**Zhipeng Zhang**

**Gender: Male Date of Birth: Sept. 17, 1993**

**Contact: (+86) 15201199242 zhangzhipeng@pku.edu.cn**

**Address: Peking University, Beijing, China 100871**

**EDUCATIONAL BACKGROUND**

**Ph.D in Computer Science (Peking Univ.) Sep.2015 to Jul.2020**

* Research interests: Big data systems, distributed machine learning system.

**B.Sc in Computer Science (Shandong Univ.) Sep.2011 to Jul.2015**

* Rank 1/15 in Taishan College.

**ACADEMIC/WORK EXPERIENCE**

**Senior Algorithm Engineer (Internship, PAI@Ali) July.2020 to present**

* Robustness and usability of Alink; Machine learning on Flink (FlinkML)

**Algorithm Engineer (Internship, Tencent) Nov.2018 to Dec.2019**

* Distributed machine learning algorithms, distributed graph embedding.

**Visiting researcher (Systems group, ETH) Jul.2017 to Jan.2018**

* Accelerate machine learning on Spark.

**OPEN SOURCE PROJECTS**

**Apache Flink**

* Co-founder and maintain of Flink ML, committer of Apache Flink.

**RESEARCH PROJECTS**

**Graph computing.**

* A taxonomy of existing SimRank solutions, a theoretical analysis of the connections of different algorithms, and a guideline on how to choose a good SimRank solution given a certain workload.

**Distributed Machine Learning System.**

* A parameter server architecture on Apache Spark and a new data abstraction. 55.6X faster than MLlib and 3.3X faster than Petuum. (Collaborated with Tencent colleagues.)
* A large-scale and high dimensional machine learning system (Athena Project: <https://github.com/ghandzhipeng/Athena>)

**Distributed Machine Learning Algorithms.**

* Model Averaging and AllReduce to accelerate GLMs and LDA. 3-orders faster than MLlib on GLM, 5X faster than Angel on LDA.
* A column-oriented framework for distributed stochastic gradient descent. 930X faster than MLlib, 63X faster than Petuum, 14X faster than MXNet.

**PUBLICATIONS**

* **Zhipeng Zhang**, Wentao Wu, Jiawei Jiang, Lele Yu, Bin Cui, Ce Zhang, ColumnSGD: A Column-oriented Framework for Distributed Stochastic Gradient Descent [CCF-A, **ICDE** 2020]
* **Zhipeng Zhang**, Bin Cui, Yingxia Shao, Lele Yu, Jiawei Jiang, Xupeng Miao, PS2: Parameter Server on Spark [CCF-A, **SIGMOD** Industry 2019]
* **Zhipeng Zhang**, Jiawei Jiang, Wentao Wu, Ce Zhang, Lele Yu, Bin Cui. MLlib\*: Fast Training of GLMs using Spark MLlib [CCF-A, **ICDE** Industry 2019]
* **Zhipeng Zhang**, Jiawei Jiang, Lele Yu, Bin Cui. Angel+: A Large-Scale Machine Learning Platform on Angel [Frontiers of Data and Computing. 2019]
* **Zhipeng Zhang**, Yingxia Shao, Bin Cui, Ce Zhang. An Experimental Evaluation of SimRankbased Similarity Search Algorithms [CCF-A, **VLDB** 2017]
* Yunyan Guo**, Zhipeng Zhang**, Wentao Wu, Jiawei Jiang,Ce zhang, Bin Cui, Jianzhong Li, Model Averaging in Distributed Machine Learning: A Case Study with Apache Spark [Submitted to **VLDBJ** 2021]
* Xiaoru Qu, Zhao Li, Jialin Wang, **Zhipeng Zhang**, …, Jun Gao. Category-aware Graph Neural Networks for Improving E-commerce Review Helpfulness Prediction. [**CIKM 2020, CCF-B**]
* Jiawei Jiang, Pin Xiao, Lele Yu, Xiaosen Li, Jiefeng Cheng, Xupeng Miao, **Zhipeng Zhang**, Bin Cui. PSGraph: How Tencent trains large-scale graphs with Spark? [CCF-A, ICDE 2020]
* Xinyi Zhang, **Zhipeng Zhang**, Bin Cui. An Reinforcement Learning-based Method for Join Optimization. [CCF-C, **NDBC** 2020, **Best Student Paper**]
* Jiawei Jiang, **Zhipeng Zhang**, Bin Cui, Yunhai Tong, Ning Xu, StroMAX: Partitioning-based Scheduler for Real-time Stream Processing System [CCF B, **DASFAA** 2017]
* Shuyang Shi, **Zhipeng Zhang**, Bin Cui. Resume Activeness Prediction in Online Recruitment Scenarios [CCF C, **NDBC** 2017]

**SKILLS**

**Professional Skills**

★★★★★Flink, Spark, Parameter Servers; Java, Scala; latex, shell, git, etc.

★★★★ TensorFlow, Petuum, MXNet; C/C++, Python; yarn, docker, etc.

**Language Skills**

★★★★★ English

**HOBBIES**

Walking, Reading, Fitting, Movies.

**SELF-ASSESSMENT**

Detail-oriented, team-spirited, responsible, and self-motivated.