ENERGY IN TIME WORKSHOP

European Research Conference: Buildings
Europäische Forschungskonferenz: Gebäude

Data Mining for Improving Building Operation

Carlos Fernández-Basso

University of Granada

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608981.

World Sustainable Energy Days
1 - 3 March 2017, Wels/Austria

Outline

- Data and Data Mining
  - Definition and techniques
- Data Mining for improving building operation
  - Prediction
  - Analysis
- Steps to implement Data Mining analysis in buildings
- Conclusions

Data and Data Mining

Data, Information & Knowledge

- **Data** - The most elementary description of things

  “0.9” “1.12” “1.30” “1.40”

- **Information** - Organized Data that has meaning and value

  “Average price for petrol from 2012 to 2015 was 0.9€ 1.12€ 1.30€ 1.40€”

- **Knowledge** - The concept of understanding information based on the recognition of patterns to provide insight to information

  “Average Petrol price increases each year”

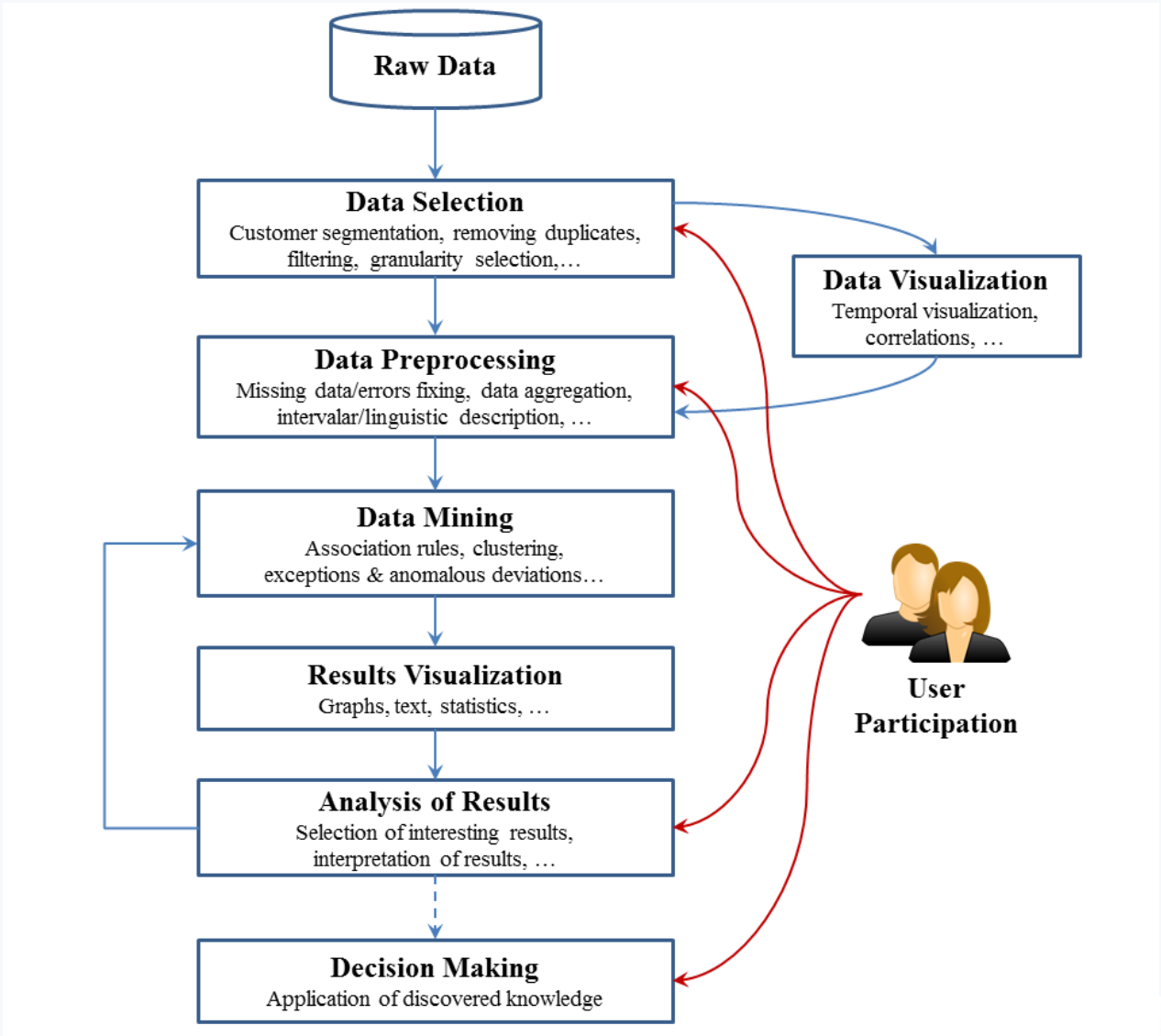
What is Data Mining?

- Data Mining is the task of discovering interesting **patterns** from large amounts of data
- Data Mining is the confluence of several technologies:
  - Databases
  - Statistics
  - Machine Learning
  - Information Science
  - Visualization
  - ….

Steps of Data Mining

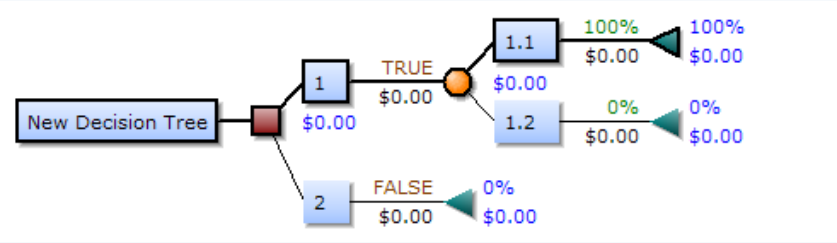
- Data Integration
- Data Selection
- Data Cleaning
- Data Transformation
- Data Mining
- Pattern Evaluation
- Knowledge Visualization

Methodology for EiT

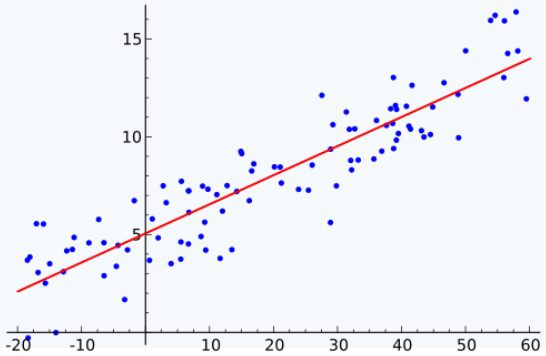


Techniques of Data Mining (I)

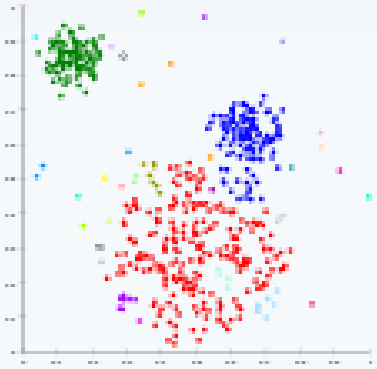
- **Classification**
  - Assigns an object to a class based on the properties of other objects or the comparison to a model
    - Fraud detection in bank transactions
    - Customers classification
    - Mortgage suitability



- **Regression**
  - Fits a equation to a dataset. Allows to predict a value based on another input value
    - Prediction of product stock levels
    - Inflation rate prediction



Techniques of Data Mining (II)



- **Clustering**
  - Gather similar objects together in a group (cluster) according to their similarity.
    - Image segmentation
    - Topic detection in text documents
- **Sequence discovery**
  - Techniques that identify statistically relevant patterns in data, whose values are distributed in order
    - Protein sequence discovery
- **Anomalies detection**
  - Identify items, events, or observations that deviate from expected patterns or from the usual behaviour of other data items
    - Bank fraud detection
- **Time Series analysis**
  - Performed on time-series data in order to obtain a model for the data that can be used to predict future values of the time series.
    - Weather forecast

Techniques of Data Mining (III)

- **Association rules**
  - Represent new information extracted from raw data that is expressed in the form of implication rules of the type $A \rightarrow B$ for decision-making.
  - Use measures of Support and Confidence
    - Market basket analysis
    - Medical diagnosis

Antecedent → Consequent

ID	Items
1	{Beer, diapers, milk}
2	{Beer, bread, eggs}
3	{Beer, diapers, milk , pizza, eggs}
4	{Beer, biscuit, milk, diapers, bread}

Diapers, milk → Beer

Association Rules

Definition

Association Rules discover hidden relations between data.

- Example:

  Sensor1 = value1 -> sensor2 = value2

  support 0.1, confidence 0.75

- Pair **<Sensor1, value1>** is called **item**
- It represents:
  - Value1 and value 2 occur together in 10% of the registered transactions along time
  - The strength of this relation is 75%
  - A transaction consists of all values for the sensors at a given time

- Possible meaning (should be given by the end-user)
  - Some values given by Sensor1 and Sensor2 are related

Why is Data Mining useful?

- Nowadays the volume of data generated by electronic devices, sensors, etc is **huge**
- Sensors in buildings provide lots of useful **information**

We have lots of DATA but we want KNOWLEDGE.

Data Mining and Big Data techniques allow for knowledge extraction from huge volumes of data

Data Mining Implementation in Buildings

How can Data Mining Improve Building Operation?

- Predicting the **energy demand** required for the efficient operation of a building
- Optimizing **building operation**
- Verifying the operational status and **failures of building equipment** and networks
- Analysing the **economic and commercial impact** of user energy consumption
- Detecting and preventing **energy fraud**.

How to Start?

Procedure

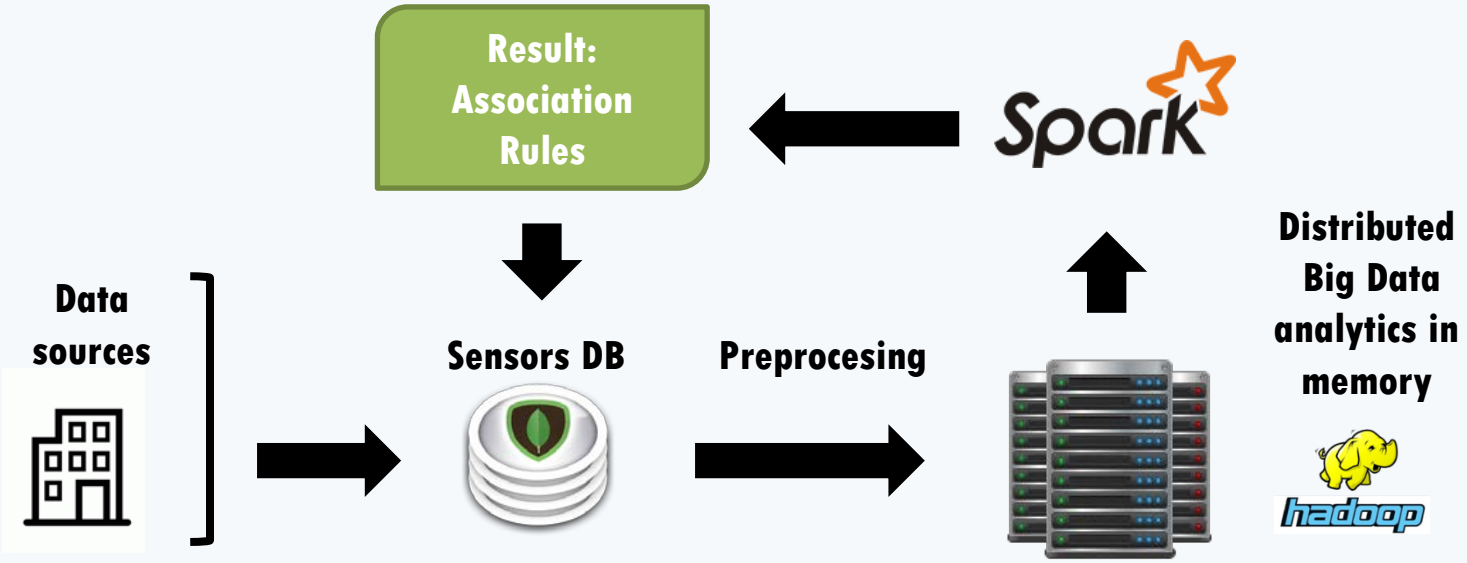
- Collect Data
- Preproccess and transform Data
- Select and apply Algorithms [Association Rules]
- Visualize Results
- Interpret Results

Critical Factors

- Storage (NoSQL)
- Processing Time (Big Data Analytics)

Data Analysis Workflow

From Building to Knowledge

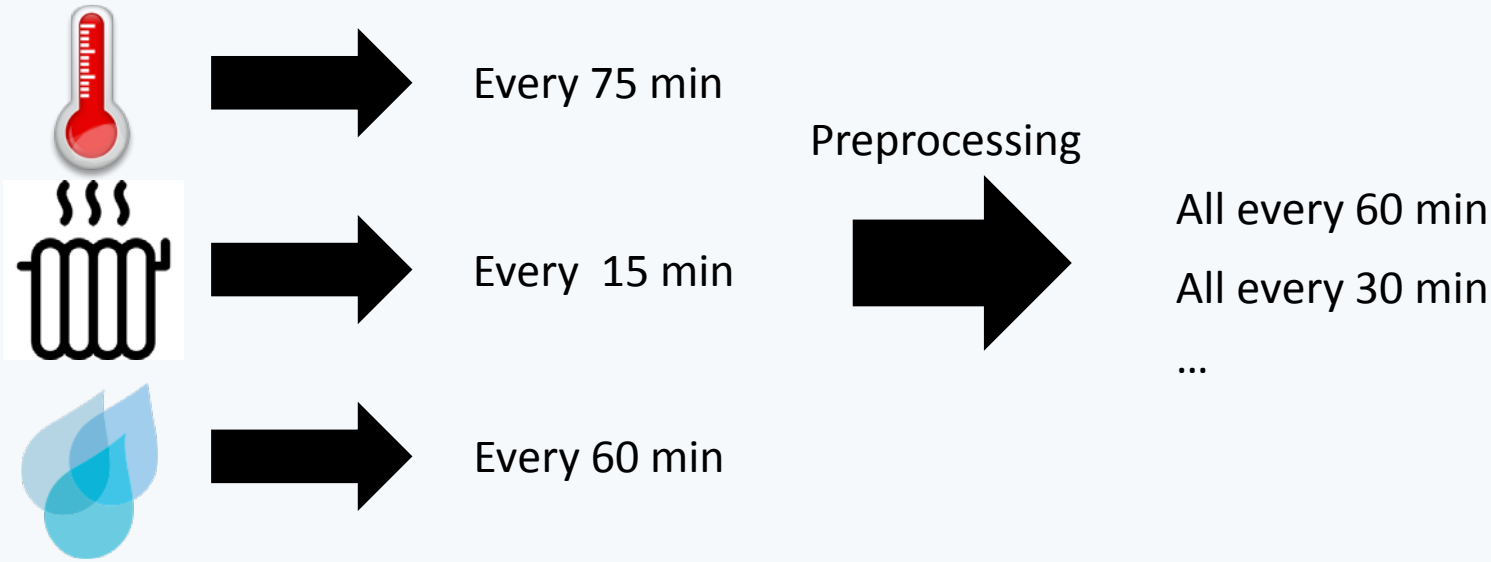


Data Collection

Data for many different sensors must be collected and stored

The sensors collect data at different frequencies

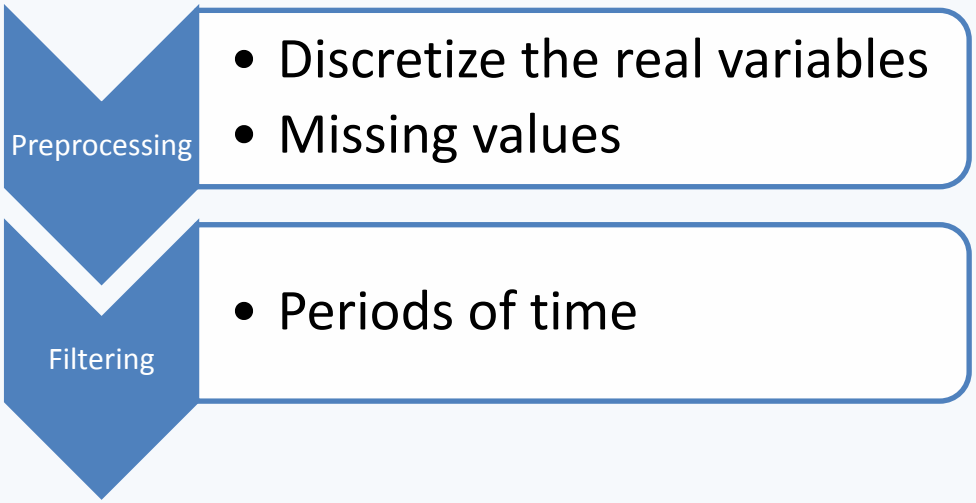
In order de properly work with the Data it must be processed and adapted



Preprocessing

Data must be adapted for processing

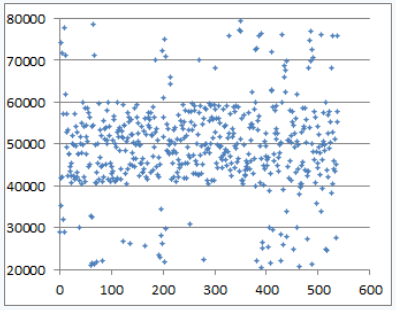
- Thousands of Variables
- Including:
  - Temperature
  - Humidity
  - Power supplied
  - Status cooling
  - Heating system
  - Energy consumptions



Preprocessing for Association Rules

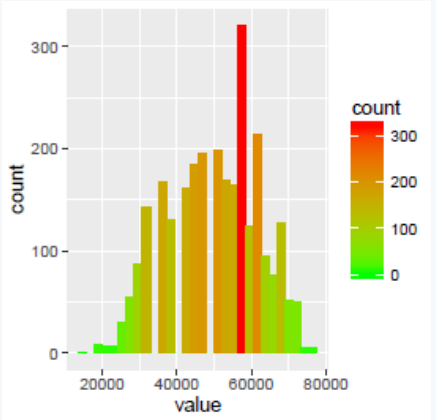
Association rules work better with interval values

All sampled data must be discretised into intervals



Floating point values

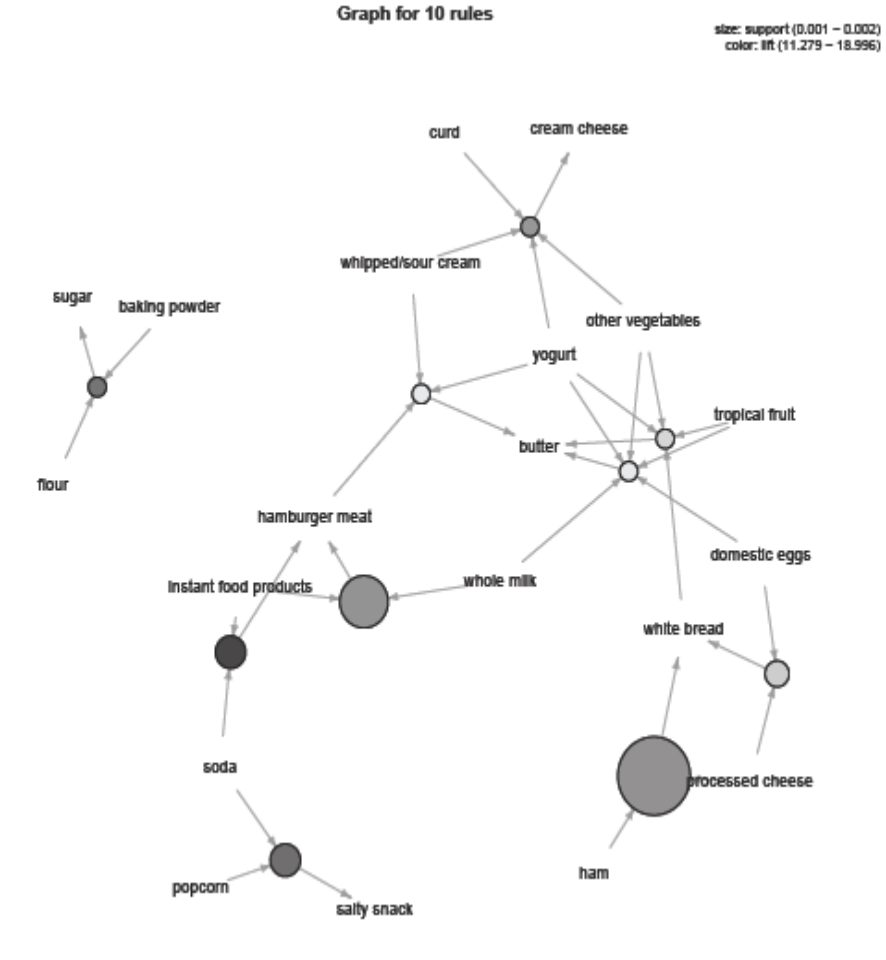
Discretise

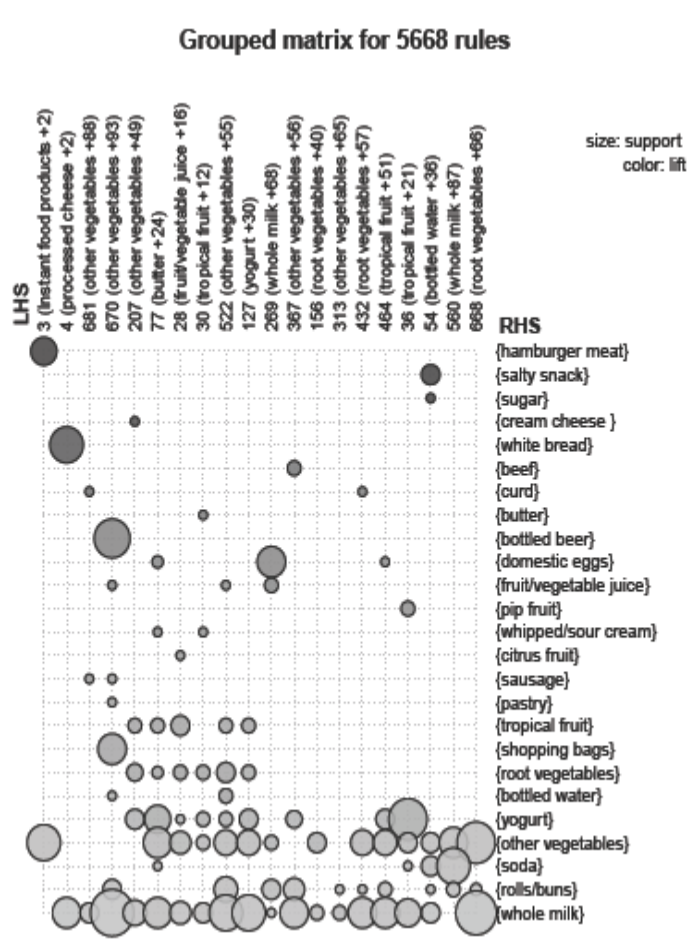


Intervals

Visualization of Association Rules

arulesViz Package (CRAN)





M. Hahsler and S. Chelluboina: Visualizing Association Rules: Introduction to the R-extension Package arulesViz. http://cran.r-project.org/web/packages/arulesViz/

Conclusions

- Data Mining techniques can improve building operation
- Necessary to know beforehand what we want to accomplish
- Data collection is fundamental
- Results must be visualized and interpreted
- NoSQL Databases and Big Data Analytics are key technologies when dealing with huge amounts of data

ENERGY IN TIME

Thanks you!

www.energyintime.eu

Carlos Fernandez-Basso

World Sustainable Energy Days
1 - 3 March 2017, Wels/Austria