VISUAL CONTINUOUS MAPS FOR ELECTRIC BILLS COMPARISON

Antonio Morán ∗,1 Juan J. Fuertes ∗ Manuel Domínguez ∗
Miguel A. Prada ∗ Serafín Alonso ∗ Pablo Barrientos ∗

∗ SUPPRESS research group, Universidad de León, Spain,
a.moran@unileon.es, jj.fuertes@unileon.es,
manuel.dominguez@unileon.es, ma.prada@unileon.es,
saloc@unileon.es, pablo.barrientos@unileon.es

Abstract: The University of León has several billing points, so all the data from the electric bills constitute a high-dimensional dataset which is quite complicated to visualize at a glance. The use of techniques for dimensionality reduction enables to obtain a two-dimensional representation of the original dataset which highlights main features in data and is easier to visualize. If these techniques are combined with interpolation methods, it can be generated continuous maps that allow comparison and interpretation of a whole range of possible electric datasets, not only the original one. These tools are used to create interactive maps that can be used by untrained people to exploit and analyze the information in electric bills, detect penalties due to a power demand excess or power factor decrease and make decisions with regard to electric contracts. Copyright CONTROLO2012

Keywords: Dimensionality reduction, Neural networks, Electric power, Data visualization

1. INTRODUCTION

In today’s society, it is necessary to carry out actions which promote savings and energy efficiency for the sake of a more sustainable development due to the significant increase of the electricity consumption during the last years. One of these actions must be the introduction of electric monitoring systems that measure, store and display the available information in electric data to define responses aimed at improving energy efficiency (Janssen, 2004). Currently, these systems are scarce and inefficient because they are only focused on the storage and visualization of a few electric variables. Usually, it is only possible to visualize the variables involved in the electric bills during a short-term.

The University of León has a system for monitoring the electric variables of all the buildings of the campus (Domínguez et al., 2011). In order to have all the consumption data available, a tool has been developed to get the electric bills from the electric company server and store these data with the rest of the electric variables. The large amount of buildings in the University makes difficult to visualize and compare data from the bills. In addition, since there are different terms of power consumption, the amount of data increases. Therefore, it is proposed to use techniques to reduce the dimensionality of data for a more intuitive display and also to create continuous maps of these data to correct the problem of discretization that such projections generate. This problem was considered in (Díaz et al., 2001) to generate continuous projection with SOM (Self-Organizing Maps) and in (García et al., 2011) the continuous maps were generated with techniques for dimensionality reduction such as the t-SNE or Isomap. The solution in both cases was based on the application of interpolation techniques to es-
2. ARCHITECTURE

The Campus of the University of León is composed of a group of buildings used for teaching, research or to provide complementary support services. The electric power consumption in these buildings, framed in the tertiary sector, corresponds basically to lighting systems, heating, ventilating and air-conditioning systems (HVAC), as well as industrial-type equipment associated with the research centers. In Table 1, all the buildings and billing points of the Campus can be observed.

The architecture adopted for the monitoring system of electric power in the buildings of the University of León is based on a three-tier structure (Eckerson, 1995). This structure follows the client-server pattern, where an intermediate layer is added to improve the modularity and the scalability. This layer manages the information provided by the Server layer.

The Server layer includes the electric meters and, conceptually, the server of the electric company. This server provides electronic bills for all the billing points of the University, which are parsed by the system. Physically this layer is composed of a private network which connects the electric meters and is isolated from the other networks by means of an internal router. In addition, this layer enables a connection, through a firewall, between the server of the electric company and a server in the middle layer for the data acquisition. The Middle layer is formed by a storage server, a data mining server, a web server and an acquisition server. This server records data both from the meters and the electronic bills and stores them orderly in the database. The storage server is composed basically of a relational database.

The Client layer contains the monitoring interface to be presented to the analysts. The system architecture allows any computer with an Internet access to be used for monitoring. There are several applications developed to visualize the data and the models of the electric data. The technologies used to develop the graphical interfaces are HTML, PHP, LabVIEW and Processing (Reas and Fry, 2007). Processing is the chosen language to develop the information visualization tool for the electric bills, because it is an open source programming language and environment that allows easy programming of visual interactive interfaces. Processing is based on Java language but its syntax is simplified for graphics programming.
It is proposed a method to generate continuous maps for visualization high-dimensional data. The first step is to generate a projection from the high-dimensional data of the input space into a 2-dimensional space which is easier to interpret for humans. Among the available techniques for dimensionality reduction, such as Isomap, Sammon, MDS, LLE, etc (Lee and Verleysen, 2007), the selected one in this case is the SNE (Stochastic Neighbor Embedding), since it gave the best visual results in the preliminary tests. The second step is to generate continuous maps to improve the visualization since the techniques for dimensionality reduction are discrete. Regression techniques are used to extrapolate data and generate continuous datasets. With regard to extrapolating techniques, GRNN (General Regression Neural Network), which can be considered a variant of the RBF (Radial Basis Function), is used. Fig. 2 shows the methodology used to generate continuous maps.

3.1 Stochastic Neighbor Embedding

Stochastic Neighbor Embedding (Hinton and Roweis, 2002) is a probabilistic method for projecting the data, or their dissimilarity. The SNE does not try to preserve the distance between data points, like the multidimensional scale methods, but the probability that the data points are neighbors. To achieve that, the conditional probabilities, \( p_{ji} \), are calculated. These probabilities indicates the possibility that \( x_i \) would pick \( x_j \) as its neighbor if neighbors were chosen according to their probability density which is estimated as a Gaussian distribution centered at the point \( x_i \). This probability is high for nearby data points and near zero when the points are separated. Mathematically, the conditional probability is given by:

\[
p_{ji} = \frac{\exp \left( -\frac{\| x_i - x_j \|^2}{2\sigma_i^2} \right)}{\sum_{k \neq i} \exp \left( -\frac{\| x_i - x_k \|^2}{2\sigma_i^2} \right)}
\]

(1)

where \( \sigma_i \) is the variance of the Gaussian function centered on the point \( x_i \). In the projected data, it is also used a similar probability denoted \( q_{ji} \) for all data points \( y_i \) and \( y_j \). It is also used a Gaussian neighborhoods but with a fixed variance. The similarity between the data in the low-dimensional space is given by:

\[
q_{ji} = \frac{\exp \left( -\| y_i - y_j \|^2 \right)}{\sum_{k \neq i} \exp \left( -\| y_i - y_k \|^2 \right)}
\]

(2)

If data have been correctly modeled, the conditional probabilities \( q_{ji} \) and \( p_{ji} \) will be equal. So, the SNE tries to find a low-dimensional projection that minimizes the mismatch between the conditional probabilities. To achieve this objective, the SNE minimizes the sum of Kullback-Leibler divergences over all data points:

\[
E_{SNE} = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}
\]

(3)

The minimization of the cost function is performed using a gradient descent method which may be interpreted as forces pulling or pushing between the points of the low-dimensional space, such that the resultant force exerted is proportional to the similarity between the points in the high-dimensional space. The gradient descent is initialized randomly and it is updated at each iteration according to the following equation:

\[
y^{(k+1)} = y^{(k)} - \eta^{(k)} \nabla E_{SNE}^{(k)}
\]

(4)

where \( \eta \) is the learning rate and \( \nabla E_{SNE}^{(k)} \) is the minimization cost function for step \( k \).

3.2 General Regression Neural Network

The main problem of the techniques for dimensionality reduction is that the projections are discrete. Due to this discretization, these techniques do not consider possible states that have not taken place. In other words, they only stores information of the known states and they do not consider the possibility of new states. For this reason, it is necessary a projection algorithm to interpolate value of each feature for new points in the visualization space.

To calculate these interpolated values, it is used the algorithm GRNN to match the extracted features and dimension reduced projections (Specht, 1991).
Fig. 3. Projection of the sum of the consumed active energy

<table>
<thead>
<tr>
<th>Peak period</th>
<th>Flat period</th>
<th>Off-peak period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak max. demand</td>
<td>Flat max. demand</td>
<td>Off-peak max. demand</td>
</tr>
<tr>
<td>Peak active energy</td>
<td>Flat active energy</td>
<td>Off-peak active energy</td>
</tr>
<tr>
<td>Peak reactive energy</td>
<td>Flat reactive energy</td>
<td>Off-peak reactive energy</td>
</tr>
</tbody>
</table>

This equation can be considered as a special case of the Nadaraya-Watson Regression Estimator (NWRE) (Zaknich, 1998)

\[
\hat{y}(x) = S_{x \rightarrow y}(x) = \frac{\sum_{i=1}^{N} \Phi(||x - x_i||) y_i}{\sum_{i=1}^{N} \Phi(||x - x_i||)} \tag{6}
\]

where

\[
\Phi(x) = \exp\left(-\frac{||x||^2}{2\sigma^2}\right). \tag{7}
\]

The width factor \(\sigma\) controls the smoothness of the mapping and can be adjusted on an empirical basis.

In the equation 6, \(\hat{y}(x)\) is an interpolated point on the visualization space, which is the low dimensional space, and \(x\) is a point of the input space whose projection is required.

Using this methodology, it is possible to obtain a reverse mapping. In other words, mapping from the visualization space to the input space \(S_{y \rightarrow x}\).

\[
\hat{x}(y) = S_{y \rightarrow x}(x) = \frac{\sum_{i=1}^{N} \Phi(||y - y_i||) x_i}{\sum_{i=1}^{N} \Phi(||y - y_i||)} \tag{8}
\]

This kind of projection enables the creation of a continuous map from a continuous grid which is generated manually. This grid contains all the coordinates of the visualization space including those coordinates which are not filled by the discrete projection. By means of the regression, a vector with data from the input space is generated for each coordinate. After obtaining these maps, it is possible to make projections that allow us to estimate the position of new features and to visualize the distribution of new data.

4. PROJECTION OF BILLING DATA.

A vector with the variables, listed in Table 2, is built. Variables are weighted in agreement with the penalties included in the bills. This way, some of them will be projected further than the normal ones in the output space. The weighting is performed as follows:

- Deviation from contracted power, i.e., the difference between the maximum demanded power \(P_{m}\) and the contracted power \(P_c\). The deviation is weighted according to the equation used
by the supplier to bill for the maximum demand power

\[
P_f = \begin{cases} 
0.85P_c & \text{if } P_m < 0.85P_c \\
\frac{P_m}{P_c} & \text{if } 0.85P_c \leq P_m \leq 1.05P_c \\
\frac{P_m + 2 (P_m - 1.05P_c)}{P_c} & \text{if } P_m > 1.05P_c 
\end{cases}
\]

(9)

- Active energy, which receives no weighting because the supplier does not penalize for it.
- Power factor. There are three possible cases
  - If \( \cos \varphi > 0.95 \), the variable is not weighted because there are no penalties and the case of a capacitive power factor is not considered whatsoever.
  - If \( 0.95 > \cos \varphi > 0.8 \), its value is multiplied by 2.
  - If \( 0.8 > \cos \varphi \), its value is multiplied by 5 proportionally to the high penalty involved.

The weighting factor is calculated according to the price penalty.
- Reactive energy, which is not weighted because this term is not subject to penalties (the penalty is applied to the power factor).

Once data are projected, the continuous map is trained using a GRNN network with \( \sigma = 2 \). The purpose is to visualize the boundaries of the areas that incur a penalty by means of contour lines.

This tool uses data from the 11 billing points shown in Table 1 during the years 2010 and 2011.

5. VISUALIZATION TOOL

The tool for visualization developed in Processing is shown in Fig. 3. This tool allows us to choose the billing point that we need to analyze so that the data of the two years and the months which are deviated can be seen more clearly. On the other hand, all the points can be projected simultaneously in order to perform comparisons and determine which are the ones with higher consumption or deviation. The tool also enables the selection of individual bills to analyze them in depth and find the periods which cause penalties. Furthermore, a chart can also be displayed to show the evolution of the daily maximum power and minimum \( \cos \varphi \). The chart reveals the days under a higher penalty and therefore leads to further analysis of the conditions that caused those behaviors.

The size of the point linked to each bill is proportional to one of the variables used in the projection. The main variables which can be visualized in the tool are the active energy, the reactive energy, the power factor and the maximum power.

**Sum of the consumed active energy** \( (P_{mp} + P_{ml} + P_{mv}) \). Fig. 3 shows the consumptions for all billing points with a contour line map that shows the sum of the consumed energy in blue for low values (left side of the map) and red for the high ones (right side of the map). It can be seen that the lowest consumptions are generally found in summer and the highest ones in winter. A comparison among points also reveals that the billing points that join more buildings (see Table 1) present higher consumptions.

**Sum of the consumed reactive energy** \( (Q_{mp} + Q_{ml} + Q_{mv}) \). Fig. 4 shows the consumption of reactive energy in each billing point. As well as in the previous figure, a contour-line continuous map is used to determine the areas of the map regarding consumption. Since some buildings have capacitor banks to compensate reactive energy, there is no coincidence with the active energy map. The billing point with the highest consumption of reactive energy in this case is P6 (Agricultural Facilities), but the dates with higher consumption do not necessarily coincide with summer time.

**Minimum power factor** \( \min(\cos \varphi_{mp}, \cos \varphi_{ml}) \). Since the power supplier computes penalties with respect to the lowest minimum factor, this is the only visualized. Contour lines for \( \cos \varphi \) values of 0.95 and 0.8 are used to display the penalty thresholds. A quick glance to Fig. 5 allows us to find how penalized each building billing point is.

In the points P2, P4, P9 (School facilities, Library and the Environmental Research Center), the power factor is rarely lower than 0.95 so there are no penalties. This is explained by the fact that they have capacitor banks, although the visualization shows that they are incorrectly sized in the library, because the power
Fig. 6. Projection of the maximum deviation factor decreases below the threshold some months. The rest of the buildings are penalized sometimes. For instance, it can be seen how the billing point P6, which was the one that consumed more reactive energy, is in the area between both thresholds. Corrective measures are necessary for the point P10 (Mining Engineering School), since its power factor is sometimes below 0.8, so it is subject to severe penalties. Although the installation of capacitor banks seems convenient, a detailed analysis should be made to estimate the saving that would be obtained.

Maximum deviation of power consumption in the three periods ($\max(P_{mp}, P_{ml}, P_{ms})$). Fig. 6 shows only three billing points and the their associated ideal power contours to facilitate visualizations. Since the power supplier penalizes the maximum powers that exceed the contracted ones in a 5% or do not reach to the 85%, the contour lines determine a band where the consumption must lie to suffer no penalties. Since the contracted maximum power is negotiated with the supplier, the visualization is a useful tool to decide for which billing points must be changed. For instance, it is not necessary to modify the contracted power in the billing point P1, because most of the points lie in the appropriate band except in summer, when consumption drops and therefore it is to make any improvement. However, it can be seen that the maximum power consumption of the billing point P5 never reaches the threshold, so the contracted power must be lowered to meet the real demand and reduce the bills.

The tool also allows displaying the deviation contour lines of an optimal contracted power in terms of penalties related to the deviation from the contracted power. This contour lines are computed by minimizing the power term function 9 with the available bills.

6. CONCLUSIONS

In this paper it is presented a method to generate continuous maps which enable the dimensionality reduction of a high-dimensional dataset alleviating the problem of discretization associated with these techniques. To generate these maps, it is used a technique for dimensionality reduction like SNE which tries to maintain the topology of the data. After the projection, it is used a regression technique, like GRNN, to interpolate the data and make the projection continuous.

This technique is validated through an application developed in Processing. This application provides a visual representation of all the data from the electric bills from all the billing points in the University of León during the years 2010 and 2011. The tool allows a fast visualization of all the data and let users infer the billing points which are subjected to penalties by the electric company. Also, the tool allows comparison of the measurement points based on the consumption charged on each bill.

7. REFERENCES


