Jinyang Liu

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

Research Interests

Recommendation System, Log Mining, NLP

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

Education

Sun Yat-Sen University (SYSU), Guangzhou, China — 2018 – Present

M.Eng in Computer Technology, expected June 2020
Supervisor: Prof. Zibin Zheng, Dean of the Software Engineering Department

Sun Yat-Sen University (SYSU), Guangzhou, China (Ranked **8th** in China) — 2014 – 2018

B.E. in Software Engineering
GPA: 3.9/5.0, ranks: 36/388 (top 10%), CET-6 score: 560

Working Experience

Chinese University of Hong Kong, HongKong, China — 2019.07 – present

Research Assistant Supervisor: Prof. Michael R. Lyu

Huawei 2012 Lab, Shenzhen, China — 2018.01 – 2018.08

Research Intern Mentor: Dr. Jieming Zhu

Research Experience

==Recommendation System (main)==

** I am cooperating with the Searching Team of Huawei Noah's Ark Lab in this field.*

Models Ensemble via Knowledge Distillation (KD)

- Investigated/implemented (1) KD models: Knowledge Distillation *[G.E. Hinton]*, Rocket*[Alibaba]*, etc. (2) Click-Through Rate (CTR) prediction models: FM, DeepFM*[IJCAI, Huawei]*, Wide&Deep*[Google]*, xDeepFM*[KDD]*, 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 proposed the use of 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. We also distilled multiple models into a single model that performs the best. The corresponding paper is in submission.

Deep Learning-based Graph Models

- Investigated/implemented algorithms related to graph neural network (GNN): GCN*[ICLR]*, RRN*[NeurIPS]*, GraphSAGE*[NeurIPS]*, PinSAGE*[KDD]*, etc.
- Existing algorithms utilize context features by either concating with user/item features (xDeepFM) or using simple attention mechanisms (DIN), which may not model the context well. We are working on constructing the context features as a graph that is integrated with user/item features by attention mechanisms, which may capture more global information.
- Graphs in reality are usually unbalanced, e.g, some items may be clicked by much more times than others, which may hinder a model to learn low-frequency items well. We are working on using GAN to reduce the frequency bias, which may help alleviate the cold start problem.

*Continue to the next page

==Log Mining==

** Most of the work is done when I was an intern in Huawei 2012 Lab.*

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 evaluated their accuracy, efficiency, and robustness on 16 datasets.
- We released the benchmark and datasets. The corresponding paper was **accepted by ICSE-SEIP 2019**.

Logzip

- We proposed logzip to optimize existing compression tools for log compression.
- We conducted iterative clustering to parse logs to generate intermediate representation that can be compressed with higher compression ratio by existing tools.
- Logzip is designed to be highly parallel. It got ∼**4.56x** the compression ratio of gzip and achieved comparable efficiency with gzip when using multiple workers. The corresponding paper was **accepted by ASE 2019**.

Huawei Phone Duplicate Issues Detection

- To detect duplicate issue reports (including issue descriptions and logs) from Huawei phone users.
- I was responsible for log parsing, log matching and feature extraction.
- The system went online and achieved more than **80%** accuracy.

==NLP==

Text Semantic Matching

- To compute the semantic similarity of two sentences.
- Investigated/implemented some deep learning-based algorithms: Decomposable Attention*[ACL, Google]*, BIMPM*[IJCAI]*, DSSM*[CIKM, Microsoft]*, etc.
- We proposed to build sentences as a graph and utilize information from adjacent nodes when inferring the similarity of two sentences, which slightly improved the accuracy of baseline models.

Text Classification [Graduation Project]

- To decide which Emoji to use in a sentence.
- We turned a classification problem into a matching problem by embedding labels (Emojis) to enrich the supervisory signal and improve model training.
- The work was rated as an excellent graduation project and similar work showed on ACL 2018.

♡ Honors and Awards

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

Skills

- Programming Languages/Frameworks: Python > C++ > Java / Pytorch == Keras > Tensorflow
- Open source projects: I am a member of LogPAI team (www.logpai.com) and a contributor of logparser where we implemented most state-of-the-art log parsing algorithms and gained more than **260** stars.

PaperList

- Jinyang Liu, Jieming Zhu, Shilin He, Pinjia He, Zibin Zheng, and Michael R. Lyu. Logzip: Extracting Hidden Structures via Iterative Clustering for Log Compression. To appear in Proceedings of the 34th IEEE/ACM International Conference on Automated Software Engineering (**ASE 2019**),

*Continue to the next page

- Jieming Zhu, Shilin He, Jinyang Liu, Pinjia He, Qi Xie, Zibin Zheng, Michael R. Lyu. Tools and Benchmarks for Automated Log Parsing. In Proceedings of the 41st International Conference on Software Engineering (**ICSE 2019**), SEIP track.
- Jinyang Liu, Jieming Zhu, Liang Chen, Gang Wang, Zibin Zheng, Yuzhou Zhang. Unifying Model Serving for CTR Prediction with Knowledge Distillation. *In submission*.
- Haicheng Xu, Jieming Zhu, Jinyang Liu, Zibin Zheng, Wuhui Chen. Linkage Embedding: Boosting FM-based CTR Prediction Models by Sharing Feature Embeddings. *In submission*.

REFEREES

- Prof. Michael R. Lyu, Chairman of the Computer Science and Engineering department, Chinese University of Hong Kong, lyu@cse.cuhk.edu.hk
- Prof. Zibin Zheng, Dean of the Software Engineering department, Sun Yat-Sen University, zhzibin@mail.sysu.edu.cn
- Dr. Jieming Zhu, Huawei Noah's Ark Lab, jamie.zhu@huawei.com

**END*