

# JINYANG LIU

✉ [liujy57@mail2.sysu.edu.cn](mailto:liujy57@mail2.sysu.edu.cn) · ☎ Tel/Wechat (+86) 136-0975-2831

## 📖 RESEARCH INTERESTS

---

### Log Mining, NLP, CTR Prediction

- Log parsing and log compression
- Text semantic matching, text classification, and word embedding
- Deep learning-based models and distillation models in CTR prediction

## 🎓 EDUCATION

---

**Sun Yat-sen University (SYSU)**, Guangzhou, China 2018 – Present

*Postgraduate student* in Computer Technology, expected June 2020

Supervisor: [Prof. Zibin Zheng](#), Dean of Software Engineering Department

**Sun Yat-sen University (SYSU)**, Guangzhou, China 2014 – 2018

*B.E.* in Software Engineering

Overall GPA Rank TOP %1, CET-6 score: 560

## ✍ WORKING EXPERIENCE

---

**Chinese University of Hong Kong (CUHK)**, HongKong, China 2019.07 – 2020.01 (Expected)

*Research Assistant* Supervisor: [Prof. Michael R. Lyu](#)

**Huawei 2012 Lab**, Shenzhen, China 2018.01 – 2018.08

*Research Intern* Mentor: [Dr. Jieming Zhu](#)

## 🔍 RESEARCH EXPERIENCE

---

### ==Log Mining==

#### Huawei Phone Duplicate Issues Detection

- Detect duplicate issue report (including issue descriptions and logs) from Huawei phone users.
- Responsible for log parsing, log matching and feature extraction.
- The system eventually achieved more than 80% accuracy.

#### Log Parser Benchmark

- Implemented, reorganized, and tuned 13 state-of-the-art log parsing algorithms : IPLoM[*KDD*], LogSig[*CIKM*], LogMine[*CIKM*], MoLFI[*ICPC*], etc.
- We evaluate their accuracy, efficiency, and robustness on 16 datasets.
- The corresponding paper is accepted by ICSE-SEIP 2019.

#### Logzip

- We proposed logzip to optimize existing compressing tools for log compression.
- We conduct log parsing before compressing a log file to decrease its entropy, and low-entropy files can be well compressed by existing tools.
- Logzip gets 4.56x the compression ratio of gzip. The corresponding paper is submitted to ASE 2019.

### ==NLP==

#### Text Semantic Matching

- Computing the semantic similarity of two sentences.
- Investigated and implement deep learning algorithms: Decomposable Attention[*ACL, Google*], BIMPM[*IJCAI*], AF-DMN[*IJCAI*], etc.

- We proposed to build sentences as a graph and utilize information from adjacent nodes when inferring the similarity of two sentences, which slightly improves the accuracy of baseline models.

### Text Classification [Graduation Project]

- Decide which Emoji to use in a sentence.
- We turn a classification problem into a matching problem by embedding labels (Emojis) to enrich label information and help model training.
- The work was rated as an excellent graduation project and similar work is shown on ACL 2018.

### Word2Matrix

- Learn multiple representations for each word from a large corpus to reduce words' ambiguity.
- We utilize attention mechanism to capture context of a word and represent each word as multiple vectors.
- Our model successfully distinguish some polysemous words like "apple", "doctor", etc.

### ==CTR Prediction==

#### Linkage Embedding for CTR

- Investigate and implement CTR algorithms: FM, DeepFM[IJCAI, Huawei], Wide&Deep[Google], etc.
- Existing algorithms embed category features as vectors and perform feature interactions by the dot product. But they ignore interactions within a field, which is not conducive to the learning of low-frequency features.
- We proposed Linkage Embedding to link the features within a field and improve baselines with 1% AUC. The corresponding paper is submitted to RecSys 2019.

#### Distillation model

- Investigate and implement neural network distillation strategies: Knowledge Distillation [Geoffrey E. Hinton], Rocket[Alibaba], etc.
- Deploying a new model in production requires to go through a long and tedious process (maybe months) of online code modification and rigorous testing for model serving.
- We propose to the use of knowledge distillation (KD) to unify the model serving for CTR prediction. We successfully obtain a unified and easy-to-deploy model that can surprisingly outperform state-of-the-art models. The corresponding paper is submitted to RecSys 2019.

## ♡ HONORS AND AWARDS

---

First Class Scholarship (Postgraduate)	2018
Mathematical Contest in Modeling, Meritorious Winner	2017
2 <sup>nd</sup> Prize in Public Governance Data Analysis Competition	2017
3 <sup>rd</sup> Prize in Guangdong Big Data Application Innovation Competition	2017
First Class Scholarship (Graduate)	2015

## ⚙️ SKILLS

- Programming Languages/Frameworks: Python > C++ > Java | Pytorch == Keras > Tensorflow
- Open source projects: I am a contributor of [logparser](#) where we implemented most state-of-the-art log parsing algorithms and gain more than **150** stars.

## 📖 PAPERLIST

- Jieming Zhu, Shilin He, Jinyang Liu, Pinjia He, Qi Xie, Zibin Zheng, Michael R. Lyu. Tools and Benchmarks for Automated Log Parsing. To appear in ICSE 2019.
- [Jinyang Liu](#), Jieming Zhu, Liang Chen, Gang Wang, Zibin Zheng, Yuzhou Zhang. Unifying Model Serving for CTR Prediction with Knowledge Distillation. Submitted to RecSys 2019.
- [Jinyang Liu](#), Jieming Zhu, Shilin He, Zibin Zheng, and Michael R. Lyu. Logzip: Optimizing System Log Compression with Inherent Structures. Submitted to ASE 2019.
- Haicheng Xu, Jieming Zhu, [Jinyang Liu](#), Zibin Zheng and Wuhui Chen. Linkage Embedding: Boosting FM-based CTR Prediction Models by Sharing Feature Embeddings. Submitted to RecSys 2019.

## REFEREES

---

- [Prof. Zibin Zheng](#), Dean of Software Engineering Department, Sun Yat-sen University, zhzibin@mail.sysu.edu.cn
- [Dr. Jieming Zhu](#), Huawei Noah's Ark Lab, jamie.zhu@huawei.com