

# Da-Wei Lee

MASTER STUDENT · NATURAL LANGUAGE PROCESSING

☎ (+86) 156-5225-0910 | ✉ dwlee@pku.edu.cn | 🌐 dwlee-personal-website.netlify.com | 📷 daviidwlee84



## 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 familiar 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.
- 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

- Participating cases 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.

## ML/NLP Project

### GCAKE: Graph and Context Attentional Knowledge Embedding

SIDE PROJECT

Oct. 2019 - Jan. 2020

- 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

SIDE PROJECT

Aug. 2019 - Oct. 2019

- Proposed a pipeline from raw lawsuit cases 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.

## Sentence Similarity in Intelligent Question Answering

SIDE PROJECT

Aug. 2019 - Sep. 2019

- Mainly used the Enhanced RCNN model (BiLSTM + Attention + CNN) and some divergences to optimize the sentence similarity calculation.
- In the scenario of Chinese intelligent question answering, finding an option which considers both the performance and complexity at the same time.

## Stanford CS224n: Natural Language Processing with Deep Learning

ONLINE COURSE

Jul. 2019 - Dec. 2019

- Reviewed the detail learned from the last two semesters.
- Projects including Neural Dependency Parsing, Neural Machine Translation, Question Answering.

## Jigsaw Unintended Bias in Toxicity Classification

KAGGLE COMPETITION

Feb. 2019 - May. 2019

- This competition aimed to classify whether a comments is toxic. Our team designed different models such as BERT, ELMo etc. as classifier and finally ensembled them. Our team reached Top 1% in rank.

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

COURSE PROJECT

Jun. 2019 - Jul. 2019

- The task was word disambiguation. Determined a meaning of a word in different sentences. There were two subtask, one was about finding the closest WordNet lemma, another was about clustering between sentences.

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

COURSE PROJECT

May. 2019 - Jun. 2019

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

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

COURSE PROJECT

Apr. 2019 - May. 2019

- Classic Chinese sequence labeling tasks. I used BiLSTM-CRF model as the main approach comparing 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 was a time-series data of a concrete pump vehicle. The goal was to predict the likelihood of each data sequence that whether a machine might fail. I used LightGBM and reached Top 5% in rank.

# Engineering Project

---

## Raspberry Pi Cluster

SIDE PROJECT

Sep. 2018

- Built a cluster with 4 Raspberry Pis with a self-made efficient script which can quickly apply popular ecosystem like Hadoop, Spark.

## Leapsy AR Glasses Video Stream Pan/Tilt Head

PIECE WORK

Jul. 2017 - Oct. 2017

- I collected sensor data on Android-based AR Glasses to capture current attitude and sent it back to Raspberry Pi to synchronize camera pan/tilt head's direction then return video stream back to glasses through Wi-Fi.
- I made pan/tilt head structure using 3D print model to contain camera and two servo motors, and designed the power supply circuit for both motors and Raspberry Pi.

## Modularized Quadcopter Architecture with Computer Vision Control

UNDERGRADUATE PROJECT

Apr. 2016 - Sep. 2016

- Built a quadcopter from scratch, including motors, sensors, power circuits, etc.
- Used PID for balancing algorithm on the control board and object detection with OpenCV on the CV board.
- Got honorable award at HOLTEK MCU Design Contest 2016 and Top 10 at ARM Design Contest 2016.