Exploring Energy Performance Certificates through visualization

Tania Cerquitelli, Evelina Di Corso, Stefano Proto, Alfonso Capozzoli, Fabio Bellotti, Maria Giovanna Cassese, Elena Baralis, Marco Mellia, Silvia Casagrande and Martina Tamburini
Context

• Energy Performance Certificates (EPCs)
  • Became mandatory with European regulations

• Open Data
  • Big volumes of collected data
  • The Piemonte region makes EPCs publicly available
Goals

• Energy characterization and mapping of the city of Turin
  • Knowledge extraction and visualization at different granularity levels
  • Present the gathered information through highly-interpretable informative dashboards, accessible also to non-domain experts.

• Create value for different stakeholders:
  • Support and improve decisional processes
  • Energy planning and targeted incentive policies
  • Make extracted information publicly accessible
Energy Performance Certificates (EPCs)

• Energy analysis of the buildings:
  • Termo-physical characteristics
    • Walls and windows characteristics
    • Building information
  • Energy characteristics
  • Geo-localization information

• Mandatory since last regulation, information of real buildings
• Open data available on the Sistema Piemonte service system¹

¹ http://www.sistemapiemonte.it/cms/privati/ambiente-e-energia/servizi/856-sistema-informativo-per-le-prestazioni-energetiche-degli-edifici-sipee
INDICE
INformative DynamIC dashboard Engine

The INDICE architecture
Data integration

- Two different data sources, merged together
  - Different regulations
  - EPCs from 2009 to 2014
  - EPCs from 2015 to June 2018

2009-2014 data extraction

2015-2018 data extraction
Expert-driven univariate analysis

• Selected EPCs: related to permanent residences (buildings type: E1-1)

Identification of the most important variables

Identification of the validity ranges for each variable

EPCs characterized by admissible values ~ 40%

EPCs that require further investigation

- Average U-value of the vertical opaque envelope
- Average U-value of the windows
- Aspect Ratio
- Efficiency of the plant subsystems
- ...

Outlier detection based on

- Knowledge of domain experts
- gESD method (generalized Extreme Studentized Deviate)
- MAD (Median Absolute Deviation)
- Boxplot
Data cleaning: address resolution

• EPCs with invalid address format
  • Typing errors
  • Incorrectly-coded characters
  • 31.6% of the addresses have generic ZIP Code information
  • Wrong longitude and longitude coordinates

• Adopted solution
  • Addresses in the DB have been compared to those stored in the Turin road list (from the City Hall Geo-portal)
  • Levenshtein distance used to compute the similarity index between the addresses in the ECPs and the reference DB.
    • If the address has been resolved, the ZIP Code and the coordinates are saved in our DB eliminating inconsistencies
    • If the address has not been resolved, the ZIP Code and coordinates are obtained through the Google geocoding API

1 https://developers.google.com/maps/documentation/geocoding/intro
2 http://geoportale.comune.torino.it/web/
Outlier detection: multivariate analysis

• Density-based clustering algorithm: **DBScan**
  • Splits the database in parts characterized by different densities (dense and sparse)
  • Density is defined by two parameters (i.e., Eps, MinPoints), that are difficult to set
  • Self-tuning strategy based on k-dist plot

Correlation analysis

• Data-driven
  • **Feature selection and removal** (correlation-based approach)
    • Simplifying the model computation
    • Improving the model performance

S/V  Aspect Ratio
U_o  Average u-value of opaque envelope
U_w  Average u-value of the windows
ETAH Average global efficiency for spacing heating
S_t  Heat transfer surface
S_f  Floor Area
Year  Construction Year
EPH  Normalized primary heating energy consumption
Selected EPCs

- E1-1 buildings, used as **permanent residence**
- EPCs issued in the period: **2009 – 2018**
- Number of selected EPCs (each identifying a single dwelling): ~ **30,000**
Self-tuning cluster analysis

• Clustering algorithms enriched by self-tuning strategies (i.e., parameter autoconfiguration)
  • Partitional algorithm: K-Means
    • Each cluster is represented by a centroid
    • The desired number of clusters is identified by the user
  • Self-tuning strategy based on the Elbow plot: quality-measure trend (e.g., SSE) vs K
    • The gain from adding a centroid is negligible
    • The reduction of the quality measure is not interesting anymore

Knowledge characterization

• Each discovered cluster of EPCs is characterized by means of:
  • Centroids, represented through radar plots
  • Data distribution of each attribute, modeled through boxplot
  • Cluster label, assigned by analyzing the EPH distribution locally

<table>
<thead>
<tr>
<th>Cluster ID</th>
<th>APE #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 0</td>
<td>811</td>
</tr>
<tr>
<td>Cluster 1</td>
<td>4,321</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>1,117</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>3,988</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>2,080</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>2,723</td>
</tr>
<tr>
<td>Cluster 6</td>
<td>2,264</td>
</tr>
<tr>
<td>Cluster 7</td>
<td>1,723</td>
</tr>
<tr>
<td>Cluster 8</td>
<td>3,369</td>
</tr>
<tr>
<td>Cluster 9</td>
<td>3,418</td>
</tr>
<tr>
<td>Cluster 10</td>
<td>2,042</td>
</tr>
<tr>
<td>Cluster 11</td>
<td>2,119</td>
</tr>
</tbody>
</table>
Clusters of EPCs: High vs Low energy performances
Cluster characterization

Automatic extraction of knowledge from data, being directly exploitable by all stakeholders (including non-experts).

**Association rules** extraction
- Exhaustive analysis of all the possible correlations, above a given threshold, among values of the attributes characterizing the EPC certificates
- Requires discretization for numeric attributes (data- / domain-driven)
- Can be performed at different granularity / aggregation levels (hierarchy definition)
- Qualitative indexes to select only the most relevant correlations
- Transparent self-describing model, directly “readable” by humans

\[
X \rightarrow Y \quad \{\text{Global Mean Efficiency} = (0.85, 1.0]\}, \{\text{Average U-value of the vertical opaque envelope} = (0.15, 0.45]\}, \{\text{Average U-value of the windows} = (1.1, 1.85]\} \rightarrow \{\text{High Energy Performance}\}
Knowledge visualization

• Maps with different spatial granularity levels
  • City
  • District
  • Neighborhood
  • Building

• Different types of maps
Knowledge visualization

• Maps with different spatial granularity levels
• Different types of maps
  • Choropleth maps
    • An aggregation metric is required
      Majority model
      Statistical functions to be defined with the domain expert
  • Maps with cluster-markers (district levels)
  • Scatter maps with EPCs-markers
    • Different spatial granularities
    • Dynamic plots to model aggregated EPCs
Knowledge visualization

- Maps with different spatial granularity levels
- Different types of maps
  - Choropleth maps
    - An aggregation metric is required
      Majority model
      Statistical functions to be defined with the domain expert
  - Maps with cluster-markers (district levels)
  - Scatter maps with EPCs-markers
    - Different spatial granularities
    - Dynamic plots to model aggregated EPCs
Knowledge visualization

- Maps with different spatial granularity levels
- Different types of maps
  - Choropleth maps
    - An aggregation metric is required
      Majority model
      Statistical functions to be defined with the domain expert
  - Maps with cluster-markers (district levels)
  - Scatter maps with EPCs-markers
    - Different spatial granularities
    - Dynamic plots to model aggregated EPCs
Dashboard overview
Work-in-progress activities

• Exploitation of supervised learning algorithms
  • to enhance the data cleaning step
  • to include a larger number of EPCs in the analysis

• Generalization of the extracted knowledge
  • through machine learning and statistical method to provide a detailed overview at the city spatial granularity level
... questions?

Stefano Proto
Department of Control and Computer Engineering
Politecnico di Torino, Italy
stefano.proto@polito.it