Yufei **Li**

Computer Science · Machine Learning Researcher

(+1) 858-291-2678 | yxl190090@utdallas.edu | yul091.github.io/liyufei.me/ | yul091 | Yufei(Ethan) Li | @CavMonkey

Summary

Research at the intersection of natural language processing (NLP), computer vision (CV), and software engineering (SE). Focus on the application of NLP in data mining and recommender systems, and investigate the explainable AI such as model uncertainty measurements.

Education

University of Texas at Dallas — *Dallas, TX*

Ph.D. in Computational Science — *2020 - Present*

Artificial Intelligence Group, CSE; Advisor: Dr. Wei Yang

University of California, San Diego — *San Diego, CA*

M.S. in Electrical and Computer Engineering — *2018 - 2020*

Thesis: Application of A* and Genetic Algorithm in TSP Path Planning Problem; Advisor: Dr. Farinaz Koushanfar

Xi'an Jiaotong University — *Xi'an, China*

B.S. in Mechanical Engineering — *2014 - 2018*

Thesis: Aircraft fuel tank oil sloshing simulation based on SPH method and center of gravity distribution analysis; Advisor: Dr. Shuai Zheng

Publications

SHARE: a System for Hierarchical Assistive Recipe Editing.

Li, S., **Li, Y.**, Ni, J., McAuley, J. | *arXiv Computing Research Repository preprint:2105.08185 (CoRR 2021)* [PDF]

- Identified a controllable recipe editing task: adapting a base recipe to satisfy a user-specified dietary constraint to assist home cooks.
- Proposed a hierarchical framework which performs simultaneous ingredient substitution via a Set Transformer, followed by conditional instruction generation using a Copy Attention network.

Estimating Predictive Uncertainty Under Program Data Distribution Shift.

Li, Y., Chen, S., Yang, W. | *arXiv Computing Research Repository preprint:2107.10989 (CoRR 2021)* [PDF]

- Defined three real-world program distribution shifts based on practical software development scenarios.
- Investigated the effectiveness of existing uncertainty measurements under program distribution shifts and provided a large-scale benchmark for their performance.
- Analyzed each method's pros and cons from a logic design perspective to inspire a domain-specific uncertainty design.

GLIB: Towards Automated Test Oracle for Graphically-Rich Applications.

Chen, K.*, **Li, Y.***, Chen, Y., Fan, C., Hu, Z., Yang, W. | *29th Foundations of Software Engineering (FSE 2021)* [PDF]

- Proposed a code-based data augmentation approach for generating UI glitch images on game apps.
- Detected UI glitches via a CNN model and localized the glitch area using a saliency map to facilitate bug fixing.

Work Experience

NEC Laboratories America, Inc. — *Princeton, NJ*

Research Intern, PyTorch — *Jun 2021 - Aug 2021*

- Annotated name entities and relations with regular expression rules on CVE texts for distant supervision.
- Incorporated the pre-trained GPT-2 backbone into a sequence labeling framework for joint entity & relation extraction.
- Proposed a bootstrap training procedure for denoising distant labels and selecting high-quality instances.

SeekTruth Scientific and Technical Corporation — *Beijing, China*

Research Intern, TensorFlow — *Jul 2019 - Sep 2019*

- Boosted an adaptive discrimination definition model for objection detection.
- Designed a light-weight CNN model for identifying the direction of videos.
- Built a joint key point & pose recognition model for character detection.

Projects

GAET, Code Embedding

Dallas, TX

Research Project, PyTorch

Jul 2020 - Sep 2020

- Proposed a low-cost offline metric for evaluating the generalizability of code embedding in SE downstream tasks.
- Patched the existing pre-trained embedding based on the semantic metamorphic relationship to improve the generalizability.

Negative Sampling with Bias Correlation, Data Mining

San Diego, CA

Research Project, PyTorch

Nov 2019 - Jun 2020

- Analyzed one limitation of the popularity-based sampling scheme in terms of non-uniform negative sampling bias.
- Corrected the bias and designed related negative sampling distributions to boost the tradition Bayesian personalized ranking (BPR).

Pet Adoption Speed Prediction, Natural Language Processing

San Diego, CA

Research Project, TensorFlow

Jan 2019 - Mar 2019

- Parsed and tokenized the description text of pets, encoded the words with word embedding.
- Designed & built a hierarchical framework to combine image features & natural language semantics for popularity prediction.

Automatic Delivery Vehicle Design, Path Planning

San Diego, CA

Research Project, Python & MATLAB

Mar 2019 - Jun 2019

- Incorporated the Courier and TSP travel agent problems into designing autonomous delivery vehicles.
- Designed & built a dynamic path planning model combining genetic algorithm and A* algorithm for path planning.

Awards

VEX Robotics International Competitions

Team Competition, C language

- Excellent award and runner-up in VEX Robotics World Championship (RECF) 2017, Louisville, KY, US.
- Excellent award and runner-up in VEX Robotics Asia Open 2016, Beijing, China.
- First class honor in VEX Robotics China Open 2016, Xi'an, China.

Scholarship Awards

Personal

- National encouragement scholarship 2015-2017, top10 of 300+ applicants.