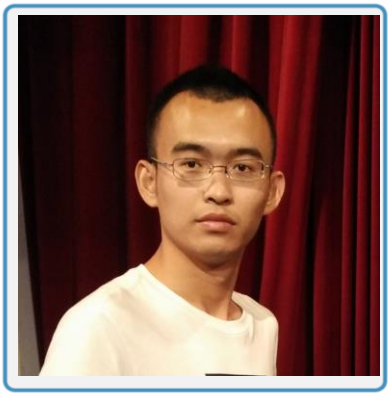
Yao Yu

Algorithm Engineer



Objective

I wish to do some related work on data mining, machine learning

Contact Info

#800, Dongchuan Road, Shanghai

(+86)17091830476

https://romanyu.github.io

csyuyao@sjtu.edu.cn

Education

- 2014 – present **Shanghai Jiao Tong University**
  master of computer science
  (expected graduation: Mar, 2017)
- 2010 –2014 **Central South University**
  bachelor of computer science

Awards

- 2015 Second-class Scholarship
  2-prize of 2015 hackathon of TouchPal
- 2014 Second-class Scholarship
- 2013 Outstanding Graduate
- 2012 National Encouragement Scholarship
  First-class Scholarship(GPA:86/100)
  Outstanding Student Leader
- 2011 Second-class Scholarship(GPA:83/100)
  Outstanding Student

Experience

Algorithm Engineer Intern

At dianping.com ♥ Shanghai

June, 2016 – Present

Focus on recommendation system, my responsibility is to provide better user experience of module "Guess You Like" of App. Our KPI is UV-CTR.

Data Development Intern

At TouchPal ♥ Shanghai

March, 2015 – Jan, 2016

Focus on big data processing, my responsibility is to provide operation data of Apps, like the number of new/upgrade users, DAU, retention rate, push messages.

Teaching Assistant

At Shanghai Jiao Tong University ♥ Shanghai

Sep, 2015 – Jan, 2016 Python

Sep, 2014 – July, 2015 Database

Projects

Item-similarity Recommendation (offline data)

Intro: http://romanyu.github.io/item_similarity_based_on_word2vec/

Recommendation strategy of module "Guess You Like" of App, computing the latent interesting item using item-based CF, firstly generate sword embedding using word2vec, then computing similarities between every two vectors. Improve KPI UV-CTR 3%

Technology: Hadoop, Java, word2vec

User-based Collaborative Filtering (offline data/realtime)

Intro: http://romanyu.github.io/user_based_collabrative_filtering/

Recommendation strategy of module "Guess You Like" of App, computing offline data using user-based CF, improve KPI UV-CTR 5%; Real-time strategy recommends item based on user similarities and 7-day click behaviors of k nearest neighbours, improve KPI UV-CTR 5%.

Technology: Hadoop, Hive

Real-time Computing Platform

Intro: http://romanyu.github.io/real-time-data-process-platform/

A real-time framework to calculate the number of new/upgrade users and DAU a.k.a daily active user, and some push messages statistics.

Technology: Spark, Cassandra, Kafka, Scala

Off-line Data Statistics

Computing DAU, retention rate, new/upgrade activations, etc.

Skills

| Java | Python | Data Structure | Algorithm| Database
SQL/NoSQL | Unix/Linux | Shell Scripts

Focus on related work on big data, familiar with Hadoop, Spark, Hbase, Hive, etc. familiar with frequently-used algorithm of data mining and machine learning