

Peng Zhou

Researcher Assistant at Tencent | jeffery.p.zhou@gmail.com
Advertising Engine & Data Algorithm Center, Shanghai, China

EDUCATION

- 2014.9~2017.3 **Tongji University**, Major:Software Engineering, GPA:82.1%,Degree:Master
- 2009.9~2013.6 **Nanjing Tech University**, Major:Computer Science, GPA:85%,Degree:Bachelor

PROJECT EXPERIENCE

Topic Modeling on APIs of Source Code Repository(2015.08~2017.3)—graduate graduation thesis

- proposed a hierarchical topic modeling method for open source APIs - AHTM, which contains
 - a.Topic Mining*: analyzed the parameters of LDA in Gibbs sampling mode and the physical meaning of the important parameters in the LDA.
 - b.API topic Filtering*: regarded the words as nodes and the relations between them as links to build a WordNet; filtered the API topics based on the relational density called centric mean in networks.
 - c.API topic Organizing*: discussed the influence of different number of topics on abstract levels of the APIs, and proposed a similarity algorithms to organize the different levels of topics.
- finally discussed the validity of AHTM by using the popular repositories in the field of data visualization, such as D3.js.

Criminal Network Analysis Based on Mobile Call Logs(2015.2~2016.8)

- suggested an analytical process with interactive strategies as a solution to analyse CNs constructed from the communication data which can be divided into three phases, namely, *a.network construction*: the generation of network structure,*b.metric design*: the core nodes and relations,*c.structure observation*: structures extraction in different levels.
- proposed interactive strategies to the analytical process, such as using various visualization layouts to configure the network, reinterpreting different measures from the social network domain to the CN domain, and controlling the community structure level with label supervision strategy.
- conducted a proof of concept study using mobile call logs.

Automatic labeling system for large-scale text training corpus — under Wechat User Profiles System(2016.6~now)

- Suggested threes methods to obtain the class seed in positive corpus:
 - a.AP Clustering*: selected the contents issued by different types of weChat official accountsas as the category of seeds of positive cases based on the results of AP Clustering in the article content;
 - b.PLSA*: the article was sorted by scoring according to the results of PLSA clustering;
 - c.Word2vec*: Combine the domain dictionary and Word2vec method to expand the word.
- applied bayesian model and logistic regression model to Co-Training algorithm and PU-Learning algorithm respectively, and the classifier would be more strong and better stability after multiple iterations.

Walrus: Advertising Big Data-Management Platform (2017.4~now)

- work on daily researches and developments of Walrus platform.
- translate the job using pig to job in spark in order to improve the efficiency for certain jobs.
- responsible for the module of monitoring the jobs running on walrus and sending their key logs to personal phones.
- finish the module of jobs' time consuming visualization using dynamic refreshing.

PUBLICATIONS/CONFERENCE PAPERS

Peng Zhou, Yan Liu, Mengjia Zhao and Xin Lou. "A Proof of Concept Study for Criminal Network Analysis with Interactive Strategies," *IJSEKE2017(International Journal of Software Engineering & Knowledge Engineering)* , May 2017.

Peng Zhou, Yan Liu, Mengjia Zhao and Xin Lou. "Criminal Network Analysis with Interactive Strategies: A Proof of Concept Study using Mobile Call Logs," *SEKE2016(The 28th International Conference on Software Engineering and Knowledge Engineering)*, July 2016.

Mengjia Zhao, Yan Liu, and Peng Zhou. "Towards a Systematic Approach to Graph Data Modeling: Scenario-based Design and Experiences." *SEKE2016(The 28th International Conference on Software Engineering and Knowledge Engineering)*, July 2016.

Huawei Liu, Peng Zhou and Yudi Tang. "Customizing Clothing Retrieval Based on Semantic Attributes and Learned Features" *MMM2017(The 23rd International Conference on Multimedia Modeling)*, Oct 2016. [Accepted but not having enough money to attend the conference, so finally rejected]

AWARDS & HONORS

National Postgraduate Mathematics Contest in Modeling – Second Prize

2016 Alibaba Music Fashion Trend prediction competition –

Microsoft Microduino intelligents: hardware training battalion

2015 First Tech Challenge(FTC) "Qualcomm Cup" robots competition – Final referee

China National Scholarship (3/60)

2016 The Excellent Student of Tongji University (5/60)

2017 The Outstanding graduate of Tongji University (3/60) – The First Honor

Tongji University Graduate Entrepreneurs Club president

Tongji University Venture Valley in Jiading Campus Operation Department Vice President

Global Entrepreneurship Week Campus Center (GCC) summit – member

China Certificate of Computer and Software Technology Proficiency – Software Designer

COURSERA verified:

Data Processing Using Python – NJU

Pattern Discovery in Data Mining – UIUC

Cluster Analysis in Data Mining – UIUC

SKILLS

Data mining:

Learned some basic knowlege about data mining and applied some classic mining algorithms.

common tools: scikit-learn, genism.

Complex Network Analysis:

Learned some basic knowledge about complex networks metrics and understood some algorithms for community detection.

common tools: Gephi, networkx, NEO4J

Loving python (most commonly used) and strong linux system programming skills.

Big Data Analysis:

well known the concepts of MapReduce, HDFS, Pig, Spark.

half yeas of developing and operating Data Management Platform.

one year of developing experience of Hadoop.

half year of developing experience of spark.

Engineering:

Backend python Django, php lavarel

Frontend known html,css,js, be able to use bootstrap,semanticUI.

Research Interests: Machine Learning, Data Mining and Complex Network Analysis