Ziyi Liu

University of Southern California (Ink Lab)
3010 S Budlong Ave,Los Angeles, CA, 90007
+1 650-713-9455
zliu2803@usc.edu

Education

M.Sc. Computer Science Aug 2021-Present
University of Southern California, Los Angeles, CA
- Focus areas: Natural Language Processing
- GPA:3.75/4
- Advisor: Xiang Ren

B.Sc. Computer Science Aug 2016-Jul 2020
Shaanxi Normal University, Xi'an, China
- GPA:88.1/100
- Rank:1/50

RESEARCH EXPERIENCE

Evaluating Explanation Regularization Methods for NLP models
Advisor: Xiang Ren Feb 2022 - Present
- Proposed ER-TEST-a unified Explanation Regularization benchmark with various intrinsic and extrinsic evaluations.
- Conducted various instance prioritization methods in instance prioritization methods (Least Confidence, Entropy, etc.) to show ER-TEST's utility in low-resource learning, given ER rationale annotation budgets.
- Conducted distantly-supervised annotating in rationales, compared the performance of lexicon-based and instance-based methods in ID and OOD datasets.

Interpretable Question Asking in Collaborative VLN
Advisor: Jesse Thomason 2022 Jan - 2022 May
- Implemented interpretable tools (LIME, Input-X-Gradient) on Vision Language Navigation models to study what affects model to raise questions in a turn-based dialogue: what to ask and when to ask.
- Proposed two methods to analyse model interpretability: TF-IDF inspired approach to study attribution scores distribution of words towards target across documents; Comparing attribution scores and prediction confidence to measure the alignment between explanation and prediction.

Human-in-the-loop Debugging of NLP models
Advisor: Xiang Ren 2021 Oct - 2021 Dec
- Automatically extracted triggers from input to provide explanation for predicting.
- Developed a HITL annotating pipeline to intervene between teacher model (explanation) and student model(learning from explanation); Included human debugging of explanation to correct the model.
- Conducted multiple settings including time budget, annotating method to study how human intervention help in model performance improvement.

PROJECTS EXPERIENCE

Google Summer of Code - Chinese Pipeline with Red Hen Lab
Advisor: Mark Turner — May 2019-Sept 2019

- Recorded Chinese TV as dataset and develops ASR, NLP pipelines on it.
- Data preprocessing and model fine-tuning: Employed WebRTCVad to split the audios according to the pause and finetuned hyperparameter; Improved acc of baseline model by 10%.
- Designed the specification of NLP pipeline, catering to the practical needs and preprocessed the ASR output.

Story Generation based on BART — Sept 2021-Nov 2021

- Inspired by PlanAndWrite, built an pipeline generating stories from story titles and story lines using BART.
- Proposed 4 criteria (Alignment, Commonsense, Relatedness and Comparison) to evaluate model performance and conducted evaluation to quantify the completeness and novelty of generated stories; results on BART outperformed the baseline.

Machine Reading Comprehension based on DuReader — Jan 2020-Apr 2020

- Data preprocessing: cleaned useless data, filtered important information, and converted data formats.
- Fine-tuning: performed system optimization by adjusting parameters and testing BERT and XLNet for pre-training.
- Evaluation: Analyzed the performance of different types of problems of tasks against benchmark, and summarized the existing model verification methods and problems in the data set.

INDUSTRY EXPERIENCE

1DATA(Algorithm Engineer Intern) — Nov 2021-Jul 2021

Document key-value extraction project

- Extracted key-value pairs of airway bill forms from specific format to fasten the process of user operations.
- Used OCR, basic rules based on domain knowledge and form format to extract key value pair; Wrote validation metrics based on commonsense and relationships between keys to calculate confidence, to ensure the correctness of high confident results.
- Achieved the accuracy of 99.9% at 70% confidence level.

Format-free document parsing project

- Extracted specific information (port, shipper, address, etc) in shipping order forms in all kinds of format to help users fill in the forms.
- Used semi-supervised method to learn domain knowledge from curated sample set.
- Combined learned knowledge of both semantic and positional relationship to parse key value pair independent of form format.
- Evaluated accuracy vs computational time tradeoffs and iterated quickly to implement best shipping order solution for customers.
- Achieved the validation accuracy of 85% overall.

Silexon AI Technology(Algorithm Engineer Intern) Jul 2020-Nov 2021

- Crawled and collated peptide-protein interaction dataset and studied on interaction prediction based on deep learning framework.
- Applied relation extraction technology to mine drug-target relations in PubMed document.

PUBLICATIONS

Brihi Joshi*, Aaron Chan*, **Ziyi Liu**, Xiang Ren. ER-Test: Evaluating Explanation Regularization Methods for NLP Models. **TrustNLP2022**

Yipin Lei, Shuya Li, **Ziyi Liu**, Fangping Wan, Tingzhong Tian, Shao Li, Dan Zhao, Jianyang Zeng. A deep learning framework for multi-level peptide-protein interaction prediction. **Nature Communication**(accepted)

HONORS (Selected)

SNNU Honor Graduate - top 1% 2020
Shaanxi Normal University

National Scholarship - top 0.3 % 2017
The Ministry of Education of P.R. China

SNNU Fellowship - top 2 % 2017
Shaanxi Normal University

COURSES

University of Southern California
CSCI570: Algorithm Analysis
CSCI544: Applied Natural Language Processing
CSCI585: Database
CSCI699: Special Topics(Grounding Natural Language)

COMPUTER SKILLS

Programming Skills:Python, PyTorch, C, C++, Java, JavaScript, HTML, SQL, MATLAB

Keywords: Natural Language Processing, Grounding Natural Language, Deep Learning, Human-in-the-loop NLP