
title: 大数据分析：某编程教育活动

tag: BigData

date: 2018-03-12 09:38:08

- 活动前
- 目的：从各个维度全方位地反映本次活动的成果。

需要设计的数据采集点：

统计本次活动学生参与总人数，总代码量。

参赛人群分析：地区占比，性别占比，年级占比，学校占比。

活动分析：点赞数，浏览数，参赛作品提交时间段分析，新增用户比（与往期活动纵向对比）

作品分析：**关键在于如何确定优秀作品**，点赞排名，收藏排名，作品tags（**在比赛中作品标签类别属于必填项**，对于没有标记tags的作品，通过机器学习进行分类？？）类型比重分析，作品改编树关系网络分析

参赛学校分析：学校数据加权分析（**需要获取参与其中的学校名单列表**，对同一学校的所有作品的点赞数之和、浏览数之和进行排名）、学校总人数参与比重（**需要获取当前学校学生总数**）、教师参与度（本次活动指导教师占全体教师的比重，**需要获取学校教师数量**），学校之间距离远近（不是直线距离而是实际交通距离）与学校之间参与度之间的关系（互助，竞争关系？），学生作品水平与当地教育水平分布图（经济情况，教育机构分布情况，交通情况）之间的关系

学生分析：重点分析点赞排名和收藏排名靠前的学生群体，建立这些学生之间的关系网络模型，分析该学生对不同作品的投稿数，确定用户画像（兴趣、爱好，创造性等指标，宏观分析用户编程习惯特征），横向对比同一班级的学生之间的兴趣爱好分布情况

班级分析：建立班级和教师关系模型，比如同一个老师教的不同班级学生（**需要获取教师教课表（每名教师所带的班级统计信息）**），班级课内学习成绩关系和活动参与程度的关系图（了解scratch编程是否有助于提高学生的学习能力，这个需要在时间上纵向开展，跟踪分析）。

教师分析：对相同指导教师的作品的统计分析：如作品整体水平、质量（点赞数，浏览数），建立优秀指导教师（优秀作品排名）之间（**所在学校，年龄，性别**）的关系网络分析模型

统计编程用时：**这里需要增加一个统计学生累计编程时长的字段（有效时长指处于编程界面下鼠标键盘有操作的情况下）**。后续功能可以累计增加经验值。

作品转化率：作品a被改编的次数/作品a被浏览次数

1. 针对首次编程周活动统计中的问题提出的建议：
2. 提交作品请不要使用“新项目”等默认名称
3. 提交作品请选择作品标签类型（使用复选框，并增加提示信息）
4. 请用户完善个人信息：性别，年级，生日，**兴趣爱好（至少填写一个兴趣爱好，为了便于统计这个可以使用多选的方式进行填写，作为构建用户画像的参考之一）**
5. 明确本次参与活动是以第一学校还是第二课堂学校作为入口，老师要明确告知学生提交作品时所选的班级。

最终，宏观统计：总数aa个学校的bb个学生在累计yy小时创作了zz个作品，累计代码行数为nn。