

# 基于大数据个性化宿舍分配系统设计

冯盛豪, 李志文

(广州大学数学与信息科学学院, 广州 510006)

**摘要:** 随着高校招生规模的不断扩大和大学生人数的快速增加, 高校后勤的宿舍管理正面临着日益严峻的挑战。为了有效应对这一问题, 满足学生日益多样化的需求, 同时减轻宿舍管理人员的工作负担, 提高宿舍分配的质量成为解决宿舍管理难题的重要途径。本文通过对大学生宿舍关系的影响因素进行预调查与文献调研, 设计了一套基于学生个性习惯的宿舍分配系统。在宿舍分配的核心算法选择上, 本文采用了基于数据的 K-means 聚类算法。这一算法能够有效地处理多维数据, 为宿舍分配提供科学而合理的依据。我们将基于学生自身偏好习惯的数据进行 K-means 聚类, 划分出习惯相近的人群, 在将他们分配到一个宿舍内, 实现基于个人习惯的宿舍分配。我们的目标是实现对学生个性化需求的满足, 为此将提供高效而智能的宿舍分配系统。这项系统可以解决当前宿舍管理面临的难题。

**关键字:** 个性化; 宿舍分配; K-Means 聚类算法

## 1 引言介绍

“大数据正在改变我们的生活以及理解世界的方式。”[1] 随着信息技术的快速发展, 计算机技术已经被运用到各个领域, 极大促进了生产力的提升, 也在高等院校的信息化管理中凸显了极大的便利。大学的学生宿舍是我国大学生日常学习生活的主要场所之一, 因此, 良好的宿舍氛围建设是大学后勤建设的重要组成部分, 并且对高校大学生的日后自身发展以及身心养成有着举足轻重的作用。[2] 随着高校招生规模的不断扩大, 传统的宿舍分配方式已经显得有些滞后, 无法满足现代学生对宿舍分配的需求。这导致学校后勤需要投入大量资源来解决因宿舍分配不合理所带来的各种问题。如何利用计算机强大的数据分析能力实现智能化宿舍分配是当下宿舍管理中相当值得思考的问题。因而在大学信息化建设的时代背景下, 宿舍管理工作需要逐渐转变意识、改变服务模式, 进一步改革加强后勤宿舍管理工作。[3]

为了提高后勤管理的效率并切合学生的需求, 我们需要转变观念, 改变服务模式, 通过对学生的个性化偏好进行预调查和信息收集来规避这些问题。在这个信息化建设的时代, 对宿舍管理进行改革是必然趋势, 而其中关键的一步就是在学生正式入住前了解他们的住宿需求和偏好。这种方法不仅可以提升后勤管理的效率, 也能推动校园的信息化建设。

因此, 本文旨在基于学生个性化偏好建立智能的大学宿舍分配系统, 这将为宿舍管理带来前所未有的改善, 并且对学生的舒适度和学习环境的优化有着积极的影响。

## 2 算法基础

### 2.1 K-means 聚类算法

1967 年 James MacQueen 提出的  $k$  均值聚类算法 (K-Means Clustering Algorithm) 是一种迭代求解的聚类分析算法 [4][5] 其步骤是, 先随机选取  $K$  个对象作为初始的聚类中心。然后计算每个对象与各个种子聚类中心之间的距离, 把每个对象分配给距离它最近的聚类中心。K-means 聚类算法中  $k$  表示聚簇的个数, 聚类结果取决于  $k$  值的选取。[6] 聚类中心以及分配给它们的对象就代表一个聚类。一旦全部对象都被分配了, 每个聚类的聚类中心会根据聚类中现有的对象被重新计算。这个过程将不断重复直到满足某个终止条件。终止条件可以是以下任何一个 [7]:

- 没有 (或最小数目) 对象被重新分配给不同的聚类。
- 没有 (或最小数目) 聚类中心再发生变化。
- 误差平方和局部最小。

K-means 算法将一组  $N$  样本  $X$  划分成  $K$  不相交的簇  $C$ , 每个都用该簇中的样本的均值  $\mu_j$  描述。这个均值 (means) 通常被称为簇的“质心 (centroids)”; 注意, 它们一般不是从  $X$  中挑选出的点, 虽然它们是处在同一个空间。

K-means (K-均值) 算法旨在选择一个质心, 能够最小化惯性或簇内平方和的标准:

$$\sum_{i=0}^N \min_{u_j \in C} (\|x_j - u_j\|^2)$$

## 2.2 一次分类

我们先设置一个循环,将聚类数  $K$  的取值范围定在  $[2,10]$  区间,然后依次选取区间中的  $K$  值。随机选取  $K$  个样本作为初始的聚类中心。接着计算每个样本与各个聚类中心之间的距离,把每个样本分配给距离它最近的聚类中心。然后将新的聚类中心移动到这个聚类所有样本的平均值处。一直重复这个过程,直到聚类中心不再移动为止。最后求出这组数据集的轮廓系数。将不同  $K$  值的轮廓系数进行比较,选取出轮廓系数最大的  $K$  值作为初始选取样本个数。设置一个标签  $A$  将同一族类的学生记为相同值。

## 2.3 二次分类

在一次分类中,已经将学生分类成  $K$  类性格习惯相同的族类,接着我们需要将这些相同族类的学生分配进宿舍(也就是对标签  $A$  值相同的学生进行分配)。选定宿舍人数为  $N$ ,我们就需要将性格习惯相同族类按宿舍学生数为  $N$  进行宿舍分配。假如这个族类有五十个学生,每个宿舍六个人,则将会分配出九个宿舍,第九个宿舍暂时只有两位同学(后期管理员可以通过对剩余学生再次聚类或其它方式进行分配调整)。将每一族类按此方式进行分配,这样使得学生都能分配到性格习惯最相近的宿舍。

# 3 宿舍智能分配系统的设计

## 3.1 基本框架

本系统采用一个轻量级的 web 框架 Flask,其使用 Python 语言编写,较其他同类型框架更为灵活、轻便、安全且容易上手。可在短时间内完成功能丰富的中小型网站或 Web 服务的实现。同时可以根据需求来添加相应的功能,如加入 pymysql 库实现系统连接数据库。在保持核心功能简单的同时实现功能的丰富与扩展。

## 3.2 web 前端设计

系统使用 Flask 框架中 Jinjia2 模板,根据需求设计出对应的网页。同时结合 Bootstrap 相关组件完成页面前端设计。web 前端主要有设计了五类模板

- (1) Login.html, 用户登录页面
- (2) Select.html, 根据用户为学生或管理员提供对应的选择页面
- (3) Habit.html, 学生问卷填写页面
- (4) Display.html, 根据用户为学生或管理员提供对应的显示页面
- (5) Out.html 根据用户为学生或管理员提供结束页面

## 3.3 问卷的设计

准确地对待调查大学生发布调查问卷,及时返回学生填写的自身偏好数据,并将其存入后勤服务宿舍分配数据库中,以保证宿舍分配的高效性与准确性 [8] 因此我们先对学生的姓名、性别等基本信息进行采集,用于收取全部学生基本信息。除去基本信息外问卷共有 11 道题,用于收取各种生活习惯数据,如睡觉的时间,兴趣爱好等。每道题中都有对应的选项,每个选项会对应一个数值(如 a,b,c,d 对应值 0-3)。问卷填写完成后会将数据传入到数据库中,将得到的数据先按性别进行分类。这些对应数值将会是宿舍个性化分配的基础数据。

## 3.4 Mysql 数据库设计

MySQL 数据库负责将网页中的数据进行收集、存储、处理等操作。同时为宿舍智能分配系统提供信息来源。本系统设计了如下三个数据表

- (1) 学生基本信息总表, 字段包括学号, 姓名, 性别

- (2) 男性学生详细信息表, 字段包括学号, 姓名, 性别. 爱好 3-爱好 14, 第一次标签, 第二次标签, 宿舍楼栋, 宿舍号, 床号
- (3) 女性学生详细信息表, 字段包括学号, 姓名, 性别. 爱好 3-爱好 14, 第一次标签, 第二次标签, 宿舍楼栋, 宿舍号, 床号

### 3.5 宿舍分配系统的设计

- (1) 系统分析: 本系统设计的宿舍智能分配分为两个身份: 学生、管理员, 每个身份所具有的功能不同。当成功系统登陆后, 系统会根据输入的账号识别登录用户为学生还是管理员。然后根据身份转向不同的页面。用户为学生系统会转向问卷填写与查看宿舍分配情况页面。用户为管理员系统会转向宿舍分配与查看页面。
- (2) 学生端功能模块: 本学生模块包括问卷填写、查看宿舍分配情况这两个基本模块。问卷填写完后系统会将数据传入数据库中, 查看宿舍分配情况时系统将数据库中分配好的结果显现在网页中。具体流程如图1所示:
- (3) 管理员端功能模块: 本管理员模块由分配宿舍, 查看宿舍分配情况这两个基本模块组成。管理人员可以在查看宿舍分配情况后选择重新分配。具体流程如图1所示:

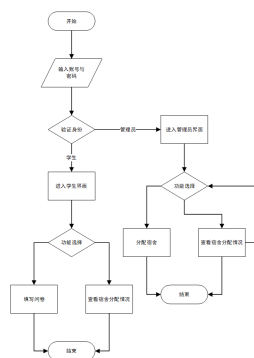


图 1: 工作流程图

## 4 宿舍分配系统的实现

本系统前后端分离的方式。前端界面使用 flask 框架及 Bootstrap 组件, 后端以 python 为开发语言, 结合 MySQL 数据库实现对数据的修改操作。

### 4.1 学生端功能的实现

学生进入该系统成功登录后, 新页面出现两项功能, 可以选择填写问卷和查看宿舍分配情况。问卷页面如图2所示:

**个人宿舍习惯调查**

注: 一人一份问卷, 填写一次, 数据将自动记录到数据库并生成统计图表

1. 学号	<input type="text"/>
2. 姓名	<input type="text"/>
3. 性别	<input type="radio"/> 男 <input type="radio"/> 女
4. 你期望宿舍的卫生环境(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
5. 宿舍的设施(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
6. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
7. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
8. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
9. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
10. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
11. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
12. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
13. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
14. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
15. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
16. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
17. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
18. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
19. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分
20. 宿舍的室友(满分5分)?	<input type="radio"/> 1分 <input type="radio"/> 2分 <input type="radio"/> 3分 <input type="radio"/> 4分 <input type="radio"/> 5分

图 2: 学生问卷 (部分截图)

当管理员完成宿舍分配后, 学生可以成功登录系统后, 点击查看宿舍分配情况。可以看到自己分配到宿舍的具体信息。如舍友, 床号, 宿舍号等。如图3所示:

宿舍分配情况			
学号: 32115217	- 姓名: 张三17	- 宿舍楼号: B1	- 宿舍号: 302 - 床号: 3
学号: 32115216	- 姓名: 张三16	- 宿舍楼号: B1	- 宿舍号: 302 - 床号: 4
学号: 32115211	- 姓名: 张三11	- 宿舍楼号: B1	- 宿舍号: 302 - 床号: 1
学号: 32115210	- 姓名: 张三10	- 宿舍楼号: B1	- 宿舍号: 302 - 床号: 2

图 3: 学生视角宿舍信息

4.2 管理员端功能的实现

管理员进入该系统成功登录后，新页面出现两项功能，可以选择进行宿舍分配和查看宿舍分配情况。当所有学生完成问卷填写后，可以进行宿舍分配操作，进行宿舍分配。宿舍分配时可以指定宿舍楼层数，楼层宿舍个数，宿舍人数。如图4所示：

请选择要分配宿舍性别

☐男

☐女

☒男和女

请输入每栋宿舍楼层数

请输入一层楼宿舍个数

请输入每个宿舍分配人数

分配

重置

图 4: 宿舍分配

管理员完成宿舍分配后，可返回上级页面点击查看宿舍分配情况功能，然后会显示学生宿舍分配情况。如图5所示。查看宿舍分配情况页面可以返回上级重新分配。

返回上级			
男宿舍分配情况			
学号: 32115206	- 姓名: 张三6	- 宿舍楼号: B1	- 宿舍号: 201 - 床号: 2
学号: 32115204	- 姓名: 张三4	- 宿舍楼号: B1	- 宿舍号: 201 - 床号: 3
学号: 32115225	- 姓名: 张三25	- 宿舍楼号: B1	- 宿舍号: 201 - 床号: 4
学号: 32115218	- 姓名: 张三18	- 宿舍楼号: B1	- 宿舍号: 201 - 床号: 5
学号: 32115205	- 姓名: 张三5	- 宿舍楼号: B1	- 宿舍号: 201 - 床号: 1
学号: 32115207	- 姓名: 张三7	- 宿舍楼号: B1	- 宿舍号: 202 - 床号: 1
学号: 32115212	- 姓名: 张三12	- 宿舍楼号: B1	- 宿舍号: 202 - 床号: 2
学号: 32115215	- 姓名: 张三15	- 宿舍楼号: B1	- 宿舍号: 203 - 床号: 1
学号: 32115221	- 姓名: 张三21	- 宿舍楼号: B1	- 宿舍号: 203 - 床号: 2
学号: 32115201	- 姓名: 张三1	- 宿舍楼号: B1	- 宿舍号: 204 - 床号: 3
学号: 32115202	- 姓名: 张三2	- 宿舍楼号: B1	- 宿舍号: 204 - 床号: 1
学号: 32115203	- 姓名: 张三3	- 宿舍楼号: B1	- 宿舍号: 204 - 床号: 2
学号: 32115228	- 姓名: 张三28	- 宿舍楼号: B1	- 宿舍号: 205 - 床号: 1
学号: 32115224	- 姓名: 张三24	- 宿舍楼号: B1	- 宿舍号: 205 - 床号: 2
学号: 32115223	- 姓名: 张三23	- 宿舍楼号: B1	- 宿舍号: 205 - 床号: 3
学号: 32115222	- 姓名: 张三22	- 宿舍楼号: B1	- 宿舍号: 205 - 床号: 4
学号: 32115223	- 姓名: 张三23	- 宿舍楼号: B1	- 宿舍号: 205 - 床号: 5

图 5: 管理员视角宿舍信息

5 结语

通过运用 python-web 编程技术和 k-means 聚类思想，并结合 Mysql 数据库，成功设计了一套基于大数据个性化宿舍分配系统。这一创新性的方案采用了综合考虑学生的基本信息、性格、爱好等多方面属性，将相似特征的学生智能地分配至同一个宿舍房间，充分尊重了每个学生的个体差异。为高校后勤管理迈向更加智能、高效的方向奠定了基础。

## 参考文献

- [1] 维克托·迈尔-舍恩伯格、肯尼思·库克耶; 盛杨燕、周涛译. 大数据时代 [M]. 浙江人民出版社, 2014: 15.
- [2] 李镜焕韩慧萍. 基于学生偏好的宿舍分配系统的设计与构建 [J]. 高校后勤研究, 2023(08): 28-31
- [3] 谢泮. 当前高校学生宿舍管理的问题和对策分析 [J]. 成都大学学报 (育科学版), 2007(08): 14-16.
- [4] 张桢巍. 基于数据挖掘技术的高校智慧宿舍系统的研究与实现 [D]. 武汉: 华中师范大学, 2019.
- [5] 孙吉贵, 刘杰, 赵连宇. 聚类算法研究 [J]. 软件学报, 2008, 19(1): 48-61.
- [6] 张璨. 学生宿舍管理系统的设计与实现 [D]. 厦门: 厦门大学, 2014.
- [7] Polykovskiy, D. and Novikov, A., Bayesian Methods for Machine Learning. Coursera and National Research University Higher School of Economics. 2018 [引用日期 2023-12-10]
- [8] 王巧巧, 韦澍. 基于贪心算法的学生宿舍分配系统设计与实现 [J]. 计算机光盘软件与应用, 2013, 16(10): 147-148.