BILLY XUANMING ZHANG

(+1) 646-243-4071 | xz2995@columbia.edu | Homepage

EDUCATION

Columbia University New York, USA
MS in Computer Science *Sep. 2021 – May. 2023 (expected)*

University of Nottingham Ningbo China (UNNC) Ningbo, Zhejiang, China
BS in Computer Science with Artificial Intelligence Sep. 2016 – Jun. 2020

- Overall GPA: 3.85/4.0
- Exchange Student at the **University of Texas**, Dallas (Aug. 2018 – May 2019; selected as top 5/100 students)
- Summer Student at the **University of California, Berkeley** (Jun. – Aug. 2018; Major 3.7/4.0)

PUBLICATIONS

1. Y. Zhao, RA. Schmidt, Y. Wang, **X. Zhang**, H. Feng "A Practical Approach to Forgetting in Description Logics with Nominals", Accepted, Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)
2. GO. Diaz, **X. Zhang**, V, Ng "Aspect-Based Sentiment Analysis as Fine-Grained Opinion Mining", Accepted, 12th Language Resources and Evaluation Conference (LREC 2020)
3. H. Chen*, C. Yang*, **X. Zhang**, Z. Liu, M. Sun, J. Jin "From Symbols to Embeddings: A Tale of Two Representations in Computational Social Science", Accepted, Journal of Social Computing
4. H. Chen, Z. Liu, M. Sun, J. Jin "Social Data Analysis during the COVID-19 Pandemic", Tsinghua University Press (composed the chapter entitled "Public Opinions: Group Polarization in the Debate Revolving around 'Wuhan Diary'")

RESEARCH EXPERIENCE

Columbia University (Natural Language Processing Group) New York, USA
Graduate Research Intern, Supervisor: Prof. Kathleen McKeown

SECAF: A Span-level Emotion Cause Analysis Framework Feb. 2022 -- May. 2022

- Implemented a general framework for span-level Emotion Cause Analysis (ECA).
- Fine-tuned BERT at the clause-level on a publicly available ECA dataset and adopted Bi-LSTM to capture the interactions between different clauses in a document for extracting the emotion causes.
- Adopted commonsense knowledge and verified its effects on emotion cause span extraction, emotion expression span extraction and emotion-cause span-pair extraction.

Generating Commonsense Enhanced Natural Language Explanations for Emotion Cause Analysis Sep. 2022 -- now

- Fine-tuned BART-base/large model on GLUCOSE dataset to make commonsense inferences on unseen stories.
- Adopted GPT-3 to generate natural language explanations as gold annotations to evaluate model performance on generating human-like explanations.

Graduate Research Intern, Supervisor: Prof. Zhou Yu

Spoken EduBot for Oral English Practice Sep. 2022 -- now

- Incorporated automatic speech recognition and text-to-speech components into the front-end user interfaces.
- Implemented a template-based chatbot by prompting the BlenderBot-3 with hand-crafted examples.
- Collected speech data from 500 users (~70 hours), engaging with the chatbot over the topics covered in the real IELTS spoken test.

Physical Activity Chatbot Oct. 2021 -- now

- Fine-tuned the Blenderbot on the intervention conversation dataset, collected from a real-world physical activity intervention program for women
- Implemented a three-step pipeline for pilot study data collection: 1) pre-survey (collect user contextual knowledge); 2) chatbot conversation (interaction between user and the trained chatbot model) and 3) post-survey (chatbot evaluation)
- Improved the chatbot response quality in terms of: 1) keywords detection for activity barriers/benefits and 2) generation model that can produce responses with less repetition, more adequate candidates and more database-enhanced (retrieval-based) answers

Tsinghua University (THUNLP Group) Beijing, China
Research Assistant, Supervisor: Prof. Zhiyuan Liu

Multilingual Country Image Detection: A Case Study of China Apr. 2021 -- Sep. 2021

- Collected Twitter tweets data generated from 2012 to 2021 via Twitter API and selected those related to China
- Programmed to auto-annotate the sentiment of each tweet towards China (containing Emoji) and used it as the training data

- Fine-tuned XLM model on auto-annotated sentiment dataset to obtain a multilingual sentiment classification model
- Inferred the sentiments of all tweets towards China using the fine-tuned XLM model, so as to analyze the country image of China reflected in different languages and country regions.

A Tale of Two Representations in Computational Social Science Jan. 2021 -- Jun. 2021

- Collected research articles published in 3 prestigious journals (Nature, Science, PNAS) and 3 top conferences (ACL, KDD and WWW) within recent 10 years and selected those related to Computational Social Science
- Categorized the selected papers into two schemes, namely symbol-based and embedding-based representation, according to the data representations of each paper
- Defined different domains of applications (e.g. Anthropology) and categories of tasks (e.g. Regression), respectively, adopting either symbol-based or embedding-based representations, and identified the form of data (i.e. text or network)
- Assisted with final manuscript that has been submitted and accepted as a journal paper in the Journal of Social Computing

Social Data Analysis during the COVID-19 Pandemic Jul. 2020 -- Feb. 2021

- Collected 12 million tweets related to COVID-19 pandemic and 3 million corresponding user data from Sina Weibo using Python Scrapy framework
- Annotated the stance of 4000+ tweets towards "Wuhan Diary" and used them as training data for automatic stance detection
- Fine-tuned BERT model on the annotated stance dataset to obtain St-BERT that can effectively detect the stance of a given tweet towards "Wuhan Diary" and outperform other baseline systems (e.g. Bi-LSTM based model)
- Inferred the stance of all tweets towards "Wuhan Diary" to analyze the group polarization phenomenon existed in online social network
- Composed the chapter "Public Opinions: Group Polarization in the Debate Revolving around 'Wuhan Diary'" in the book entitled "Social Data Analysis during the COVID-19 Pandemic"

UT Dallas (Human Language Technology Research Institute) Dallas, TX, USA

Researcher, Supervisor: Prof. Vincent Ng

Aspect Based Sentiment Analysis (ABSA) Sep. 2018 – Dec. 2018

- Developed a model to examine product reviews on e-commerce sites to help buyers make purchasing decisions
- Summarized all review info and trained the classifier to differentiate between subjective and objective feedback
- Implemented the existing polarity classification system

Opinion Target Based Sentiment Analysis Jun. 2019 – Sep. 2019

- Analyzed the fine-grained sentiment both the source of the sentiment as well as its propagation
- Contributed to literature review, experimental design proposal and drafted framework for the group's prototype
- Implemented selected baseline systems, assigned prototype modules, and evaluated scripts
- Assisted with final manuscript and accepted to the 2020 *Language Resources and Evaluation Conference*

Nanjing University, School of Artificial Intelligence Nanjing, China

Research Intern (remote), Supervisor: Prof. Yizheng Zhao

A Practical Approach to Forgetting in Description Logics with Nominals Jun. 2018 – Aug. 2018

- Aimed to test research group's forgetting and reasoning tools in order to fully assess their viability as back-end technology in Babylon's knowledge base interface for their ontology analysis and tracking
- Tested the reasoning methods and the accompanying tool to assess their viability as back-end technology in Babylon Health's knowledge base interface
- Built a bespoke ontology comparison and tracking system, allowing high-level automation in creating and maintaining Babylon's knowledge base with a time reduction by at least 20%
- Realized direct saving for the company in staff costs
- Accepted to the 2020 *Association for the Advancement of Artificial Intelligence (AAAI-2020)*

INTERNSHIP EXPERIENCE

Educational Testing Service AI Lab Princeton, NJ, USA

Collegiate Associate, Mentors: Rahul Divekar and Rutuja Ubale

Repair and Grounding in Conversations for Language Learning Jun. 2022 -- now

- Conducted literature reviews on Repair and Grounding (R&G) in conversations, defined a list of R&G types and identified gaps in the current task-oriented dialogue datasets for language learning.
- Designed a conversational task that is conducted between Native Speakers (NS) and Non-Native Speakers (NNS) to collect a new dataset that incorporates R&G elements.
- Implemented a web application (GrounDialogue) using Node.js, MySQL and AWS services for data collection on Amazon Mechanical Turk (AMTurk). GrounDialogue functions to pair the speakers (NS-NNS) and send/store both textual and spoken responses from the speakers.

- Devised an AMTurk data collection pipeline that encapsulates the consent form, conversational task and the post-chat survey.

JD Cloud & AI Beijing, China
Researcher
JD Customer Service Dialog System Dec. 2020 -- May. 2021

- Trained a FastText model to achieve Intent Detection (task-oriented or chatting) using dialogue data from JD e-commerce customer service platform
- Implemented Hierarchical Navigable Small World (HNSW) model to effectively retrieve relevant documents given input queries for task-oriented dialogues
- Adopted LightGBM to train a Learning-to-Rank (L2R) model to re-rank the retrieved documents according to the degree of similarity between input queries and the retrieved texts, using several similarity measurements (such as Edit Distance, Cosine Similarity, Jaccard Similarity, BM25) as features, along with deep learning based features extracted by BERT
- Trained a BERT model on *Large-scale Cleaned Chinese Conversation* dataset, under sequence-to-sequence framework, to automatically generate chatting response to user input

Marketing Text Generation Aug. 2020 -- Nov. 2020

- Aimed to automatically generate marketing texts for JD products, given data from JD e-commerce platform (such as product titles and product attributes)
- Implemented the baseline system composed of traditional sequence-to-sequence model with attention mechanism, using Pytorch framework
- Implemented Pointer-Generator-Network (PGN) model and realized the coverage mechanism to enhance the quality of the generated marking texts
- Optimized the beam search algorithm for decoder module via length normalization, coverage normalization and end-of-sentence normalization
- Applied weight tying and scheduled sampling to optimize the performance of PGN model
- Augmented training data through words switching and back translation

JD Book Classification Jun. 2020 -- Jul. 2020

- Aimed to automatically classify the book into one of the 33 categories listed in JD e-commerce platform, according to the textual description and the cover of each book
- Adopted TF-IDF, Word2Vec, FastText to extract the features of each textual description, together with several hand-crafted features (such as part-of-speech)
- Utilized ResNet to extract the visual features of each book cover, paired with the textual features for each book description
- Trained different machine learning models (such as LightGBM and SVM) and deep learning models (such as Bi-LSTM and LSTM) using the extracted textual and visual features, in order to compare their performances
- Applied SMOTE algorithm to address problems related to unbalanced labels

SELECTED AWARDS AND HONORS

- Zhejiang Provincial Excellent Graduates (top 4%, academic and overall performance) 2020
- Provincial Scholarship, Zhejiang Province (top 1%, academic performance) 2019
- Head's Scholarship, UNNC (top 10%, academic performance) 2019
- Ningbo Government Scholarship (top 5%, for academic performance in a foreign country) 2018
- Dean Scholarship, UNNC (top 5%, academic performance) 2017
- Outstanding Student Award (top 1% academic and overall performance) 2017
- GYLTLC, Outstanding Youth Leader Award (5/50) 2017
- GYLTLC, Most Outstanding Individual Award (1/50) 2017
- Third Place, Present Around the World Competition (Institute of Engineering/Technology speech contest) 2017

ADDITIONAL INFORMATION

Interests

- Passionate about reading (biographies, psychology)
- Sports (Captain, university men's basketball team; Gold medal, Sino-foreign International Sport Tournament)
- English debates: Participated in 3 British Parliamentary debate competitions in 2017

Languages

- Native Chinese Speaker
- Fluent English (TOEFL score 112: Writing 30, Speaking 29)

Computer and Language Skills

- C, Java, Python, C++, MySQL, Matlab, Linux, AWS Workspace, CSS, PHP, Javascript, HTML