

教育经历

2014.09-至今	大连理工大学	数据挖掘	硕士
	免试保送（推免排名第1）		
2010.09-2014.7	大连理工大学	软件工程	学士
	专业排名3.3%		

主要能力

- **编程语言:** 了解C++, Python。接触过C#, Java。
- **算法:** 掌握常用的基础算法。
- **机器学习:** 常用机器学习算法, 包括分类、聚类、关联规则、推荐、主题模型等。

科研及项目经历

2014.11-2015.05 Conditional Discriminative Pattern Mining: Concepts and Algorithms

- 差异模式是一种在正项数据集频繁, 在反项数据集不频繁的模式。与频繁项集挖掘类似, 普通的差异模式挖掘算法往往会产生冗余模式。为此, 本研究提出了一种基于FP-Growth的去除冗余的算法。我主要负责算法设计、实验和论文撰写。目前已投**Information Science**期刊(CCF B, SCI IF = 3.364), 已按审稿人要求修改两次。

2014.3-2014.7 基于置换检验的聚类结果评估

- 针对聚类结果的评估。我们提出了一种基于置换检验的评估方法, ECP。该算法直接对聚类结果进行置换检验从而计算出p-value。我主要负责算法设计、实验和论文撰写。该论文已被《智能系统学报》(中文核心期刊)录用。

2012.11-2012.12 汉字识别

- 该项目对所写汉字形成的图片经行灰度化、二值化处理, 然后从各个方向提取特征向量。用KNN分类算法对汉字进行识别。 我主要负责算法的设计和识别汉字部分代码的编写。

实习经历

2016. 05-至今 滴滴研究院·策略七组

- 滴滴订单数目预测。将每个城市划分为几千个直径为1km的正六边形格子, 预测每个格子未来十分钟的订单数目。该项目利用每个格子的历史订单数目构造了66个属性, 用线性回归进行预测。误差率为22.7%。我在该项目中负责数据处理、模型的训练调优等。该项目现已上线。

2016. 01-2016. 05 亚马逊·Capacity Plan

- Economic Operation Plan. 该项目考虑库房容量、劳动力成本、7商品在各个环节运输成本等条件简历线性规划的模型为美国库房制定供应链计划。 我在项目中主要对数据的缺失和异常做填充和平滑, 开发模型评估工具。该项目现已上线。

2013. 10-2014. 02 百度·商务搜索部·反作弊

- 人工评估平台脚本开发(主要有PC检索、无线检索、PC点击, 无线点击、网盟等模块)。项目用Map Reduce对点击检索日志进行数据抽样并统计处理, 评估反作弊略, 提高反作弊的效率。我主要负责了平台的部分脚本开发。

2013. 06-2013. 09 微软·STC部·Bing

- 天猫商家数据实时抓取分析。针对天猫一商家每五秒钟抓取一次其在量子恒道的数据, 将抓到的数据导入数据库, 对数据进行处理分析, 并用网页将结果展示。 我负责项目的设计与开发。

个人博客

<http://www.cnblogs.com/gufeyang/>