CITIES
Centre for IT Intelligent Energy Systems
smart-cities-centre.org @CITIES_Centre

Innovation Fund Denmark

AALBORG UNIVERSITET

DTU Danmarks Tekniske Universitet

Clustering-Based Analysis

clustering-based analysis for residential district heating data

BY P. GIANNIOU[1], X. LIU[1], A. HELLER[1], P. S. NIELSEN[1], C. RODE[1]
[1] Danmarks Tekniske Universitet, Lyngby, Denmark

INTRODUCTION

Information and communications technologies (ICTs) are revolutionizing today's energy management systems. Their defining characteristic is to use smart meters to monitor energy consumption. The wide use of smart meters enables collection of a large amount of fine-granular time series, which can be used to improve the understanding of consumption behaviour and used for consumption optimization.

INVESTIGATION

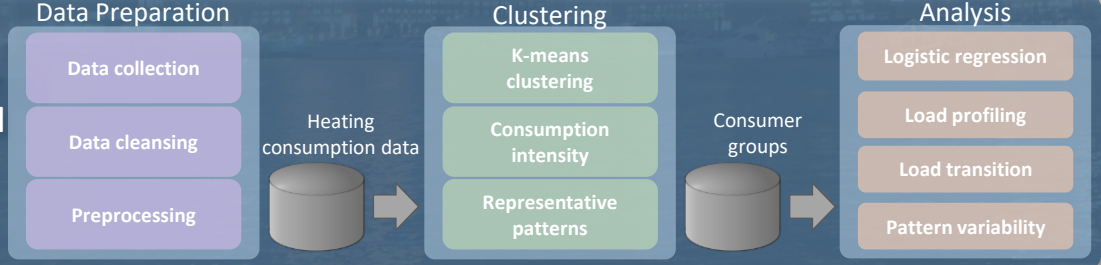
1. A clustering-based knowledge discovery is presented in a database method to analyze residential heat consumption data and evaluate information included in the national building databases. The proposed method uses the K-means algorithm to segment consumption groups based on consumption intensity and representative patterns and ranks the groups according to daily consumption.

2. This paper also examines the correlation between energy intensity and the characteristics of buildings and occupants, load profiles of households, consumption behaviour changes over time, and consumption variability.

METHOD

A three-stage approach is used to process and analyze district heating consumption data. The K-means algorithm is employed for the clustering to segment different consumer groups. The goal of the clustering is to identify and segment the consumers with similar load intensity and consumption patterns, and, secondly, to specifically look at the consumption patterns on normalised data. A correlation study is conducted to investigate the influencing factors (buildings' and occupants' features) for each of the consumption intensity groups based on the clustering results. Therefore, a variability study is conducted to investigate the consumption behaviour change over time for households.

Overview of the clustering-based analysis



RESULTS AND FINDINGS

The district heating consumption data are from 8293 single-family households in Aarhus, Denmark, with hourly time resolution. Single-family houses represent 44% of the total number of residential households in Denmark. Moreover, the utility, Aarhus Affaldvarme, provided the data for research purposes after anonymizing the data like removing personally identifiable information. The additional data include building information from the Danish National Building Register (BBR) such as the construction years, sites and areas.

The clusters were characterized by fairly constant load profiles with two weak peaks in the early morning and in the evening, respectively. In addition, the clusters were labelled with the alphabets, A–E , to represent heating consumption levels ranging from low to high.

The results indicated that building age, area and family size had a pronounced impact on the consumption, whereas the impact of age of occupants was less pronounced. It is therefore appropriate to use the first two factors to categorize Danish housing stock in building classification schemes with regards to energy consumption intensity, in particular single-family households can be used.

DISCUSSION

The paper studied the load profile characteristics for a single customer using the clustering approach, and based on the clustering results, illustrated the consumption transition probability over time and quantified the consumption variability using entropy methods. The results showed that the consumption patterns for each individual customer were serially correlated, and the higher consumption groups had lower variability in terms of the patterns.

For future work, clustering-based short-term energy demand forecasting will be investigated. In this study, it has been found that the majority of the customers have regular and predictable consumption behaviours.

The proposed techniques can easily be implemented on large data sets. The selected algorithm, K-means, has been successfully used on large data sets because of its simplicity and linear time complexity.

ABOUT THE AUTHORS:

- **Panagiota Gianniou** is PhD from Department of Civil Engineering at DTU and associate at AWA Holding
- **Carsten Rode** is a professor in Department of Civil Engineering at DTU. car@byg.dtu.dk
  Danmarks Tekniske Universitet, Building 118, 2800 Kgs., Lyngby, Denmark
- **Xiufeng Liu** is a senior researcher in Department of Management Engineering at DTU. xiuli@dtu
- **Per Sieverts Nielsen** is a senior researcher in Department of Management Engineering at DTU. pernn@dtu.dk
  Danmarks Tekniske Universitet, Building 426, 2800 Kgs., Lyngby, Denmark
- **Alfred Heller** was an associate professor in Department of Civil Engineering at DTU now works for NIRAS. ahr@niras.dk

Acknowledgments: Danish Innovation found. Grant DSF 1305-00027B