

“卷皮网”用户画像及精准推荐实战

大学生创新训练项目中期答辩

答辩人：孙嘉轩 黄爽 龙海文 杨岳浩 许艳

指导老师：张千帆

华中科技大学管理学院

2018 年 10 月 14 日

**”There is nothing about doing data analysis that is neutral.
What and how data is collected, how the data is cleaned
and stored, what models are constructed, and what questions are asked
– all of this is political.”**

Danah Boyd, NYU

目录

- 1 项目进展
- 2 现阶段问题
- 3 财务情况
- 4 后续工作安排



专业知识与技能

- 机器学习与数据挖掘基本概念与使用方法
- 从商业分析角度探究用户画像在企业中的应用
- 使用R/Python等编程语言进行分析与研究



企业合作

- 与卷皮网员工进行交流
- 爬取并分析卷皮网相关数据
- 发现数据潜在商业价值与市场应用



方案优化

项目执行中的问题

- 卷皮网购物数据质量较低，用户类型单一，不具有普遍意义。
- 一个项目无法全面展现机器学习中丰富的研究方法。
- 不同应用场景下的用户画像有很大区别。
- 报告的形式无法很好的呈现研究结果。

方案优化

解决方案与改进措施

- 增加研究领域，扩大研究范围，全方位解读用户画像的企业应用场景。
- 增加对于不同数据类型、不同方法的对比研究。对比不同编程语言及程序包的优劣。
- 通过建立网页版教程，系统归纳团队研究过程及心得，给入门用户提供学习的平台。
- 编写入门级机器学习实践案例，增强教程的实践性与互动性。

前言

用户画像从何而来

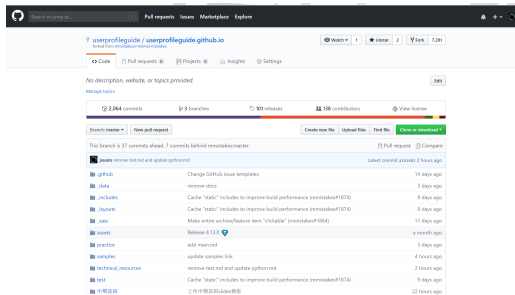
在大数据、智能算法蓬勃发展的当今，传统业态正在面临新技术的冲击而重新洗牌，用户画像应运而生。人们开始关注如何从海量的数据中挖掘商业价值，切实的通过大数据的方法整合，优化用户体验，增加企业盈利能力，打造品牌价值。通过一个数据完美的抽象出一个用户的信息全貌，这用户画像最原始的想法，也是企业应用大数据的根基。

什么是用户画像

用户画像是根据用户的属性、偏好、习惯、用户行为等信息抽象出的标签化用户模型，通俗地说就是给用户打标签，而标签是通过对用户信息分析而来的高度精练的特征标识。

如今，用户画像被广泛的使用和研究，现在已经产生的较为成熟的方法包括：

userprofileguide.github.io



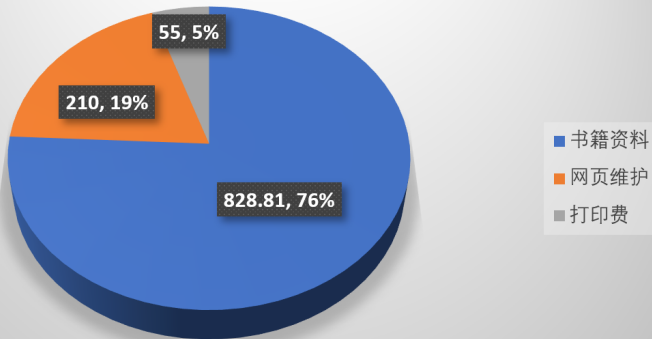
教程框架设计

- 技术原理：对概念和基础知识点的介绍、程序包的使用方法介绍、算法原理介绍等。
- 案例分析：全面记录团队项目开发过程和在其中使用到的研究方法、编程技巧、优化思路，并进行探讨比较。
- 入门练习：提供让入门读者快速上手实践并理解机器学习和用户画像的案例，数据与源代码。

现阶段问题

- 企业数据获得较为困难，企业严格限制API的调用。
- 已有中文语义分析技术尚不成熟，对中文用户刻画精确度较低。
- 尚未明确针对不同行业设计数据清理方案的思路。

财务情况统计



后续工作安排

- 拓宽研究领域，完成更多项目开发案例。
- 继续完善教程主体部分和延申部分，增强内容的完整度与可读性。
- 与企业进行合作，进一步探讨用户画像的商业价值与伦理学边界。
- 扩大影响力，打造可持续发展编程社区，不断引入更多优质内容，吸引越来越多的同行加入。

感谢聆听!

