Wang Yang

Target: Software Engineer
Email: khil.wang@gmail.com
Cellphone: +86 13552133119

WORK EXPERIENCE

Senior Software Engineer, Tencent   2012 - present

1. Online Advertising Training System. 2015.1 - now

I developed a distributed model training system. The model is used to predict Click Through Rate(CTR) of Advertisement. The system is based on Storm/Kafka-like infrastructure. The system get streamed training data and then feed into one-pass algorithm(FTRL/AdPredictor/FM/MLP) and periodically dump the model for serving. In order to equally placed the memory consumed component I also designed custom scheduler of storm.

2. Training Data Generation System. 2013.1 - 2014.12

I developed training data generation system for both online and offline model training. Firstly I got the origin click/impression log from distributed queuing system(Tube), then use storm divided it into different Ad position and extract keys to join with user/ads/context features, and finally dump training data into HDFS and Tube for model training purpose. I use 32*8 CPU to handle 15 Billions impression for each cluster set(I have 2 set, one for production, another for experiment)

3. Online Training Monitor 2015.1 - 2015.5

I developed plenty of metric for monitoring task, for the model specific: Delay AUC, Real-time AUC, Loss, Miss-classification-rate, model-size; for the system specific: memory-used, batch-process-time. This project initially base on Graphite then migrate to a more featured monitor system.

4. Distributed Database Engine Optimizer 2012.6 - 2013.1

I developed Bool Expression Analyzer which is the fundamental component of Optimizer in SQL compiler. I implemented the Quine McCluskey algorithm to generate conjunctive normal form(CNF) and use Petrick Method to get Minimal-CNF. Based on the bool expression analyzer, I developed Predicate Push-down Optimizer for logical plan. And I also developed Reduce Shuffling Optimizer for physical plan.

Side Project

1. Aliyun Based API Platform for Cellphone Signaling Analyzing. 2016.9 - 2016.10

I developed a API platform based on Aliyun use Tornado as server. Encapsulate UDAF and SQL to simple API. `https://github.com/wang-yang/cell_signalling_analysis`

2. Online Image Caption Generation and Tech blog about RNN. 2016.6 - 2016.8

I developed a web service of image caption based on Neuraltalk2. `http://willwy.com`

EDUCATION

MSc. Computer Science   2009-2012
Chinese Academy of Science Software Institute
BSc. Telecommunication Engineering   2005-2009
Queen Mary, University of London
Beijing University of Posts Telecommunications

SKILLS

Languages   Java, Python, C++, Unix-like (*nix) Shell
Knowledge   Data structure and Algorithm, Machine Learning, Distributed System