

Da-Wei Lee

MASTER STUDENT · NATURAL LANGUAGE PROCESSING

☎ (+86) 156-5225-0910 | ✉ dwlee@pku.edu.cn | 🏠 dwlee-personal-website.netlify.com | 📷 daviddwlee84



Summary

A Master student at Peking University. Enjoy being a **maker** - think of any creative idea and try hard to make it come true. A Vim addict, loved discovering new technologies, and problem-solving. Used to famalier with Embedding System Design. Recently focus on Natural Language Processing and Knowledge Graph related research using Deep Learning.

Education

Peking University

MASTER IN SOFTWARE ENGINEERING

Sep. 2018 - Present

- Focus on NLP and some related application in Knowledge Graph recently.
- Core member of PKU Open Source Association in ML/NLP department.

Taiwan Tech (National Taiwan University of Science and Technology)

BACHELOR IN ELECTRONIC AND COMPUTER ENGINEERING

Sep. 2013 - Jun. 2017

- Focus on Embedding System Design and any other cool engineering stuff like APP, Web, Game design.
- The graduate project is Quadcopter with CV Control. Got two school awards and other two competition awards.
- Got 4 individual piecework experiences and 5 engineering competitions based on different topics.

Experience

Microsoft Search Technology Center Asia Carina Group

ALGORITHM INTERN

Jul. 2020 - Present

- AI Writer for automatic article generation and rewriting by generation models and information retrieval methods.
- Writing Assistant which offers such as copyediting, typesetting, proofreading, indexing, page makeup.

Microsoft Research Asia Knowledge Computing Group

RESEARCH INTERN

Dec. 2019 - May. 2020

- Generation of slides from academic paper.

Peking University National Engineering Research Center of Software Engineering

RESEARCH INTERN

Jul. 2019 - Present

- Doing case of Anti-healthcare fraud and Medical record analysis.
- Including research of Knowledge Extraction, Knowledge Graph

Taiwan Industrial Technology Research Institute

EMBEDDING SYSTEM DESIGN SOFTWARE INTERN

Jul. 2016 - Aug. 2016

- In the self-driving group, I mainly handled the STV0991 development board which was going to carry the computer vision algorithms.

Project

GCAKE: Graph and Context Attentional Knowledge Embedding

ACL (THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS) IN PROGRESS

Oct. 2019 - Present

- A knowledge graph representation learning using self-attention (a Transformer like architecture) encoded context not only between triples but also considered the description context and graph structure.

Similar Cases Recommendation via Legal Knowledge Graph Construction and Representation

WWW (THE WEB CONFERENCE) UNDER REVIEW

Aug. 2019 - Oct. 2019

- Propose a pipeline from raw lawsuit case into dense representation to compare and recommend the top k most similar cases.
- Including subtask of Named-entity Recognition, Relation Extraction, Knowledge Graph Embedding. Using jointly-trained multitask model of fine-tuned BERT.

Stanford CS224n: Natural Language Processing with Deep Learning

ONLINE COURSE

Jul. 2019 - Dec. 2019

- To review the detail learned from the last two semesters.
- Project including Neural Dependency Parsing, Neural Machine Translation, Question Answering.

Jigsaw Unintended Bias in Toxicity Classification

KAGGLE COMPETITION

Feb. 2019 - May. 2019

- This competition is aim to classify whether a comments is toxic. Our team design different models such as BERT, ELMo etc. as classifier and finally ensemble them. Our team reach Top 1% in rank.

SemEval-2013 Task 13: Word Sense Induction for Graded and Non-Graded Senses

COURSE PROJECT

Jun. 2019 - Jul. 2019

- This task is word disambiguation. Determine a meaning of a word in different sentences. There are two subtask, one is finding the closest WordNet lemma, another is clustering between sentences.

SemEval-2018 Task 7: Semantic Relation Extraction and Classification in Scientific Papers

COURSE PROJECT

May. 2019 - Jun. 2019

- This task is the relation classification task. The scenario is based on the abstract of lots of papers. For each sentence, given two entities, and we have to determine which relation might between the entities.

Chinese Word Segmentation, Part-of-speech Tagging, Named-entity Recognition

COURSE PROJECT

Apr. 2019 - May. 2019

- The classic Chinese sequence labeling tasks. I used the BiLSTM-CRF model as the main approach compare with other baselines.

Failure Prediction of Concrete Piston for Concrete Pump Vehicles

DIGITAL CHINA INNOVATION CONTEST 2019

Jan. 2019 - Mar. 2019

- In this competition, each sample is a time-series data of a concrete pump vehicle. The goal is to predict the likelihood of each data sequence that whether a machine might fail. I used LightGBM and reach Top 5% in rank.

AI BOT of the Game 2048 using MCTS and Reinforcement Learning

COURSE PROJECT / NATIONAL TAIWAN UNIVERSITY 2048 BOT COMPETITION IN 2014

Nov. 2018 - Dec. 2018

- In 2014, I used MCTS with alpha-beta pruning to select best action. And score each state(board) with our own designed evaluation function.
- In 2018, I built a Reinforcement Learning model and test performance between Policy Gradient and Deep Q-Learning.