Affect Information Diffusion, advised by Prof. David Jurgens

- Cleaned data on more than 200M social relationships, filtered mistakenly annotated data based on keyword extraction and verified with web crawler results
- Constructed time-series data visualization infrastructure supporting personalized confidence interval selection, hierarchy categorized layout with Bootstrap Algorithm, Matplotlib and Seaborn
- Trained classifiers on social relationships based on information diffusion, topics, reciprocality, time series features, etc to detect the difference of information diffusion in topic, word and character level between social relationships

Side Project, Ann Arbor, MI Dec. 2017 – Jan. 2018
Chinese news search engine with recommendation system based on JIEBA, Flask

- Got ~10K news documents by extracting URL, analyzing HTML and constructing web crawler based on XML and BeautifulSoup
- Constructed index system with SPIMI-INVERT algorithm, used TFIDF model to select news posts, and built recommendation system based on JIEBA text classification and keyword extraction
- Built a web search engine based on flask, supporting news searching and ranking selection

Industrial and Operation Engineering, University of Michigan, Ann Arbor, MI Jan. 2018 - Mar. 2018
Bundled Payment vs. Fee-For-Service: Impact on Active Surveillance of Low Risk Prostate Cancer, advised by Prof. Brian Denton

- Developed rationally assumption, objective function and constraints for a Hidden Markov Model(HMM) on Low-Risk Prostate Cancer treatments
- Simulated the HMM with Viterbi Algorithm, tuned the hyperparameters and optimized ~20% expenses for Prostate Cancer Patients according to JHU Prostate Cancer Database

AWARDS AND HONORS

3rd place, Bloomberg Coding Challenge, **University of Michigan, MI** Nov. 2017