



Tony Lin

+86-185-1197-6446 | sss950123@gmail.com

Beijing / Taiwan | WeChat: s950123 | github.com/tnlin

Expected to graduate in 2020 | Objective: Applied Scientist (NLP)



SUMMARY

- Computer science major, the research direction is natural language processing, having published paper in ACL as the first author.
- 3 years of experience in data science, having completed more than 10 projects, and can independently lead the team to complete the task

EDUCATION

Tsinghua University (THU)

Sep 2017 - Jun 2020

Master, Computer Science, State Key Laboratory of Intelligent Technology and Systems

Beijing

- GPA: 3.62 / 4.0 | Awards: The Special Prize of Hong Kong, Macau and Overseas Chinese Scholarship (**Top 8 students**)
- Related Courses: Advanced Machine Learning, Frontier Research in Information Retrieval, Computational Linguistics
- TA Courses: Data Mining: Concepts and Techniques (2018-present), Industrial Data Mining and Analysis (2018-present)

National Chiao Tung University (NCTU)

Sep 2012 - Jun 2016

Bachelor, Electrical and Computer Engineering

Taiwan

- GPA: 3.65 / 4.0 | Award: Outstanding Contribution Award (**Top 1% of graduates**)
- Second place of Japan TUAT startup competition and receive a full scholarship to visit Stanford Research Institute (SRI) for a week

PUBLICATION

Research Topic: Discover new intents of dialogue systems

- **Ting-En Lin**, Hua Xu, Hanlei Zhang. Discovering New Intents via Constrained Deep Adaptive Clustering with Cluster Refinement, AAAI 2020.
- **Ting-En Lin**, Hua Xu. Deep Unknown Intent Detection with Margin Loss. ACL 2019.
- **Ting-En Lin**, Hua Xu. A Post-processing Method for Detecting Unknown Intent of Dialogue System via Pre-trained Deep Neural Network Classifier. Knowledge-based Systems 2019. (IF=5.1)

SKILL

- NLP: Multi-turn dialogue intent classification / Anomaly detection / Unsupervised & Semi-supervised clustering / Active learning
- Programming Language : Python (5 years) / SQL (3y) / C++ (2y) / C (2y) / Java (1y) / Bash (1y)
- Libraries : Pandas (4y) / Sklearn (4y) / Matplotlib (4y) / Keras (2y) / Pytorch (1y) / Tensorflow (1y)
- Others : Linux (6y) / Git (4y) / MySQL (4y) / Docker (3y) / MongoDB (1y) / Tableau (1y)

WORK EXPERIENCE

MoBagel

Sep 2016 - Dec 2017

Consultant & Intern, Data Science

Taiwan

- Participate in GB~TB level data mining projects of Fortune Global 500 companies, including churn / malfunction / sales prediction... etc
- Participate in Automated Machine Learning (AutoML) project, including the development of testing tool, SDK, middleware, and client software

PROJECT

Cross-modal sentiment classification (Fortune Global 500 companies)

Mar 2019 - Present

Leader, Predicting the sentiment of the customer review audio is positive or negative

Beijing

- Audio: MLP + statistical features
 - 12% improvement in accuracy compared with LightGBM+MFCC features (54.4% -> **67%**)
 - Do error analysis by feature importance and confusion matrix
 - Use Librosa to segment audio into sentence-level audio (2~20 seconds), based on loudness threshold, minimum silence duration...etc.
 - Use Opensmile to extract 1500+ statistic features from audio, including loudness, MFCC, audible probability, etc.
- Text: pre-trained BERT (Bidirectional Encoder Representations from Transformers)
 - 23% improvement in accuracy compared with text CNN (72.6% -> **96%**)
 - Use text CNN (with different kernel sizes), BiLSTM (with dropout) as the baseline
 - Use Jieba to segment Chinese text (with the external dictionary) and build vocabulary
 - Use pre-trained Chinese word vectors (CA8, Tencent) to initialize the weights of the embedding layer & fine-tune it during training
- Modal fusion: audio & text fusion at the decision level to improve robustness of prediction
- Write a 16-page English report and develop the corresponding tools for the company (Web application, using Flask to develop the backend)

OTHERS

- **Competitions and Misc.:** 2018 Tencent Advertising Algorithm Competition - CTR Forecast (ranked 51/1500+), 2017 Booking.com hackathon (champion, collaborative travel planning tool), 2016 PokemonGo-TSP (100+ Stars on Github, using simulated annealing to solve traveler salesman problems in PokemonGo), 2015 NCTU Hackathon (champion), 2014 NCTU Project Competition (champion)
- **Activities:** THU Taiwan Graduate Student Association (VP), THU MOOC Association (VP), NCTU Student Union (Head of Information)