JIAZHAO LI

1839 Shirley Ln, Ann Arbor, MI. 48105
(+1)734 604 1596, jiazhaol@umich.edu, https://jiazhaoli.github.io

EDUCATION

University of Michigan, Ann Arbor U.S *Sept. 2020 – Present*
Ph.D in Informatics GPA: -

University of Michigan, Ann Arbor U.S *Sept. 2017 – Apr. 2019*
M.S in Electrical Computer Engineering GPA: 3.826
Courses: Information Retrieval, NLP, Data Mining, Machine Learning, Computer Vision

Nankai University China *Sept. 2013 – June. 2017*
B.S in Electrical Engineering GPA: 85.26/100
Courses: Data Structures and Algorithms, Computer Principle, Foundational of Computer Network, C++

RESEARCH EXPERIENCE

PharmMT: A Neural Machine Translation Approach to Simplify Prescription Directions.
[EMNLP21'] *Nov 2018 - Dec 2019*

- Built Nerual Network-based MT model between Prescription and Pharmacy directions corpus.
- Augmented model using MIMIC-III domain-specific pre-trained word embedding, external information from Drug/ Strength.
- Applied ensemble learning and numerical checking to improve accuracy and avoid fictitious generations.
- Applied BLEU score and SARI score to do automatic evaluation on MT performance and developed web app to do manual evaluation by pharmacists.

Identify Medication Relations from Clinical Narratives [Paper]

· Identifying medication relations between drugs and associated attributes automatically from clinical narratives to develop advanced tools for decision support. This is part of 2018 national clinical NLP challenge.

- Developed and shared python tokenization package for pre-processing MIMIC clinical notes for team.
- Feature engineering including part-of-speech tag, named-entities-recognition tag, pre-trained token embedding and bi-direction relative position of target entities pair.
- Developed Bi-LSTM models to extract 8 relations between drug names and adverse events associated concepts with F1 0.892 outperforming CNN and SVM model.

Baby blues: Analyzing Facebook and health forums on Pregnancy [Paper]

- Systematically analyzed questions posted on pregnancy forums by young mothers and contrasted it to a unique dataset of Facebook posts by expectant adolescent mothers.
- Implemented the topic model Latent Dirichlet Allocation (LDA) to extract the top 10 themes of questions in different trimesters during pregnancy.
- Concluded that Facebook is chosen as a self-expression place to seek emotional support while health forums served as professional information providers.

Movie Revenue prediction with Hierarchical Model [Poster] Jan. 2019 - Apr. 2019

- Observed movie revenue following two mixed Gaussian Distribution: High revenue and Low revenue.
- Based on observation, trained High-Low binary Random Forest Classifier based on labels resulted from Gaussion Mixture Model (GMM) clustering.
- Trained movie Gradient Boosting Regression (GBR performed best) separately on two revenue group and using Back-Off strategy to solve cold start problem.

Video Segments Retrieval System based on Attentive CNN [Report] Sep. 2018 - Nov. 2018

- Enhanced video clip embedding with attentive-weighted contextual video segments embedding.
- Generated cross latent feature between video clip embedding and corresponding video content description text embedding through outer product.
- Trained ACNN model on TACoS dataset with loss function on video-text similarity and offset of video clips achieved 0.347 (IoU=0.5) and 0.719 (IoU=0.1) in Top10.

Attention Based CNN: Sentiment Analysis for Yelp Reviews Sept. 2018 - Dec. 2018

- Combined Hierarchical CNN and Attention Mechanism to build a sentimental analysis classifier based on the Yelp review
- Applied NLTK and Word2Vec to implement text data processing, and rescale the label based on the data distribution
- Built CNN to capture N-gram information of text and introduced Attention Mechanism to improve the performance.
- The final accuracy of this model on the sentence level prediction is up to 69.85%.

PageRank++: European Soccer Team Ranking Prediction [Report] Jan. 2018 - Apr. 2018
Best poster of Graph Data Mining Course Project *University of Michigan*

· Designing PageRank++ algorithm to make a prediction on team ranking in the league before the season begin through re-defined directed graph of team.

- Applied K-mean to classify soccer player into 4 groups based on scores in 33 features in 5 fields of performance.
- Used PageRank iteration to predict rank where nodes representing feature vector of scores of players component for each team and edges of graph representing probability of win (directed).

PUBLICATIONS

Jiazhao Li, Corey Lester, Xinyan Zhao, Yuting Ding, Yun Jiang, and V.G.Vinod Vydiswaran. PharmMT: A Neural Machine Translation Approach to Simplify Prescription Directions. In Findings of EMNLP, the 2020 Conference on Empirical Methods in Natural Language Processing. (**Accepted**).
Zhao X, **Li J**, Lester C, Ding Y, Jiang Y, Vydiswaran VGV. *Focused representation with lexical constraints for parsing prescription instructions to decrease medication error.* (Under review)
Zhao X, **Li J**, Lester C, Jiang Y, Vydiswaran VGV *Focused representation models for transcribing prescription instructions.* (Poster MIDAS 2019 Symposium)
Li J, Liu J, Vydiswaran VGV. *Neural Network Models to Identify Medication Relationsfrom Clinical Narratives.* (MLHC 2019 Submitted)
Li J, Vydiswaran VGV *Baby blues: Analyzing Facebook and health forums on Pregnancy.* (WWW 2019 Submitted)

WORK EXPERIENCE

Research Associate(Aug 2019 -present)	NLP4Health Group	Prof. VG Vinod Vydiswaran
TA Grader (Jan 2019- Apr 2019)	Information Retrieval	Prof. Rada Mihalcea

SKILLS

Languages	Python, SQL, C++, JavaScript, HTML, MatLab
Framework/OS	Tensorflow, PyTorch, Hadoop, Linux

HONORS AND AWARDS

- Sep. 2016. Third Prize Scholarship of the University (10%), NKU
- Sep. 2016. Merit Student of the University (5%), NKU