Identifying Customer Profiles in Power Load Time Series Using Spectral Clustering

Carlos Alzate, Marcelo Espinoza, Bart De Moor, and Johan A. K. Suykens

Department of Electrical Engineering, ESAT-SCD-SISTA
Katholieke Universiteit Leuven
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
{carlos.alzate,johan.suykens}@esat.kuleuven.be

Abstract. An application of multiway spectral clustering with out-of-sample extensions towards clustering time series is presented. The data correspond to power load time series acquired from substations in the Belgian grid for a period of 5 years. Spectral clustering methods are a class of unsupervised learning algorithms where the solutions can be obtained from the eigenvectors of a Laplacian matrix derived from the data. Nonlinearity can easily be added to the analysis by the use of nonlinear similarity functions that can be regarded as Mercer kernels. In this paper, a weighted kernel PCA formulation to spectral clustering is used to find interpretable customer profiles underlying the power consumption load time series. The main advantage of the interpretation as kernel PCA is the extension of the clustering model to out-of-sample points. The clustering model can be trained, validated and tested in a learning framework working directly with the data and without the use of pre-modeling steps. The experimental results with real-life data demonstrate the applicability of the multiway spectral clustering method compared to an existing method pre-modeling the data based on periodic autoregressions (PAR).

1 Introduction

Spectral clustering provides a powerful unsupervised learning tool to find similar patterns underlying the data. These methods are typically expressed as relaxations of NP-hard graph partitioning problems. Although several spectral clustering methods have been proposed in the literature [1,2,3], they all share the use of information contained in the eigenspectrum of Laplacian matrices derived from the data. Spectral clustering can be also interpreted as a kernel method when the chosen similarity gives rise to pairwise similarity matrices with positive eigenvalues. Kernel-based modeling is a powerful and elegant way of introducing nonlinearity into the analysis. However, in practice, the models can become too general for the application at hand with a high risk of overfitting. A careful and systematic parameter selection is thus needed in order to obtain relevant and informative groupings among the data. A recent approach linking spectral clustering with a form of weighted kernel PCA has been proposed in [4]. This

method fits into the least squares support vector machine (LS-SVM) [5] framework for kernel-based learning. One of the main advantages of this approach is the possibility to extend the clustering model to out-of-sample points without the need of approximation or ad-hoc schemes.

This paper contains an application of the multiway spectral clustering with out-of-sample extensions introduced in [4] towards clustering time series. Time series data pose a challenge to standard clustering methods due mainly to the high-dimensionality. The particular dataset corresponds to readings of the power load measure at 245 different substations in the Belgian grid for a period of 5 years. The objective is to identify types of costumers underlying the data. This task is particularly important for short, mid and long-term planning in the electricity sector. An approach to cluster this type of data was proposed in [6]. This approach uses a pre-modeling scheme based on periodic autoregressions (PAR) in order to reduce the dimensionality together with representing the time series independent of seasonal and calendar variations. In a later stage, k-means was used to find the clusters. Pre-modeling can be useful for interpretability and dimensionality reduction but there might be a loss of information while trying to pre-model the complete set of load time series with the same template. The proposed approach aims at obtaining interpretable customer profiles of power load consumption by clustering the data directly in a learning framework with model selection and without the use of pre-modeling steps. This paper is organized as follows. Section 2 contains a short description of the dataset used in the sequel. Section 3 summarizes classical spectral clustering. In Section 4, the multiway formulation to spectral clustering introduced in [4] is described. Section 5 reviews a model selection method. Section 6 contains the clustering results of an existing methodology and the proposed approach and in Section 7, we give conclusions.

2 Power Load Time Series

A time series dataset containing readings of the power load measured at 245 different high voltage - low voltage substations in the Belgian grid is used in this paper. The power load was measured every hour for a period of 5 years. The data characterize different profiles of load consumption such as business, residential and industrial. Each time series was normalized to have zero mean and unitary standard deviation. Figure 1 shows one particular time series of the dataset. The yearly cycles are visible together with the daily cycles. From the daily cycle it is also possible to visualize morning, noon and evening peaks.

3 Classical Spectral Clustering

3.1 Spectral Graph Theory

Given a set of N data points $\{x_i\}_{i=1}^N, x_i \in \mathbb{R}^d$ and some similarity measure $s_{ij} \geq 0$ between each pair of points x_i and x_j, an intuitive form of representing

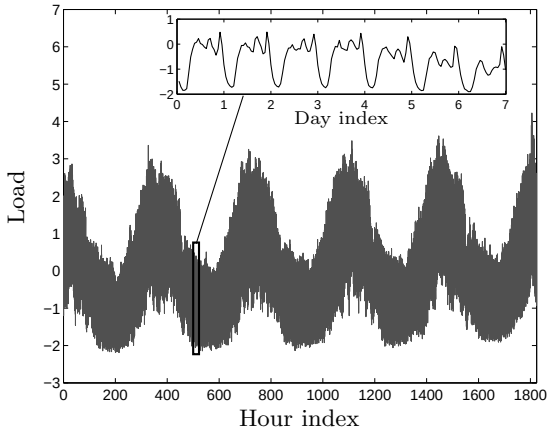

Fig. 1. Example of time series from the power load dataset. Yearly and daily cycles are visible. It is also possible to visualize morning, noon and evening peaks.

the data is using a graph $G = (\mathcal{V}, \mathcal{E})$. The vertices $\mathcal{V}$ represent the data points and the edge $e_{ij} \in \mathcal{E}$ between vertices v_i, v_j has a weight determined by s_{ij}. The affinity matrix of the graph is the matrix S with ij-entry $S_{ij} = s_{ij}$. The degree of vertex v_i is defined as $\deg_i = \sum_{j=1}^{N} s_{ij}$, the degree matrix D is the matrix $D = \mathrm{diag}([\deg_1, \ldots, \deg_N])$ on the diagonal.

A basic problem in graph theory is the graph bipartitioning problem, that is to separate the graph G into two disjoint sets $\mathcal{A}, \mathcal{B}$ based on a cut criterion. The resulting sets should be disjoint: $\mathcal{A} \cap \mathcal{B} = \emptyset$ and $\mathcal{A} \cup \mathcal{B} = \mathcal{V}$. This problem has been extensively studied [7] and several cut criteria have been proposed [8,1]. The normalized cut *NCut* [1] is a common bipartitioning criterion and is defined as $\mathrm{NCut}(\mathcal{A}, \mathcal{B}) = \mathrm{cut}(\mathcal{A}, \mathcal{B})/\mathrm{Vol}(\mathcal{A}) + \mathrm{cut}(\mathcal{A}, \mathcal{B})/\mathrm{Vol}(\mathcal{B})$, where $\mathrm{Vol}(\mathcal{A}) \equiv \sum_{i \in \mathcal{A}} d_i$ denotes the volume of the set $\mathcal{A}$.

3.2 Multiway Cut Criteria

A more general problem in graph theory is the so called multiway cut. In this case, the aim is to partition the graph into $k > 2$ disjoint parts $\mathcal{A}_1, \ldots, \mathcal{A}_k$, with $\mathcal{A}_1 \cap \ldots \cap \mathcal{A}_k = \emptyset$. The multiway NCut problem as introduced in [9] is $\min_G k - \mathrm{trace}(G^T \hat{L} G)$ such that $G^T G = I_k$, where $\hat{L} = D^{-1/2} L D^{-1/2}$ is the normalized Laplacian and $G \in \{0, c_L\}^{N \times k}$ is a cluster indicator matrix with one non-zero entry per row (indicating cluster membership) and c_L is a normalization constant. Minimizing the multiway Ncut is NP-hard. Relaxing G and allowing to take real values leads to a special form of Fan's theorem [10] where the relaxed optimal solution is given by:

$$\tilde{G}^{\star} = BR_L,$$

where $B \in \mathbb{R}^{N \times k}$ is any orthonormal basis of the k-th principal subspace of $\tilde{L}$ and R_L is an arbitrary rotation matrix. Roughly speaking, the relaxed solution of the multiway NCut is contained in the eigenvectors corresponding to the largest k eigenvalues of the normalized Laplacian up to arbitrary rotations.

Several methods to convert the partitioning information contained in the eigenvectors to clusters have been proposed [1,11]. A popular method introduced in [1] consists of *reclustering* and consists of computing the top k eigenvectors of the multiway *NCut* problem and then applying k-means onto them. Reclustering works well only if the clusters in the space represented by the eigenvectors are spherical and well separated.

3.3 Markov Random Walks

A Markov random walks view of spectral clustering was discussed in [12]. A random walk on a graph consists of random jumps from vertex to vertex. The ij-th entry of the stochastic transition matrix $P = D^{-1}S$ represents the probability of moving from node i to node j in one step. The corresponding eigenvalue problem becomes $Pr = \xi r$. Maximizing the random walks model is equivalent to minimizing the binary *NCut* and can be interpreted as finding a partition of the graph in such a way that the random walk remains most of the time in the same cluster with few jumps to other clusters.

In the ideal case of a graph with k disconnected components, the transition matrix P will have k eigenvalues equal to 1. The eigenvectors corresponding to these unitary eigenvalues are the indicator functions to the respective disconnected components and the clustering is trivial because each cluster is represented as a single point in $\mathbb{R}^k$. This also corresponds to the piecewise constant property of eigenvectors discussed in [12].

4 Spectral Clustering via Weighted Kernel PCA

In this Section, we summarize the multiway formulation for spectral clustering introduced in [4]. This algorithm considers weighted kernel PCA with D^{-1} as the weighting matrix to perform spectral clustering. The formulation fits into a constrained optimization framework typical of least squares support vector machines (LS-SVM). The choice of D^{-1} as the weighting matrix is motivated by the random walks model of spectral clustering. Moreover, [4] can be interpreted as a weighted centered version of the random walks algorithm with bias terms and out-of-sample extensions.

4.1 Multiway Formulation

Given training data $\{x_i\}_{i=1}^N, x_i \in \mathbb{R}^d$, the primal problem is formulated as follows:

$$\min_{w^{(l)},e^{(l)},b_l} J_P = \frac{1}{2N}\sum_{l=1}^{k-1}\gamma_l e^{(l)^T} D^{-1} e^{(l)} - \frac{1}{2}\sum_{l=1}^{k-1} w^{(l)^T} w^{(l)} \quad (1)$$
$$\text{such that } e^{(l)} = \Phi w^{(l)} + b_l 1_N, l = 1, \ldots, k-1$$

where k is the number of desired clusters, $\gamma_l \in \mathbb{R}^+$ are regularization constants, $e^{(l)} = [e_1^{(l)}; \ldots; e_N^{(l)}]$ is the l-th score variables vector, b_l are bias terms, Φ is the $N \times n_h$ feature matrix $\Phi = [\varphi(x_1)^T; \ldots; \varphi(x_N)^T]$, $\varphi(\cdot) : \mathbb{R}^d \to \mathbb{R}^{n_h}$ is the mapping to a high dimensional feature space $\mathcal{F}$ of dimension n_h (which can be infinite dimensional) and $l = 1, \ldots, k-1$. Thus, each training data point $x_i \in \mathbb{R}^d$ is represented as point in the score variables space $e_i \in \mathbb{R}^{k-1}$, $e_i = [e_i^{(1)}, \ldots, e_i^{(k-1)}]$. After binarization, the score variables become cluster indicators $q_i^{(l)} = \text{sign}(e_i^{(l)}), i, \ldots, N, l = 1, \ldots, k-1$. The Lagrangian of the constrained optimization problem (1) is expressed as $\mathcal{L}(w^{(l)}, e^{(l)}, b_l; \alpha^{(l)}) = J_P - \sum_{l=1}^{n_e} \alpha^{(l)^T}\left(e^{(l)} - \Phi^{(l)} w^{(l)} - b_l 1_N\right)$ where $\alpha^{(l)} \in \mathbb{R}^N$ are Lagrange multiplier vectors. Using the KKT optimality conditions $\partial\mathcal{L}/\partial e^{(l)} = 0, \partial\mathcal{L}/\partial w^{(l)} = 0, \partial\mathcal{L}/\partial\alpha^{(l)} = 0, l = 1, \ldots, k-1$, eliminating $w^{(l)}, e^{(l)}, b_l$ leads to $(\gamma_l/N) D^{-1} M_D \Phi\Phi^T \alpha^{(l)} = \alpha^{(l)}$ where

$$M_D = I_N - \frac{1}{1_N^T D^{-1} 1_N} 1_N 1_N^T D^{-1}. \quad (2)$$

Applying the kernel trick $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$, and defining $\lambda_l = N/\gamma_l$ leads to the following eigenvalue problem:

$$D^{-1} M_D \Omega \alpha^{(l)} = \lambda_l \alpha^{(l)}, j = 1, \ldots, k-1 \quad (3)$$

where Ω is the kernel matrix with ij-th entry $\Omega_{ij} = K(x_i, x_j)$. Note that the matrix $D^{-1} M_D \Omega$ is not symmetric but it is the product of two symmetric positive semidefinite matrices therefore its eigenvalues are real [13]. The bias terms become $b_l = -\frac{1}{1_N^T D^{-1} 1_N} 1_N^T D^{-1} \Omega \alpha^{(l)}, l = 1, \ldots, k-1$, and the projections of the training points onto the eigenvectors can be expressed in terms of the dual variables: $e_i^{(l)} = w^{(l)^T} \varphi(x_i) + b_l = \sum_{j=1}^N \alpha_j^{(l)} K(x_i, x_j) + b_l$.

4.2 Out-of-Sample Extension and Encoding / Decoding

In classical spectral clustering, the cluster indicators obtained from the eigenvectors are defined only for training data. Extensions to compute cluster indicators for unseen points are typically done in an ad-hoc manner using the Nyström approximation [14]. However, due to the primal-dual nature of the weighted kernel PCA framework, it is possible to extend the clustering model to out-of-sample points in an straightforward way using the projections onto the eigenvectors.

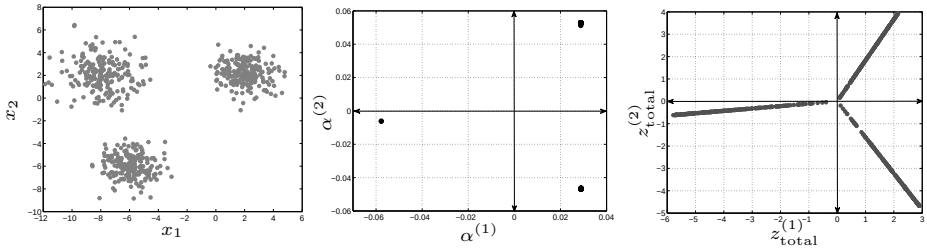

Fig. 2. Toy dataset illustrating piecewise constant eigenvectors and projections. **Left:** Training data consisting of 3 Gaussian clouds and $N = 600$ points. **Center:** Piecewise constant eigenvectors $\alpha^{(1)}, \alpha^{(2)}$ obtained by solving (3) with an RBF kernel, $\sigma = 0.5$. **Right:** Corresponding projections of the full dataset $N_{\text{total}} = 1,200$. Note that the clusters are represented in the eigenvectors and the corresponding projections in distinct quadrants.

Given a set of N_{test} test points $\{x_t^{\text{test}}\}_{t=1}^{N_{\text{test}}}$, the score variables for these points become: $z_t^{(l)} = w^{(l)^T}\varphi(x_t^{\text{test}}) + b_l = \sum_{j=1}^{N} \alpha_j^{(l)} K(x_t^{\text{test}}, x_j) + b_l$ and the cluster indicators are given by $\breve{q}_t^{(l)} = \text{sign}(z_t^{(l)}), l = 1, \ldots, k-1, t = 1, \ldots, N_{\text{test}}$.

The final clustering is obtained through encoding the $k-1$ binary indicators. Each training point $x_i \in \mathbb{R}^d$ has an associated binary encoding $q_i \in \mathbb{R}^{k-1}$. From the clustering indicators q_i a codebook is formed with the k binary encodings with most occurrences [4]. Figure 2 illustrates a simple example with 3 Gaussian clouds in a $2D$ space. Note that, the eigenvectors are piecewise constant thus representing each cluster as a single point in a different quadrant. The out-of-sample extension induced by the projections displays a line structure which can be used for model selection as will be discussed in the sequel. The decoding step consists of comparing the cluster indicators for out-of-sample data with each codeword in the codebook and selecting the encoding that minimizes the Hamming distance.

5 Model Selection

Model selection is a critical issue in unsupervised learning methods such as spectral clustering. Typically, the number of clusters k and the kernel parameters are chosen in a heuristic way. However, by means of the out-of-sample extensions, it is possible to perform model selection in an intuitive and straightforward way. A model selection criterion called the Balanced Line Fit (BLF) was introduced in [4]. The clusters are represented as lines in the score variables space for well-chosen parameters. Thus, it becomes important to characterize the fitness of a particular cluster to a line. The BLF as defined in [4] is a weighted sum of a collinearity measure called the linefit and a measure of the balance of the obtained clusters: $\text{BLF}(k) = \eta\text{linefit}(k) + (1-\eta)\text{balance}(k)$, where η controls the importance given to the collinearity measure with respect

to the balance, $0 \leq \eta \leq 1$, $\text{linefit}(k) = \frac{1}{k}\sum_{p=1}^{k} \frac{k-1}{k-2}\left(\frac{\zeta_1^{(p)}}{\sum_l \zeta_l^{(p)}} - \frac{1}{k-1}\right)$ where $\zeta_1^{(p)} \geq \ldots \geq \zeta_{k-1}^{(p)}$ are the ordered eigenvalues of the sample covariance matrix $C_{\tilde{Z}}^{(p)} = (1/|\mathcal{A}_p|)\tilde{Z}^{(p)^T}\tilde{Z}^{(p)}, p = 1, \ldots, k$, $\tilde{Z}^{(p)} \in \mathbb{R}^{|\mathcal{A}_p| \times (k-1)}$ is the matrix representing the zero mean score variables for out-of-sample data assigned to the p-th cluster. The term $\zeta_1^{(p)}/\sum_l \zeta_l^{(p)}$ indicates how much of the total variance is contained on the top eigenvector of $C_{\tilde{Z}}^{(p)}$. In this way, if the validation score variables for the p-th cluster are collinear, then $\zeta_1^{(p)}/\sum_l \zeta_l^{(p)}$ equals 1. The additional terms in the linefit normalize the criterion between 0 and 1. The balance measure is defined as $\text{balance}(k) = \min\{|\mathcal{A}_1|, \ldots, |\mathcal{A}_k|\}/\max\{|\mathcal{A}_1|, \ldots, |\mathcal{A}_k|\}$. The balance index equals 1 when the clusters have the same number of elements and tends to 0 in extremely unbalanced cases. Note that, the linefit is defined only for $k > 2$. For $k = 2$, a slight modification of the BLF was discussed in [4].

6 Experimental Results

6.1 Existing Methodology

An approach to cluster the load time series was proposed in [6]. This method computes a Typical Daily Profile (TDP) from the identified parameters in a Periodic Autoregression (PAR) system estimated for each times series, later performing k-means on the TDPs. PAR models allow the autoregression parameters to vary according to cyclic patterns (such as hours, days, years, seasons). These models have been used for forecasting in economical modeling and electricity price modeling. The method proposed in [6] aims at representing the time series as an hourly PAR model of order 48 with additional exogenous variables accounting for monthly and weekly cyclic variations together with temperature effects. The PAR parameters can be obtained by solving an Ordinary Least Squares (OLS) system. A TDP was defined in [6] as a convergence vector of the PAR(48) model after extracting all exogenous information. The main advantage of a TDP is that it represents a daily load profile independent of the seasonal and calendar variations. In summary, [6] creates a TDP for each time series which implies a dimensionality reduction from $43,824$ to 24. PCA is then applied over the TDP to capture 99% of the information explained by the variance. This corresponds to a further reduction to 9 dimensions. The last step consists of applying k-means over the 9 principal components of the typical daily profile. Figure 3 shows the clustering results. Several types of customers can be identified from the results, particularly clusters 1 and 2 which corresponds to *residential* profiles, clusters 6 and 7 which display a *commercial* profile, clusters 4, 5, 8 show a profile typical of street lighting and certain industrial activities taking place during the night.

6.2 Proposed Clustering Scheme

The proposed scheme consists of clustering directing the time series in a learning framework with model selection without any premodeling steps. The training

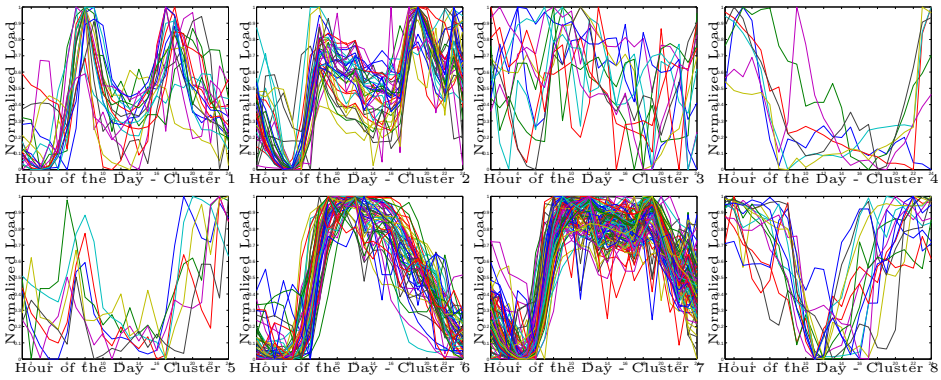

Fig. 3. Typical daily profile (TDP) clustering results using the method proposed in [6] which corresponds to preprocessing the TDP with PCA to reduce the dimensionality to 9 and then applying k-means.

scenario consists of 123 time series for training and 122 times series for validation. Each time series is a $43,824$-dimensional vector with zero mean. The training / validation sets were randomized 10 times and the model selection results report average values. We used the BLF for tuning the number of clusters and the RBF kernel parameter σ. The η parameter was set to 0.5 hence, giving equal weight to the linefit and to the balance. The tuning range for the RBF kernel parameter was varied from 500 to $2,000$ in 100 steps. The number of clusters k ranged from 2 to 12. The choice for the 500 as the minimum value of the range is motivated by the fact that with lower values the RBF kernel matrix was tending to the identity matrix which has trivial eigenspace with no discriminatory information about the clusters. On the other hand, values greater than $2,000$ led to a almost constant-valued kernel matrix independent of the data. The maximum of the BLF gives the optimal number of clusters and the optimal value for σ. Figure 4 shows the model selection results. The BLF detected 7 clusters and the optimal σ value was 200. For visualization purposes, each time series is transformed into a mean daily pattern by averaging the corresponding $1,824$ 24-hour non-overlapping windows. This mean daily pattern is similar to the TDP but differs in the sense that we did not model any periodicity or exogenous variables. Moreover, the mean daily pattern is computed *after* the clustering has been performed and serves only for visualization and clustering interpretation purposes. Figure 5 shows the mean daily patterns for the 245 time series using the multiway spectral clustering with parameters $k = 7, \sigma = 200$ obtained through model selection with the BLF. The obtained clusters also display typical customer profile information. Namely, cluster 1 suggests *residential* behavior with morning and evening peaks. Cluster 5 characterizes low consumption during the day increasing during the night. Clusters 2, 6 and 7 suggest *commercial* or *business* profiles while cluster 3 and 4 display load behavior with no particular associated profile.

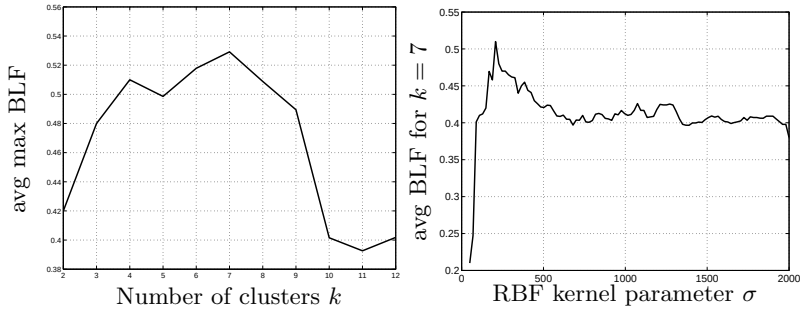

Fig. 4. Model selection using the BLF. **Left:** Average maximal BLF to determine the number of clusters k. **Right:** Average BLF for $k = 7$ to determine the optimal value of the RBF kernel parameter σ. Tuned parameters are $k = 7$ and $\sigma = 200$. The plots show average BLF values after 10 randomizations of the training and validation sets.

7 Conclusions

We have presented an application of the multiway spectral clustering method with out-of-sample extensions towards grouping types of customers in power load time series. The proposed scheme can be applied directly to the raw data without the need of pre-modeling steps. The parameters of the clustering model are obtained in a learning framework using training and validation scenarios. The experimental results show the applicability of the proposed methodology using a real-life high-dimensional dataset compared to an existing approach using periodic autoregressions.

Acknowledgements: This work was supported by grants and projects for the Research Council K.U.Leuven (GOA-Mefisto 666, GOA-Ambiorics, several PhD / Postdocs & fellow grants), the Flemish Government FWO: PhD / Postdocs grants, projects G.0240.99, G.0211.05, G.0407.02, G.0197.02, G.0080.01, G.0141.03, G.0491.03, G.0120.03, G.0452.04, G.0499.04, G.0226.06, G.0302.07, ICCoS, ANMMM; AWI;IWT:PhD grants, GBOU (McKnow) Soft4s, the Belgian Federal Government (Belgian Federal Science Policy Office: IUAP V-22; PODO-II (CP/01/40), the EU(FP5-Quprodis, ERNSI, Eureka 2063-Impact;Eureka 2419-FLiTE) and Contracts Research/Agreements (ISMC / IPCOS, Data4s, TML,Elia, LMS, IPCOS, Mastercard). Bart De Moor and Johan Suykens are professors at the K.U.Leuven, Belgium. The scientific responsibility is assumed by its authors.

References

1. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Machine Intell. **22**(8) (2000) 888–905
2. Meila, M., Shi, J.: Learning segmentation by random walks. Advances in Neural Information Processing Systems 13, MIT Press (2001)

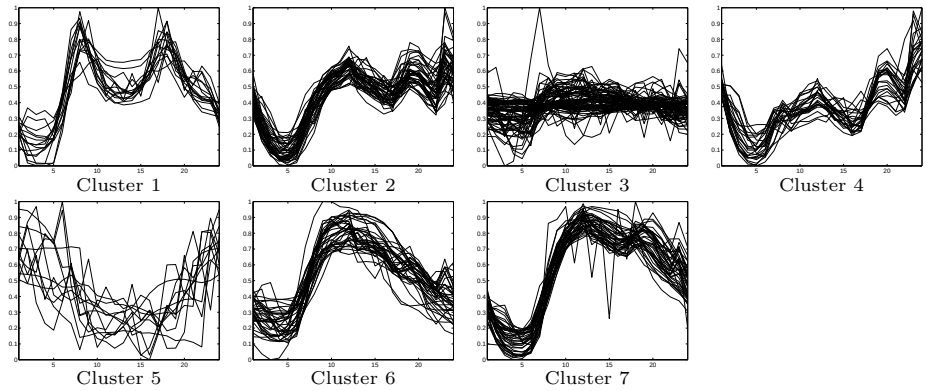

Fig. 5. Clustering results using the proposed methodology. The results show typical customer profiles while clustering the data directly without pre-processing. The x-axis corresponds to the hour of the day while the y-axis is the normalized load.

3. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: Analysis and an algorithm. Advances in Neural Information Processing Systems 14, MIT Press (2002) 849–856
4. Alzate, C., Suykens, J.A.K.: Multiway spectral clustering with out-of-sample extensions through weighted kernel PCA. IEEE Transactions on Pattern Analysis and Machine Intelligence. In Press (2009)
5. Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J.: Least Squares Support Vector Machines. World Scientific, Singapore (2002)
6. Espinoza M. and Joye C. and Belmans R. and De Moor B.: Short-term load forecasting, profile identification and customer segmentation: A methodology based on periodic time series. IEEE Trans. on Power System **20**(3) (2005) 1622–1630
7. Chung, F.R.K.: Spectral Graph Theory. American Mathematical Society (1997)
8. Fiedler, M.: A property of eigenvectors of nonnegative symmetric matrices and its applications to graph theory. Czech. Math. J. **25**(100) (1975) 619–633
9. Gu, M., Zha, H., Ding, C., He, X., Simon, H.: Spectral relaxation models and structure analysis for k-way graph clustering and bi-clustering. Tech. report, Penn. State Univ, Computer Science and Engineering. (2001)
10. Fan, K.: On a theorem of Weyl concerning eigenvalues of linear transformations. Proc. Natl. Acad. Sci. USA **35**(11) (1949) 652–655
11. Bach, F.R., Jordan, M.I.: Learning spectral clustering. Advances in Neural Information Processing Systems 16, MIT Press (2004)
12. Meila, M., Shi, J.: A random walks view of spectral segmentation. In: Artificial Intelligence and Statistics AISTATS. (2001)
13. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press (1990)
14. Fowlkes, C., Belongie, S., Chung, F., Malik, J.: Spectral grouping using the Nyström method. IEEE Trans. on Pattern Analysis and Machine Intelligence **26**(2) (2004) 214–225