YUFEI LI

Phone: (858) 291-2678 | Email: yxl190090@utdallas.edu
Personal Website: https://yul091.github.io/liyufei.me/ | LinkedIn: www.linkedin.com/in/yufei-li

PERSONAL STATEMENT

- First-year Ph.D. student at UT Dallas advised by Prof. Wei Yang, have the experience of Computer Vision (CV) algorithm internship and Recommender System (RecSys) & Natural Language Processing (NLP) & software engineering (SE) research experiment. My interest lies in the application of Deep Learning (DL) in NLP and SE fields and analysis of model uncertainty
- Proficient in writing Python / MATLAB language, designing deep learning algorithm and solving relevant problems efficiently, proficient in using PyTorch / TensorFlow and other models

EDUCATION

Doctor of Philosophy in Computer Science, the University of Texas - Dallas   07/2020
- Relevant Research: GAET: Towards Automatically Characterizing the Generalizability of Code Embeddings, Analysis of Uncertainty Metrics in Programming Language

Master of Science in Electrical and Computer Engineering, University of California - San Diego   09/2018
- Relevant Research: Accommodating Dietary Restrictions via Hierarchical Recipe Editing, Negative Sampling with Bias Correlation for Pairwise Personalized Ranking
- Relevant Courses: Machine Learning & Data Science, Linear Algebra and Application, Statistical Learning, Mathematics for Engineers, Programming for Data Analysis, Neural Networks/Pattern Recognition, Learning Algorithms, Random Process, Deep Learning and Neural Network, Computer Vision for Machine Learning, Recommendation System

Bachelor of Science in Mechanical Engineering, Xi'an Jiao Tong University   09/2014
- GPA: 3.5/4.3(85.7/100)
- Relevant Courses: Probability Theory & Mathematical Statistics, C language, C# language and application

INTERNSHIP EXPERIENCE

SeekTruth Scientific and Technical Corporation - Algorithm Intern   Beijing, China   06/2019-09/2019

- **BMW video project (OpenCV): design an adaptive discrimination definition model and improve its performance**
  Combined Laplace transform and gaussian filter to calculate the mean square deviation of image edges, also used adaptive threshold changing method to quickly judge the clarity of video files. The model could be trained in real time and showed decent recognition efficiency. The judgment time per frame was 5 ~ 10ms

- **BMW image recognition project (TensorFlow-Keras): design small convolutional network to identify the direction of video**
  1) Classify character orientation in video by simple model, enhance and clean the data set so that the model can quickly identify character orientation in different actual scenes (lift, sign, character)
  2) Independently design a small four-layer convolutive network, which is comparable to the effect of a large model network. Through experiments, the model showed good robustness under complex background conditions. Accelerate the model by optimizing the loss and training mode, which presented obvious optimization effect, up to millisecond level

- *SeekTruth* **AI pose recognition project (TensorFlow): designed and trained key points and pose recognition models for characters, and optimized the previous version of Caffe model**
  Selected Images, key points of COCO data set and obtained CNN convolution network, heatmap and coordinates of key points according to DSNT matrix algorithm, also optimized MSE loss and training method (considering Shannon entropy of heatmap and target gaussian map)

RESEARCH PROJECTS

GAET: Towards Automatically Characterizing the Generalizability of Code Embeddings – Natural Language Processing & Software Engineering
Dallas, US   07/2020-09/2020
- We observe that existing works of code embeddings miss the evaluation of generalizability of code embedding and we identify two major challenges in performing such an evaluation. To address the challenges, we propose a low-cost offline metric GAET for evaluating generalizability of code embeddings in SE applications
- Our study motivates the use of pre-trained embeddings and demonstrates the poor generalizability of some existing embeddings. Our study also makes a key finding that the semantic metamorphic relationship among vectors in an embedding can be an indicator of the embedding's generalizability
- The experiment results show that our method is more effective and efficient than the existing techniques and GAET is not sensitive to the hyperparameters
- We also apply GAET for patching a pre-trained embedding, the results show that the generalizability of the vectors have a significant increase after patching

Hierarchical Recipe Editing – Natural Language Processing
San Diego, US   02/2020-06/2020
- We annotate recipe with one of the seven dietary restriction categories (low sugar, low fat, low carb, low calorie, low gluten, low dairy, vegetarian) using curated regular expressions on recipe names and create match pairs considering their edit distance
- To leverage recipe ingredient overlap, we explore two types of edit operations: *insertions* and *deletions* from the base ingredient set to create target ingredients. Our ingredient prediction model realizes this by defining Bi-LSTMs (Encoder) and Linear projection.
- Our hierarchical framework allows us to train a step predictor based on previous predicted ingredients, we apply Pointer Attention to ingredients at each time step to further improve the Step Transformer Encoder-Decoder model

Negative Sampling with Bias Correlation – Data Mining
San Diego, US   11/2019-06/2020
- The performance of unsuccessful popularity-based sampling scheme can mainly be explained by sampling bias introduced from non-uniform negative sampling process
- We propose two ways to correct the bias and design related negative sampling distributions to improve accuracy from tradition Bayesian Personalized Ranking (BPR)
- Our experiments show the effectiveness of our method, which could be helpful in the future design of negative sampling schemes

Pet Adoption Speed Prediction – Computer Vision & Machine Learning
San Diego, US   01/2019-03/2019
- Given a database of 15000 cats and dogs' features (gender, fur color, size, etc. description and adoption speed), need to develop algorithms to predict the adoptability of pets
- Responsible for description and comparison of ML algorithms in image processing. Train sentences by using NLP and Word Embedding to transfer words into vectors, compare algorithms like ResNet, DenseNet, SVM and CNN

Automatic Delivery Vehicle Design – Path Planning Algorithm
San Diego, US   03/2019-06/2019
- Application of the problem of Courier and TSP travel agent in autonomous delivery vehicle
- In the genetic algorithm, each order has a relevant fitness, that is, the sum of the paths between each two destinations. The path between each two points is to be reached by the dynamic calculation of the heuristic function through the A* algorithm. Individuals with higher fitness have a higher probability of inheriting to their children
- A dynamic path planning model combining genetic algorithm and A* algorithm is designed. This model has the advantages of real-time training and detection without GPU calculation

PRIZE & Publications

Certifications
- First Class Honor in VEX Robotics China Open 2016-2017, Xi'an, China, 12/2016
- Excellent Award in VEX Robotics Asia Open 2016-2017, Beijing, China, 02/2017
- VEX Robotics Championship 2016-2017, Louisville, KY, US, 04/2017