Yun Cheng Wang

Los Angeles, CA, USA
213-379-0669 yunchenw@usc.edu linkedin GitHub

Education

University of Southern California **Jan. 2021 – Present**
Ph.D. in Electrical and Computer Engineering, Advisor: Prof. C.-C. Jay Kuo *Los Angeles, CA*

University of Southern California **Aug. 2018 – Dec. 2019**
M.S. in Electrical and Computer Engineering, GPA: 4.0/4.0 *Los Angeles, CA*

National Taiwan University **Sep. 2014 – Jun. 2018**
B.S. in Electrical Engineering, GPA: 3.8/4.3 (A+: 4.3) *Taipei, Taiwan*

Research Interests

- **Machine Learning**: Representation Learning, Feature Extraction, Explainable Machine Learning
- **Semantic Web**: Knowledge Base Completion, Knowledge Acquisition, Information Extraction

Publications

Peer Reviewed Papers

- Bin Wang, Fenxiao Chen, Yun-Cheng Wang, C.-C. Jay Kuo, "Efficient Sentence Embedding via Semantic Subspace Analysis", *International Conference on Pattern Recognition (ICPR), 2020.*
- Fenxiao Chen, Yun-Cheng Wang, Bin Wang, C.-C. Jay Kuo, "Graph representation learning: A survey", *APSIPA Transactions on Signal and Information Processing, 2020.*
- Bin Wang, Angela Wang, Fenxiao Chen, Yun-Cheng Wang, C.-C. Jay Kuo, "Evaluating word embedding models: Methods and experimental results", *APSIPA Transactions on Signal and Information Processing, 2019.*

Papers Under Review

- Yun-Cheng Wang, Xiou Ge, Bin Wang, C.-C. Jay Kuo, "KGBoost: A Classification-Based Knowledge Base Completion Method with Negative Sampling", *AAAI Conference on Artificial Intelligence (AAAI), 2022.*

Experience

USC Media Communication Lab (MCL) **Aug. 2018 – Present**
Student *Los Angeles, CA*

- Devoted to develop lightweight, feed-forward, and high-performance machine learning models.
- Predicting missing relations in the knowledge graphs. While previous work solves the problem via training end-to-end embedding models, we developed an modularized system including, entity feature extraction, triple preprocessing, and classification on missing links.
- The modularized design not only allows the classifiers to operate under a low-dimensional setting, but also gives users the freedom to use different entity encoders and integrate multiple negative sampling strategies.
- Prior to the knowledge graph research, we have done comprehensive survey on graph and word representation learning models.

Taiwan Semiconductor Manufacturing Company **Feb. 2020 – May. 2020**
Defect Analysis Engineer *Hsin-chu, Taiwan*

- Responsible for identifying the possible causes of the defects by analysing data in the manufacturing process, including defect images, station chemicals and temperatures, and time intervals in the stations.

Taboola, Inc. **Jun. 2019 – Aug. 2019**
Data Science Intern *Los Angeles, CA*

- Responsible for building large-scale knowledge graphs based on trending topics from multiple sources.
- The main challenges include deduplication of entities, incrementally update the knowledge base, and building an automation pipeline from the sources to the products.
- The whole pipeline is ran daily and is able to process 5,000 articles within an hour.

Projects

YouTube Video Content Retrieval System **Spring 2019**

- Users sometimes would like to look for a certain object in the video or a certain phrases in a speech. However, the current searching system on Youtube doesn't support such a feature.
- Developed a retrieval system that takes users' queries, such as "dolphin" and returns videos containing dolphins.
- The system is powered by the NLP and Vision APIs on Google Cloud Platform.

Video Summarization System **Summer 2017 – Summer 2018**

- This project is a collaboration project between KKTV, an exclusive OTT media service provider in Taiwan, and MPAC lab at NTU. The project goal is to understand the storylines in the dramas with machine learning models and extract the important frames to form video summaries.
- While key frames are extracted based on the qualities of frames, the system further process the transcripts to capture the plots and climaxes in the dramas.
- We are able to generate video summaries that are capable of "story-telling", and have done multiple subjective testing.
- The results are presented in front of over 50 data scientists in an annual company meeting.

Relevant Coursework

- Digital Image Processing
- Multimedia Data Compression
- Machine Learning from Signals
- Mathematical Pattern Recognition
- Building Knowledge Graphs
- Natural Language Processing
- Digital Signal Processing
- Digital Speech Processing
- Machine Learning Fundamentals
- Algorithms

Honors, Teachings, and Members

- IEEE Student Member in Signal Processing Society, 2019–2021
- Course Mentor for Graduate-level Probability (EE503), Fall 2019
- USC EE Master Honors Program (Highly Selective), Dec. 2019

Technical Skills and Languages

Languages: Chinese (Native), English (Fluent)
Programming: Python, C++, Java, Matlab, SQL
Software: Pytorch, scikit-learn, XGBoost, Scrapy, Neo4j
Infrastructure: Kubernetes, Apache Spark
Cloud Computing Services: Google Cloud Platform, IBM Watson