

基于K-MEANS聚类算法的智能手机APP使用行为研究

苏 焯¹ 吴友权² 李柯岐² 韦思诗² 朱治衡^{2*}

1.湖南科技大学 湖南湘潭 411200; 2.桂林理工大学南宁分校 广西崇左 532100

摘 要: 基于K-MEANS聚类算法的智能手机APP使用行为研究旨在深入了解智能手机用户的应用程序使用模式。该研究采集了大规模的APP使用数据,包括应用程序类型、使用频率、使用时段等信息。随后,研究利用K-MEANS聚类算法对这些数据进行分析 and 分类。研究结果揭示了不同用户群体之间存在明显的APP使用模式差异。通过K-MEANS聚类,将用户分为若干群体,每个群体都有其独特的APP使用行为。这些群体可以进一步用于个性化推荐系统的改进,以提高用户体验。

关键词: K-MEANS聚类算法; 智能手机; APP使用行为; 用户模式; 数据分析

Research on Smartphone APP Usage Behavior Based on K-MEANS Clustering Algorithm

Ye Su¹, Youquan Wu², Keqi Li², Sisi Wei², Zhiheng Zhu^{2*}

1.Hunan University of Science and Technology Hunan, Xiangtan, 411200, China

2.Guilin University of Technology Guangxi, Chongzuo, 532100, China

Abstract: The study of smartphone APP usage behavior based on K-MEANS clustering algorithm aims to gain insights into the app usage patterns of smartphone users. The study collected large-scale app usage data, including information such as app type, usage frequency, and usage period. Subsequently, the study utilized the K-MEANS clustering algorithm to analyze and classify these data. The results of the study reveal that there are obvious differences in APP usage patterns among different user groups. Through K-MEANS clustering, we classify users into several groups, each of which has its own unique APP usage behavior. These groups can be further used for the improvement of personalized recommendation system to enhance user experience.

Keywords: K-means clustering algorithm; Smartphone; App usage behavior; user pattern; Data analysis

1 引言

智能手机已经成为现代生活不可或缺的一部分,而移动应用程序在其中扮演着至关重要的角色。为了更好地理解和优化智能手机用户的应用程序使用行为,本研究基于K-MEANS聚类算法进行了深入探究。此研究旨在识别和分析不同用户群体之间的APP使用模式,以便更好地满足用户需求,提高用户体验。

收集了大量的APP使用数据,包括应用程序类型、使用频率、使用时段等信息。通过K-MEANS聚类算法,将用户分为若干群体,每个群体都表现出独特的应用程序使用习惯。这有助于了解用户的需求,以及他们如何与移动应用互动 [1]。

这项研究的结果不仅可以为个性化推荐系统提供有力支持,还可以为应用程序开发者提供宝贵的指导,帮助他们更好地满足用户的期望 [2]。总之,本研究旨在深入挖掘智能手机用户的APP使用行为,以推动移动应用领域的发展,提高用户体验,并为未来的研究和应用提供有力的基础。

2 K-MEANS 聚类

K-Means算法是一种常用的聚类算法,通过将数据集分成K个

不同的簇(cluster)来对数据进行聚类。K-Means算法的目标是最小化所有样本到其所属簇中心的平方距离之和,称为簇内平方和(WCSS, Within-Cluster Sum of Squares)。因此,该算法也被称为最小平方和聚类 [3]。K-Means算法的优点包括实现简单、计算效率高以及在大数据集上具有可扩展性。然而,它也有一些限制,如对初始聚类中心的敏感性、对离群点(outlier)的不稳定性以及对簇数K的事先设定等。在实际应用中,K-means算法常用于图像分割、市场分析、用户行为分析等领域 [4]。同时,还可以通过调整参数和结合其他算法进行改进,如K-Means++初始化方法、谱聚类等。

肘部法则则是用来帮助选择聚类分析中最优簇数的一种常见方法。它基于不同簇数所对应的总方差,帮助找到一个最优簇数。对于不同的k值(或簇数),计算其聚类结果的总内部方差 w_k ,将计算得到的每个簇的 w_k 累加,得到总方差 T 。计算各个k值对应的平均方差变化:

$$d_k = \frac{W_{k-1} - W_k}{W_1 + W_2 + \dots + W_n} \quad (1)$$

其中k表示当前聚类的簇数, w_{k-1} 和 w_k 分别表示k-1个和

k 个簇的总内部方差。 n 为聚类的最大簇数, $w_1 + w_2 + \dots + w_n$ 即为总方差 T 。

依据肘部法则, 得出上述平均方差变化随 k 变化的曲线图, 如下图所示。

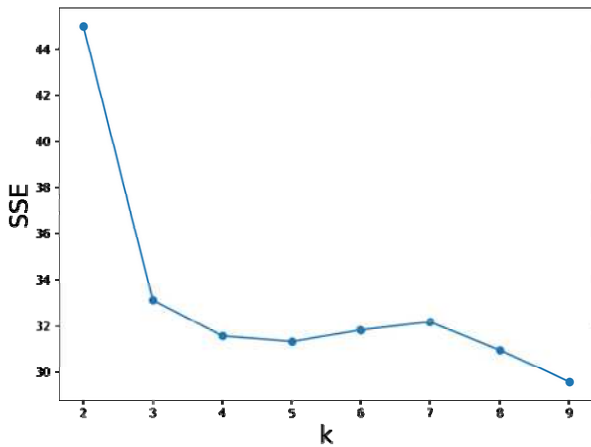


图1 聚类数量对比

图1用于选择较好的聚类数量, 横坐标是聚类个数, 纵坐标是 K 均值聚类的损失函数是所有样本到类别中心的距离平方和, 也就是误差平方和 (值越小说明聚类效果越好) [5]。在聚类数量达到4后, 下降的速度变缓, 形成一个“肘状”的拐点。该肘点即最优簇数。

使用聚类分析 $K - means$ 算法, 经过数据处理以4为参数, 将 n 个对象分为4个簇, 使得簇内对象具备相对较高的相似度, 并且簇与簇之间的相似度较低。

需要输入聚类个数 K , 以及包含的 n 个数据对象的数据库, 通过聚类分析 $K - means$ 算法得出满足方差最小标准的 K 的聚类。

首先, 从 n 个数据对象中任意选择9个用户的数据作为初始的聚类中心, 选择欧式距离作为数据相似度计算的方法, 它的定义为:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2)$$

通过欧式距离根据每个聚类对象中所有对象的均值计算每个对象与这些中心对象之间的距离, 并根据最小距离重新对相应对象进行划分, 并从中不断得出新的聚类中心, 重复上两步操作, 使得每个聚类不再发生变化, 即为最终的分类。将数据导入到 Python 当中, 最终得出了聚类点中心坐标的偏移情况。

3 模型求解

将各个用户的数据从1开始进行编号, 运用 Python 软件进行求解, 得出各个用户数据字段的差异性分析, 通过字段差异性分析可以得出各个用户使用 APP 时长的数据与各类别的匹配程度, 对各条数据进行分类, 使之具有一定的相关性。同时, 聚类点中心也会随

之发生变化, 此时, 再重复进行上一步的操作, 进行字段差异性分析, 并对各个数据进行重新的划分, 如此往复执行, 直到聚类点中心不再发生变化为止。

4 结论

首先, 通过 $K - MEANS$ 聚类, 成功将智能手机用户划分为不同的群体, 每个群体具有独特的 APP 使用模式。这些群体反映了用户之间的差异, 包括应用程序类型偏好、使用频率以及活跃时段等。这为个性化推荐和定制化用户体验提供了坚实的基础。

其次, 研究揭示了智能手机用户对应用程序的不同需求和期望。某些用户更注重娱乐应用, 而其他人可能更关心生产力工具或社交媒体应用。这种深入理解用户需求的洞察力对于应用程序开发者和市场策略制定者至关重要, 可帮助他们更好地满足用户的期望。

最后, 本研究为移动应用领域的未来发展提供了有益的指导。通过优化应用程序、改进用户界面和提供个性化服务, 可以提高用户满意度, 增加应用的用户粘性, 并促进市场竞争力的提升。

总之, 本研究强调了 $K - MEANS$ 聚类算法在深入理解智能手机 APP 使用行为方面的潜力, 为开发更智能、更具吸引力的移动应用提供了宝贵的见解。这对于满足用户需求、提高用户体验以及推动行业进步都具有积极意义。

参考文献:

- [1]Sun, H., Chen, Y., Lai, J., Wang, Y., & Liu, X. (2021). Identifying tourists and locals by K-means clustering method from mobile phone signaling data. *Journal of Transportation Engineering, Part A: Systems*, 147 (10), 04021070.
- [2]Shrivastava, G., & Kumar, P. (2021). Android application behavioural analysis for data leakage. *Expert Systems*, 38 (1), e12468.
- [3]Kang, S., & Kim, S. K. (2021). Behavior analysis method for indoor environment based on app usage mining. *The Journal of Supercomputing*, 77 (7), 7455-7475.
- [4]Chen, X., Li, W., & Jiang, Y. (2021). K-means clustering algorithms used in the evaluation of online learners' behaviour. *International Journal of Continuing Engineering Education and Life Long Learning*, 31 (3), 394-404.
- [5]Ribeiro, J., Saghezchi, F. B., Mantas, G., Rodriguez, J., & Abd-Alhameed, R. A. (2020). Hidroid: prototyping a behavioral host-based intrusion detection and prevention system for android. *IEEE Access*, 8, 23154-23168.