

# 李萍

目标岗位：算法工程师

✉ lipingict@gmail.com · ☎ (+86) 130-200-48987 · 🌐 https://qiuwuyiye.github.io

## 🎓 教育背景

中国科学院计算技术研究所，计算机，免试保送，北京	2014 – 2017
吉林大学，计算机科学与技术，1%，长春，吉林	2010 – 2014

## 👨‍💻 实习经历

✓ 百度凤巢	数据挖掘工程师	2015 年 5 月 – 2015 年 8 月
--------	---------	-------------------------

项目概述 在 FCR\_Model 实习期间主要负责模型评估、特征挑选以及特征调研。

- 特征分析工具: 基于凤巢特征抽取系统 Adfea 和模型训练系统 Platform 开发一个特征评价工具以助力特征挑选。
- Sug 策略评测: 评估时间衰减策略在不同衰减因子以及时间窗口上的 Sug CPM 预测效果。
- 特征调研: 调研 Kaggle CTR Prediction 比赛中优胜选手使用的特征、连续特征值的处理方式以及使用的模型。

✓ 微软亚洲工程院	研发工程师	2017 年 4 月 – 2017 年 5 月
-----------	-------	-------------------------

项目概述 负责 Bing Map 中 Chain Store 数据的采集以及 Cortana Answer 在若干应用场景下的 Bug 修复。

- Chain Store 数据采集: 构建连锁店基本信息采集 workflow，目前可稳定为 Bing Map 提供连锁店数据。
- Cortana Answer Bug 修复: 修复了 Cortana Answer 中天气以及连续剧等应用场景下的若干 Bug。

## 👨‍💻 项目经历

✓ 图形化大数据机器学习平台 BDA Studio	2015 年 8 月 – 2016 年 10 月
---------------------------	--------------------------

项目概述 图形化机器学习平台由可拖拽大数据机器学习平台 BDA Studio 及大数据机器学习库 BDALib 构成。

- 图算法: 开发三个图算法 (Pagerank, ICmodel, KShell), 相比于 graphx 原生算法可收敛、可扩展性好、速度快, 可支持上 10 亿规模顶点的图数据挖掘。
- 推荐算法: 实现 Factorization Machine 和 NMF 算法的单机及分布式版本, 在 movie-lens 上 RMSE 约 0.80。
- ETL: 完成 BDA Studio 的 ETL 功能, 支持 Mysql, Hive 等异源数据的导入。

✓ 大规模矩阵分解算法研究与实现	2016 年 11 月 – 2017 年 4 月
------------------	--------------------------

毕业论文 研究及实现基于图计算框架的大规模矩阵分解算法。

- 在线图划分算法: 提出并实现了在线图分割算法 BiEdgePartition2D, 在保持子图平衡性的同时降低了复制因子。
- 基于模型分布的分布式矩阵分解实现: 将待分解矩阵用二部图的形式存储, 利用 BiEdgePartition2D 算法进行分割, 基于 Pregel 图计算框架实现矩阵分解算法, 最终减少了约 20% 的平均迭代时间。

## 🏆 比赛经历

✓ 天翼大数据算法应用大赛	冠军 (1/1111)	2015 年 12 月 – 2016 年 3 月
---------------	-------------	--------------------------

- 概述: 使用前 7 周用户每天点击 10 个视频网站的统计数据, 预测用户第八周每天点击视频网站的数量。
- 职责: 特征调研, 特征抽取, 特征评价。

✓ Google Girl Hackthon 2017	最有影响力奖	2017 年 3 月
-----------------------------	--------	------------

- 概述: 在三天时间内独自设计并实现了基于表达学习的人才检索系统 Talent Search。
- 模型训练: 基于表达学习, 将用户的评论信息和人才需求映射到相同的空间, 进行人才匹配。

## ♥ 论文及获奖情况

分布式算法实现比较研究: 数据分布与模型分布, CCIR2016	2016 年 6 月
Ease the Process of Machine Learning with Dataflow, CIKM2016	2016 年 5 月
Session Segmentation Method Based on Naive Bayes, ISTP 检索	2012 年 9 月
Session Segmentation Method Based on COBWEB, EI 检索	2012 年 6 月
连续三年国家奖学金, 吉林大学	2011 年 – 2013 年