Data Analysis Techniques for Monitoring District Heating Substations

Shahrooz Abghari
shahrooz.abghari@bth.se

Co-authors: Veselka Boeva, Christian Johansson, Jens Brage
Håkan Grahn, and Niklas Lavesson

September 10, 2019
District heating domain

- Faults in substations [1]:
  - Comfort problem and physical faults
  - Known faults but unsolved
  - Faults which require **new fault detection** methods

Hybrid method

- Higher Order Mining [2]: mining from *derived data rather than raw data*
- Combination of methods:
  - Sequential Pattern Mining -> to find *frequently occurring patterns* with respect to a user-specified threshold.
  - Clustering Analysis,
    - Weekly clustering -> to group extracted patterns based on their similarities
    - Consensus clustering, -> to group patterns that are coming from two weeks explaining the same phenomenon.
  - Minimum Spanning Tree -> to detect *outlying* patterns

Data & preprocessing

- Hourly numerical measurements, 2-year data
- Feature selection
  - *Outdoor temperature* and five features:
    - ‘sec_delta_t’, ‘prim_delta_t’, ‘consumed_energy’, volume_flow’, ‘station_effectivness’
    - \[ \text{station_effectivness} = \frac{\text{prim_delta_t}}{\text{primary supply temperature} - \text{secondary return temperature}} \]
- Data categorization
  - Transform the continues data into *four* categories
  - *Four* seasons = low, low_medium, medium_high, high

<table>
<thead>
<tr>
<th>sec_delta_t</th>
<th>prim_delta_t</th>
<th>consumed_energy</th>
<th>volume_flow</th>
<th>station_effectivness</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>high</td>
<td>medium_high</td>
<td>medium_high</td>
<td>high</td>
</tr>
</tbody>
</table>
Proposed method

Data segmentation → Frequent sequential patterns finding & Mapping → Cluster analysis
Given two clustering solutions of datasets $X$ and $X'$:

- $C = \{C_1, C_2, ..., C_n\}$
- $C' = \{C'_1, C'_2, ..., C'_m\}$

$$S_w(C, C') = \frac{\sum_{i=1}^{n}(\min_{j=1}^{m} w_i \cdot d(c_i, c'_j))}{2} + \frac{\sum_{j=1}^{m}(\min_{j=1}^{n} w'_j \cdot d(c_i, c'_j))}{2}$$

- $w_i = \frac{|c_i|}{X}$
- $d$, distance measure
Proposed method

Data segmentation

Frequent sequential patterns finding & Mapping

MST building

Cluster analysis

Cluster analysis

Week 1
Week 2
...
week n

Data segmentation

Frequent sequential patterns finding & Mapping

MST building

Cluster analysis

Cluster analysis
Minimum spanning tree (MST)

- Given an undirected, connected, and weighted graph $G=(V,E)$
  - a spanning tree of the graph $G$ is a tree that covers all the nodes of $G$
  - a tree with the minimum cost of traversing is called minimum spanning tree

- MST is applied on top the clustering solutions
  - $V =$ exemplars of the clusters
  - $E =$ dissimilarity between $V$

- Removing the longest edges
  - Smaller sub-trees are labeled as outlier
Proposed method (cont.)

Data segmentation

Frequent sequential patterns finding & Mapping

Deviating cluster(s)

MST building

Cluster analysis

Cluster analysis
Results:
Substation’s bi-weekly profile, 2017
Results:
Comparison of two profiles, 2017
Results:
Weekly clustering models and consensus model
Future work

- We aim to pursue further analysis and evaluation of the proposed approach
- In the long-term perspective,
  - Label weekly patterns with some *performance indicators*
Thank You for listening!

Shahrooz Abghari
shahrooz.Abghari@bth.se