Energy Consumption Pattern Discovery Using Clustering and Association Rule Mining

Alex Jurcich

Abstract

As smart energy metering expands, apartment energy data has become important for understanding and improving energy efficiency. My project aims to apply modern data techniques to identify patterns in daily electricity use across apartments located at UMASS. Using this dataset, which provides 3 years of electricity and weather records, the objective is to uncover behavioral clusters and learn relationships between environmental conditions and energy demands. I will combine different clustering algorithms with association rule mining to generate interpretable insights into apartments' energy usage. Performing this analysis can help develop energy usage patterns and understand how the seasons relate to energy usage.

Introduction

The increase in smart meter data has transformed the ability of data professionals to analyze the behavior of energy usage. Apartment-level energy consumption is influenced by numerous factors, such as temperature, season, occupancy, and appliance usage. These relationships are often nonlinear and difficult to pinpoint using summary statistics. Data Mining allows the use of powerful tools for discovering hidden patterns and structure within a dataset.

This project focuses on applying clustering and association rule mining to multiple years of UMASS apartment electricity data records. Unlike many other energy studies that examine household totals or forecast future energy consumption, this project will try to identify behavioral patterns, such as group apartments with similar daily energy consumption/patterns, and combinations of environmental factors that shift energy consumption. I will use clustering algorithms such as K-means, DBSCAN, and Gaussian Mixture Models to segment apartments. I will also use association rule mining with the intent to reveal co-occurring conditions such as temperature, humidity, and total energy consumption. Altogether, these techniques can provide actionable insights and explain why electricity usage varies. This can lead to supporting informed energy management and planning.

2 Methods

The dataset used is the **Smart* Data Set for Sustainability** from UMASS Trace Repository, containing daily electricity usage for 114 single-family apartments for the period 2014-2016, as well as weather data. This project will proceed in three main stages:

1. **Data Preprocessing:** For the preprocessing of this dataset, I will perform it in three ordered steps:
    a. Data Cleaning, where I will combine over 300 .csv files, parse and convert timestamps to daily summaries (mean, min, max, standard deviation, total sum), remove missing data and outliers, and apply IQR-based filtering to reduce the extremes in the dataset.
    b. Feature Engineering, where I will create categorical variables for the season (winter, spring, summer, fall) and weekend vs weekday.
    c. Dimensionality Reduction, where I will standardize numerical features and apply PCA to be able to visualize the clustering results.

2. **Cluster Analysis:** Apply unsupervised methods, including K-means, DBSCAN, and GMM, to identify natural groupings amongst apartments. Clustering will be evaluated using the appropriate metrics depending on the method, as well as visually.

3. **Association Rule Mining:** Perform the Apriori and FP-Growth algorithms to uncover co-occurring patterns between energy usage and environmental conditions.

References

[1] Smart*: An Open Data Set and Tools for Enabling Research in Sustainable Homes. Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, and Jeannie Albrecht. Proceedings of the 2012 Workshop on Data Mining Applications in Sustainability (SustKDD 2012), Beijing, China, August 2012. Pdf Smart* | UMass Trace Repository

[2] Tan, P., Steinbach, M., Karpatne, A., & Kumar, V. *Introduction to Data Mining (2nd ed.)*. Available via the U. Minnesota web site. College of Science and Engineering

[3] “MMDS” (Massive Data Mining Series) – http://www.mmds.org/

[4] Course slides from CSCI-B 565 (Data Mining)