CMPE- 255 Data Mining

The Ocho

Clustering Individual Household Electric Power Consumption

Presented to

Professor Gheorghi Guzun

By

David Montes

Aaron Choi

Wil Carrasco

Contents

[Introduction 3](#_Toc73389415)

[Motivation 3](#_Toc73389416)

[Objective 3](#_Toc73389417)

[System Design & Implementation details 3](#_Toc73389418)

[Algorithms considered/selected 3](#_Toc73389419)

[Technologies and tools used 3](#_Toc73389420)

[System design/architecture/data flow 3](#_Toc73389421)

[Experiments / Proof of concept evaluation 3](#_Toc73389422)

[Dataset 4](#_Toc73389423)

[Preprocessing 4](#_Toc73389424)

[Methodology 4](#_Toc73389425)

[Graphs 4](#_Toc73389426)

[Analysis 4](#_Toc73389427)

[Discussion & Conclusions 4](#_Toc73389428)

[Decisions made 4](#_Toc73389429)

[Difficulties faced 4](#_Toc73389430)

[Things that worked 4](#_Toc73389431)

[Things that didn’t work well 4](#_Toc73389432)

[Conclusion 4](#_Toc73389433)

[Project Plan / Task Distribution 5](#_Toc73389434)

# Introduction

Clustering Individual Household Electric Power Consumption

## Motivation

Power usage and utilization is a metric that every household generates. Can patterns exist in time lapsed measurands and can that data be used to predict high demand or categorize usage? Patterns can be found in reoccurring daily tasks and making a large set to analyze these patterns can be beneficial to consumers and providers.

## Objective

Our group proposes to use the Individual household electric power consumption data set to look for power consumption trends over time. We plan on clustering the data using descriptive methods to discover patterns and trends.

# System Design & Implementation details

## Algorithms considered/selected

* K-Means
* Bisecting K-Means
* PCA
* L-2 Normalization
* SVD
* Feature selection / Extraction
* DBScan

## Technologies and tools used

* Python 3.x
* Jupyter Notebooks
* VSCode
* Numpy
* Scipy
* Sklearn
* Matplotlib

## System design/architecture/data flow

# Experiments / Proof of concept evaluation

## Dataset

https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption

The dataset itself contains 2,075,259 power consumption measurements from a

house in Sceaux, France between December 2006 and November 2010. The measurements

were taken every minute and consist of the following:

* Date in dd/mm/yyyy
* Time in hh:mm:ss
* Global Active Power: household global minute-averaged power (kilowatt)
* Global Reactive Power: household minute-averaged reactive power (kilowatt)
* Voltage: minute-averaged voltage(volt)
* Global Intensity: household global minute-averaged current intensity (ampere)
* Sub Metering 1: Energy sub-metering which corresponds to the kitchen, containing a
* dishwasher, an oven, and a microwave. (watt-hour of active energy)
* Sub Metering 2: Energy sub-metering which corresponds to the laundry room, containing
* a washing-machine, a tumble-drier, a refrigerator, and a light
* Sub Metering 3: Energy sub-metering which corresponds to an electric water heater and
* an air-conditioner

### Preprocessing

The first step was to read the raw data and remove unwanted delimitators and next line characters. We opted to remove any uncaptured data from the original dataset as this lack of information would not be useful in clustering. Since the time data was in string format, we also converted that information into a numeric ratio and reduced, merging them into a single dimension instead of two. The remaining data was normalized by the max value in each column checked the global power to check that the trends were the same. We applied multiple dimensionality reduction techniques such as PCA, SVD, and feature selection and determined that a combination of feature selection/extraction worked best.

## Methodology

## Graphs

## Analysis

# Discussion & Conclusions

## Decisions made

## Difficulties faced

## Things that worked

## Things that didn’t work well

## Conclusion

# Project Plan / Task Distribution

|  |  |
| --- | --- |
| **Task** | **Responsibility** |
| Dataset Selection | ALL |
| Data Exploration | ALL |
| Data Cleaning | David, Wil |
| Data Preprocessing | David, Wil |
| Time ration conversion | David |
| Dimensionality Reduction | Wil, Aaron |
| Graphs | All |
| Bisecting K-Means | Wil |
|  |  |
|  |  |
| Report | All |
| PowerPoint Presentation | TBD |