Jinyang Liu

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

⚑ Research Interests

Log Mining, NLP, CTR Prediction

- Log parsing, log compression
- Text sematic matching, text classification, word embedding
- Deeplearning-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

Huawei 2012 Lab Shenzhen, China   2018.01 – 2018.08

Research Intern Mentor: Dr. Jieming Zhu

Inplus Lab SYSU, Guangzhou, China   2017.09 – Present

Postgraduate

ℹ Research Experience

==Log Mining==

Huawei Phone Duplicate Issues Detection

- Detect duplicate issue report (including issue description 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 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 can get average 4.56x compression ratio than gzip.

==NLP==

Text Sematic Matching

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

- We proposed a graph-based model. We 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.
- My 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 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.

Distillation model

- Investigate and implement neural network distillation strategies: Knowledge Distillation *[Geoffrey E. Hinton]*, Rocket*[Alibaba]*, etc.
- There is a great demand for on-line inference speed of CTR models in the industry, and many deep models are time-consuming, so we are trying to use different distillation strategies to compress the model and get faster inference speed under the condition of guaranteed effectiveness.

♡ Honors and Awards

First Class Scholarship (Postgraduate)	2018
Mathematical Contest in Modeling, Meritorious Winner	2017
2nd Prize in Public Governance Data Analysis Competition	2017
3rd Prize in Guangdong Big Data Application Innovation Competition	2017
First Class Scholarship (Undergraduate)	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 International Conference on Software Engineering (ICSE), 2019.
- Jinyang Liu, Jieming Zhu, Shilin He, Zibin Zheng, and Michael R. Lyu. Logzip: Optimizing System Log Compression with Inherent Structures. In submission.
- 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 SIGIR 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