JIACHENG LI

8681 Via Mallorca #37, La Jolla, CA, 92092
(858)-346-3311 ⋄ J9LI@ENG.UCSD.EDU

EDUCATION

University of California, San Diego Sep 2020 - Present
Doctor of Philosophy. Advisor: Julian McAuley.
Department of Computer Science and Engineering.

University of California, San Diego Sep 2018 - June 2020
Master of Science.
Department of Computer Science and Engineering.

Nanjing University of Posts and Telecommunications Sep 2014 - June 2018
Bachelor of Engineering.
College of Computer Science.

INTERESTED AREA

Information extraction, Natural language generation, Recommender system

WORK EXPERIENCE

Machine Learning for Natural Language Processing Intern June 2020 - Sep 2020
Bosch USA.
Project: Weakly supervised named entity recognition by rules.

Graduate Student Researcher-Machine Learning, NLP Sep 2019 - June 2020
Center for Microbiome Innovation
Project: Dual Relation Network for Few-shot NOTA Relation Extraction.

Summer Research Volunteer June 2019 - Sep 2019
Advisor: Julian McAuley
Project: Time Interval Aware Self-Attention for Sequential Recommendation.

PUBLICATIONS

Yang Jiao*, Jiacheng Li*, Jiaman Wu, Dezhi Hong, Rajesh Gupta and Jingbo Shang. SeNsER: Learning Cross-Building Sensor Metadata Tagger. (Findings of EMNLP) 2020.

Jiacheng Li, Yujie Wang, Julian McAuley. Time Interval Aware Self-Attention for Sequential Recommendation. International Conference on Web Search and Data Mining (WSDM) 2020.

Jianmo Ni, Jiacheng Li, Julian McAuley. Justifying Recommendations using Distantly-Labeled Reviews and Fine-grained Aspect. Empirical Methods in Natural Language Processing (EMNLP) 2019.

PROJECTS

Weakly Supervised Relation Extraction with Few-shot and Rules Expansion.
Ongoing Research Project. **Pytorch**. **Snorkel**

- An iterative framework to extract relation in a sentence given two entities from a few seeding sentences.
- Generate and select shortest path and entity pair rules from raw text. Apply rules on raw text to expand more entities.
- To deal with low recall problem from the rule module, Few-shot method is applied to find the nearest neighbors in the low dimension space.

- Train a Bert based relation classifier.

Weakly Supervised Named Entity Recognition with both Symbolic and Neural Representations
Ongoing Research Project. **Pytorch**. **Snorkel**

- An iterative framework to recognize entities starting from a few seeding entities.
- Generate and select rules from raw text and predicted entities automatically. Apply rules on raw text to expand more entities.
- An independent entity instances selector to reduce the noises in the expanded entities.
- Span-based named entity recognition.

SeNsER: Learning Cross-Building Sensor Metadata Tagger
Research Project. **Pytorch**.

- SeNsER learns a sensor metadata tagger for a new building based on its raw metadata and some existing fully annotated building.
- At the character level, it employs bidirectional neural language models to capture the shared underlying patterns between two buildings.
- At the word level, it leverages as features the k-mers existing in the fully annotated building.
- During inference, it incorporates the information obtained from sources such as Wikipedia as prior knowledge.

Time Interval Aware Self-Attention for Sequential Recommendation
Research Project. **Tensorflow**.

- Designed a novel time interval aware self-attention (TiSA) mechanism to learn the weight of different items, absolute position and time intervals to predict the following items.
- We proposed to view users interactions history as a sequence with different time intervals, and model different time intervals as relations between any two interactions.

Justifying Recommendations using Distantly-Labeled Reviews and Fined-Grained Aspects
Research Project. **Pytorch**.

- Defined each segment as an Elementary Discourse Unit (EDU) which corresponds to a sequence of clauses.
- Annotate data and trained GRU-based (Gated Recurrent Unit) and BERT-based text classifier to identify justifications from user reviews and compared the results.
- Reference-based Seq2Seq Model: A natural language generation model based on the seq2seq model. We apply two-layer bidirectional GRU as the encoder and decoder. Use attention mechanism to incorporate aspect information to improve controlment of generation.
- Aspect Conditional Masked Language Model: The masked language model in the pretrained BERT model as our sequence decoder and add attention over the aspect encoders output.

Joint Force in Managing Zambezi River
2017 Mathematical Contest in Modeling (MCM).
Outstanding Winner, SIAM Award.

- Design a new dam system to replace the old Kariba Dam on the Zambezi River.
- Investigate different facts on water flow and consider various situations under different weathers and terrains.
- Formulated the water flow balance equations.
- Used dynamic programming to get the best positions of dams implemented with Lingo.
- Used the particle swarm optimization in MATLAB optimization tools to get the most optimized volume of water should be scheduled.