

张随远

✉ a_dayuanzi@163.com · ☎ (+86) 136-5115-6109

教育背景

中国科学院计算技术研究所 武汉理工大学	网络数据重点实验室 计算机学院	硕士 (推免) 学士 (2/110)	2015 – 至今 2011 – 2015
------------------------	--------------------	-----------------------	--------------------------

获奖情况

天池-智慧交通预测挑战赛	1 / 1716	2017.07 – 2017.08
用户流失率预测	1 / 200+	2016.07 – 2016.09
酒店未来 30 天产量预测	1 / 200+	2016.07 – 2016.09
Kaggle-Quora 文本匹配相似度大赛	31 / 3304	2017.04 – 2017.06
天池-阿里音乐流行趋势预测大赛	16 / 5475	2016.05 – 2016.07
全国大学生数学建模竞赛	一等奖 (top 1%)	2013.09

实习经历

滴滴出行科技有限公司	地图事业部	算法工程师新锐实习生	2017.06 – 至今
<ul style="list-style-type: none">熟悉了滴滴司机端送客线路推荐框架, 利用 <code>pyspark</code> 在海量数据下进行开发实验建立道路 <code>link</code> 拓扑结构, 利用图卷积神经网络模型 (GCN) 进行路况挖掘			
中国平安集团	万江龙商业地产	室内定位工程师实习生	2014.09 – 2015.01

项目经历

分布式检索系统	核心成员	2016.09 – 至今
---------	------	--------------

c++ 编写, 已上线到多个实际应用场景, 多个系统的底层系统

- 倒排索引的数据结构, 支持实时索引, 一篇文档从开始建索引到被检索到不超过 1 分钟
- 支持节点热插拔, 负载均衡。支持文档分片, 提高效率, 采用 `Thift` 框架通讯更快
- 采用 `mmseg+HMM` 的分词算法, 基于余弦相似度作为排序依据, 支持 `DBSCAN` 算法对结果聚类

携程云海-科赛大数据竞赛	队长	2016.07 – 2016.09
--------------	----	-------------------

根据历史一年的酒店夜量, 预测 500 家酒店在 2016 年 2 月份的每天的需求量

- 分步建模, 将问题抽象为两个子问题, 选取酒店属性、对历史数据进行窗口化的统计特征、酒店 `id` 等
- 设计针对星期窗口化的双分裂点选取的随机森林算法, 较随机森林和 `XGBoost` 有显著提高

Trec 2017 OpenSearch	队长	2017.04 – 至今
----------------------	----	--------------

根据论文检索网站 `SSOAR` 的反馈信息的论文排序系统, 分为 `Online` 和 `Offline` 两部分, 同样的特征工程与模型在刚刚结束的 `Kaggle` 平台 `Quora` 问题匹配大赛获得成绩 0.133 (`logloss`), 31/3304

- 主题模型采用 `SVD`、`NMF` 矩阵分解算法, `LDA` 主题模型, 利用 `wordnet` 抽取同义词、近义词特征
- `WordToVector`、`DocToVector`、`LDAToVector` 算法将句子转化成向量表达
- 基于 `ngram` 采用余弦相似度、编辑距离、`jaccard` 距离等, 结合句法依存关系提取地点、数字、疑问词等特征
- `LightGBM` 直方图算法速度更快、样本分布更均衡, 尝试 `FFM` 和 `FM` 算法来学习组合特征
- 基于时序建模的 `LSTM` 和基于卷积的 `CDSSM` 深度神经网络, 模型融合采用 `stacking` 方式
- 基于 `PairWise` 的排序学习模型采用 `LambdaMart`, 在线学习采用 `DBGD` 算法, 以满足实时性需求

IT 技能

- 编程语言: `Python`、`C++`、`Java`

技术分享

- 技术博客: <http://blog.csdn.net/chedan541300521> (`XGBoost` 源码阅读笔记)
- `GitHub`: <https://github.com/dayuanzi> (双分裂点随机森林算法)