

本科毕业设计说明书

题 目：基于聚类的居住用户画像

及其簇特点分析

院 （部）： 计算机科学与技术学院

专 业： 软件工程

班 级： 软件143班

姓 名： 弭娜

学 号： 20141114072

指导教师： 宋玲

完成日期： 2018年5月25日

目 录

[摘 要 III](#_Toc484643463)

[ABSTRACT IV](#_Toc484643464)

[1前 言](#_Toc484643465)

[1.1研究意义 1](#_Toc484643466)

[1.2本文主要内容 1](#_Toc484643467)

[1.3本文主要工作 1](#_Toc484643468)

[1.4本章小结 1](#_Toc484643469)

[2相关研究](#_Toc484643470)

[2.1基于聚类的分析与应用相关研究 2](#_Toc484643471)

[2.2聚类方法的相关研究 2](#_Toc484643472)

[2.3 本章小结 2](#_Toc484643475)

[3基于机器学习的](#_Toc484643476)聚类分析

[3.1数据预处理 3](#_Toc484643477)

[3.2聚类算法 3](#_Toc484643478)

[3.3评价函数 3](#_Toc484643480)

[3.4本章小结 3](#_Toc484643481)

[4基于k-means的聚类分析系统的设计与实现](#_Toc484643487)

[4.1数据预处理 4](#_Toc484643488)

[4.1.1数据集的转换 4](#_Toc484643489)

[4.1.2 数据集标准化 4](#_Toc484643490)

[4.2聚类模型 4](#_Toc484643491)

[4.2.1 应用k-means建立聚类模型 4](#_Toc484643492)

[4.2.2参数调优 4](#_Toc484643493)

[4.2.3聚类结果的比较与分析 4](#_Toc484643494)

[4.3 本章小结 4](#_Toc484643495)

5[结 论 5](#_Toc484643496)

[谢 辞 6](#_Toc484643497)

[参考文献 7](#_Toc484643498)

# 摘 要

建筑环境作为能源消耗的主要载体，通过科学、合理、准确地预测建筑运行能耗是设定合理、明确的节能目标，制定建筑节能政策、法规，以及开展建筑节能工作的重要前提条件，也是实现建筑能耗需求同其他经济领域协调、可持续发展的重要保障。在这种宏观趋势下,加强对建筑能耗的预测和分析,具有重要的理论和实际意义。而在建筑能耗模拟过程中，居住用户的画像是预测建筑能耗的关键因素之一。

本文工作的主要概括如下：针对用户的每日的活动数据进行数据提取和处理，然后进行K-means聚类分析，根据聚类的结果进一步统计分析得出不同行为模式的用户人群对于日常活动轨迹的偏好和习惯特征。

关键词：用户活动；数据处理；聚类；K-means

**Analysis of residential user portrait and its cluster characteristics based on Clustering**

# ABSTRACT

As the main carrier of energy consumption, building energy consumption by scientific, reasonable and accurate prediction is the goal of setting reasonable and clear energy saving, making building energy saving policies, laws and regulations, as well as the important precondition of building energy saving work. It is also the coordination of building energy consumption demand with other economic fields and the important guarantee for sustainable development. Under this macro trend, it is of great theoretical and practical significance to strengthen the prediction and analysis of building energy consumption. In the process of building energy consumption simulation, the portrait of residential users is one of the key factors to predict building energy

The main work of this paper is summarized as follows: according to the daily activity data of users, data is extracted and processed, and then K-means clustering analysis is carried out. According to the results of clustering, the preference and customary characteristics of users with different behavior patterns for the path of daily activities are obtained.

**Key words**: user activity; data processing; clustering; K-means

# 1前 言

目前，发达国家建筑能耗占国家总能耗30％--40％，其中建筑运行能耗占30％以上。我国建筑能耗占社会商品总能耗的27％左右，暖通空调能耗约又占建筑能耗的50％～70％，建筑环境设备优化操作和管理可实现20％-30％的能源节约。有研究指出，建筑消耗了约40％的全球能源，其温室气体排放量约占全球温室气体排放的33％。随着全球气候变化、能源短缺、大气污染日益严重，特别是近几年，全国大范围出现的雾霾天气和PM2．5值严重超标，给我国能源生产和消费方式提出了严峻挑战。

1.1研究意义

1.2本文主要内容

1.3本文主要工作

本文的工作可概括为——采用聚类算法将用户的活动数据进行聚类分析（主要采用k-means算法），然后根据聚类结果得到最合适的聚类数量并针对确定下来的每一簇进行分析，得到每一簇用户活动的特点。

具体的工作流程为:

1. 数据预处理

1.4本章小结

2相关研究

2.1基于聚类的分析与应用相关研究

2.2聚类方法的相关研究

2.3 本章小结

# 3基于机器学习的聚类分析

3.1数据预处理

3.2聚类算法

3.3评价函数

3.4本章小结

# 

# 4基于K-means的聚类分析系统的设计与实现

4.1数据预处理

4.1.1数据集的合成

4.1.2数据集标准化

4.2聚类模型

4.2.1 应用k-means建立聚类模型

4.2.2参数调优

4.2.3聚类结果的比较与分析

4.3 本章小结

# 5结 论

# 谢 辞

# 

# 参考文献