Electricity consumption data clustering for load profiling using generalized self-organizing neural networks with evolving splitting-merging structures

Marian B. Gorzałczany, Jakub Piekoszewski, and Filip Rudziński
Department of Electrical and Computer Engineering
Kielce University of Technology
Al. 1000-lecia P.P. 7, 25-314 Kielce, Poland
Email: {m.b.gorzalczany, j.piekoszewski, f.rudzinski}@tu.kielce.pl.

***Abstract*—The main goal of the paper is the application of our clustering technique based on proposed by us generalized self-organizing neural networks with evolving tree-like splitting-merging structures to the clustering of electricity consumption data collected as a part of a smart metering pilot study conducted by Irish Commission for Energy Regulation (CER). First, the Irish CER data are briefly characterized. Then, the operation of our clustering technique is outlined and illustrated using a benchmark data set. In turn, the application of our approach to the Irish CER data clustering is presented, evaluated, and discussed as well as a comparative analysis with several alternative approaches is performed.**

I. Introduction

Rapid development of smart meter installations in various countries in recent years [1] has led to large amounts of electricity customer consumption data. The availability of such data can substantially improve various aspects of electricity production and distribution. One of them is load profiling which exhibits different behaviors and characteristics of customers. A load profile is a graph of the variation in the electrical load (the amount of electricity used by customers) versus time over some time period. Effective and efficient mining of the variability of electricity consumption shape - essential in load profiling - can be performed by means of clustering algorithms, in particular those rooted in the Computational Intelligence field. They aim at partitioning of the initial electricity consumption data into a set of clusters defined by assigning consumers with the most similar behavior (i.e., the most similar load variability shapes) into the same cluster and consumers with dissimilar behavior into different classes (cf. [2]).

The main goal of this paper is the application of our original data clustering technique based on generalized self-organizing neural networks (S-ONNs) with evolving tree-like splitting-merging structures (see also [3]–[6]) to the clustering of the data collected as a part of an electricity smart metering pilot study conducted by the Irish Commission for Energy Regulation (CER) [7]. First, a brief characteristics of the aforementioned data set is presented. Then, the operation of our clustering technique is outlined and illustrated using a benchmark data set. In turn, the application of our approach to the Irish CER data clustering is presented. Finally, a comparative analysis with several alternative approaches applied to the same data set is carried out.

II. Irish CER electricity consumption data set

The Irish CER initiated the CER Smart Metering Project, supplying 6445 electricity customers with smart meters. The study took place from July 14, 2009 (00:00h) to December 31, 2010 (23:59h), i.e., 536 days. Electricity consumption (in kW) supposed to be measured half-hourly, so the data should consists of $536 \cdot 48 = 25728$ recordings per series (a customer profile). 6445 customers were subcategorized as (i) *Residential* - consisting of 4225 customers, (ii) *Small-to-Medium Enterprises* (*SME*) - consisting of 485 customers, and (iii) *Others* - consisting of 1735 customers. However, it was observed on inspection that in the *Others*-category, the electricity consumption data for 1170 customers are not available for the entire duration of the recordings [8] and thus, the *Others*-category is not considered in experiments reported in the literature. Moreover, some number of customer profiles contain missing values (sometimes, significant amounts of them). For this reason, we decided to remove from the original data set, all customer profiles containing missing values. Finally, we obtained a data set of 4066 customer profiles consisting of 3639 *Residential* profiles and 427 *SME* profiles. This highly imbalanced data set (which poses additional challenge for any clustering method) will be used in our experiments reported later in the paper.

III. An outline of our clustering technique based on generalized S-ONNs with evolving tree-like splitting-merging structures

In this section, we outline main features, main objectives, general concept, implementation, and benchmark-based illustration of our clustering approach (more details can be found in [3]–[6]; see also [9]–[13] for earlier versions of our approach).

Main features: Our approach works in a fully unsupervised way, i.e., (i) it does not need to predefine the number of clusters and (ii) it uses unlabeled data. It is worth emphasizing that the knowledge on the assignments of the data samples to clusters is by no means used by our approach. However, after the completion of the clustering process, we can use that knowledge for the verification of the obtained results.

Main objectives: (i) an automatic determination of the number of clusters in a given data set and (ii) an automatic generation of multi-point prototypes for particular clusters.

General concept: Original S-ONNs (also referred to as self-organizing maps (SOMs)) [14] are used, in general, to visually display topological structures of high dimensional data in lower (usually 2-dimensional) space rather than for clustering, i.e., partitioning of these data into groups [15]. The proposed generalized S-ONNs, however, are also equipped with 3 additional mechanisms (allowing for data segmentation) such as: (i) automatic adjustment of the number of neurons in the network (removing low-active neurons and adding new neurons in the areas of the existing high-active neurons), (ii) automatic disconnection of the tree-like structure into subnetworks, and (iii) automatic reconnection of some of the subnetworks (preserving the no-loop spanning-tree properties). Such a generalized S-ONN is able to detect data clusters of various shapes and densities by locating a single disconnected subnetwork in the area of the data space occupied by a given cluster. Hence, the number of automatically generated subnetworks is equal to the number of clusters. Moreover, a set of neurons in a given subnetwork is a multi-point prototype of the corresponding clusters. Such a prototype can be directly used in clustering/classification tasks using the well-known nearest multi-prototype algorithms [16].

Implementation: We start with a conventional S-ONN with 1-dimensional neighborhood (S-ONN with 1DN), i.e., the neuron chain. Assume that the network has n inputs $x_1, x_2, \ldots, x_n$ and consists of m neurons; their outputs are $y_1, y_2, \ldots, y_m$, where $y_j = \sum_{i=1}^{n} w_{ji} x_i$, $j = 1, 2, \ldots, m$ and w_{ji} are weights connecting the i-th input of the network with the output of the j-th neuron. Using vector notation ($\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$, $\boldsymbol{w}_j = (w_{j1}, w_{j2}, \ldots, w_{jn})^T$), $y_j = \boldsymbol{w}_j^T \boldsymbol{x}$. The learning data consists of L input vectors $\boldsymbol{x}_l$ ($l = 1, 2, \ldots, L$). In the first stage of any Winner-Takes-Most (WTM) learning algorithm that can be used in the learning process of the considered network, the neuron $j_{\boldsymbol{x}}$, which wins in competition of neurons when the learning vector $\boldsymbol{x}_l$ is presented to the network must be determined. Assuming that the normalization of learning vectors is performed, the winning neuron $j_{\boldsymbol{x}}$ is selected in the following way:

$$d(\boldsymbol{x}_l, \boldsymbol{w}_{j_x}) = \min_{j=1,2,\ldots,m} d(\boldsymbol{x}_l, \boldsymbol{w}_j), \quad (1)$$

where $d(\boldsymbol{x}_l, \boldsymbol{w}_j)$ is the Euclidean distance measure between $\boldsymbol{x}_l$ and $\boldsymbol{w}_j$. The WTM learning rule is formulated as follows:

$$\boldsymbol{w}_j(k+1) = \boldsymbol{w}_j(k) + \eta_j(k) N(j, j_{\boldsymbol{x}}, k)[\boldsymbol{x}(k) - \boldsymbol{w}_j(k)], \quad (2)$$

where k is the iteration number, $\eta_j(k)$ is the learning coefficient, and $N(j, j_{\boldsymbol{x}}, k)$ is the neighborhood function of the $j_{\boldsymbol{x}}$-th winning neuron. Most often the Gaussian-type neighborhood functions are used, i.e.:

$$N(j, j_{\boldsymbol{x}}, k) = e^{-\frac{d_{tpl}^2(j, j_{\boldsymbol{x}})}{2\lambda^2(k)}}, \quad (3)$$

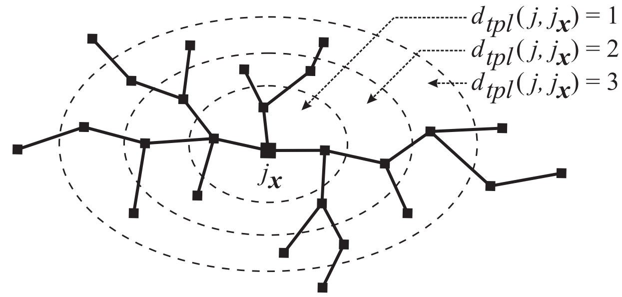

Fig. 1. Illustration of neighborhood of the $j_{\boldsymbol{x}}$ neuron in the tree-like neural structure [3]–[6]

where $\lambda(k)$ is the neighborhood radius and $d_{tpl}(j, j_{\boldsymbol{x}})$ - the topological distance between the $j_{\boldsymbol{x}}$-th and j-th neurons. In case of the conventional S-ONN with 1DN, $d_{tpl}(j, j_{\boldsymbol{x}}) = |j - j_{\boldsymbol{x}}|$. Once the earlier mentioned mechanisms of splitting and merging of the neural network structure are activated, the conventional S-ONN with 1DN evolves towards a tree-like structure. As a result of that, the concept of the neuron's neighborhood is also modified as shown in Fig. 1.

The mechanisms (i), (ii), and (iii) mentioned in the *General-concept*-part of this section are implemented by conditional activation - after each learning epoch - of 4 successive operations:

1. The removal of single, low-active neurons: neuron No.$j_{\boldsymbol{x}}$ is removed from the network (preserving the network continuity), if its activity - measured by the number of its wins win_{j_r} - is below an assumed level win_{min}, i.e., $win_{j_r} < win_{min}$. win_{min} is an experimentally selected parameter.

2. The disconnection of the network (subnetwork) into 2 subnetworks: the disconnection of two neighboring neurons j_1 and j_2 takes place if the following condition is fulfilled: $d(\boldsymbol{w}_{j_1}, \boldsymbol{w}_{j_2}) > d_{coef} d_{avr}$ where $d_{avr} = \frac{1}{P} \sum_{p=1}^{P} d_p$ is the average distance between two neighboring neurons for all pairs p, $p = 1, 2, \ldots, P$, of such neurons. d_{coef} (a distance coefficient) is an experimentally selected parameter.

3. The insertion of additional neurons into the neighborhood of high-active neurons in order to lower their activities: a new neuron, labeled as (new), is inserted between two neighboring and high-active neurons j_1 and j_2 (i.e., their numbers of wins win_{j_1} and win_{j_2} are above an assumed level win_{max}: $win_{j_1}, win_{j_2} > win_{max}$). win_{max} is an experimentally selected parameter.

4. The reconnection of two selected subnetworks: two subnetworks S_1 and S_2 are reconnected by introducing topological connection between neurons j_1 and j_2 ($j_1 \in S_1$, $j_2 \in S_2$) after fulfilling condition $d(\boldsymbol{w}_{j_1}, \boldsymbol{w}_{j_2}) < d_{coef} \frac{d_{avr_{S_1}} + d_{avr_{S_2}}}{2}$. $d(\boldsymbol{w}_{j_1}, \boldsymbol{w}_{j_2})$ and d_{coef} are the same as in operation 2. $d_{avr_{S_1}}$ and $d_{avr_{S_2}}$ are calculated for subnetworks S_1 and S_2, respectively, in the same way as d_{avr} is calculated in operation 2 for the considered network.

Benchmark-based illustration: The clustering of the benchmark set, referred to as *TwoDiamonds* data set, from the so-called Fundamental Clustering Problem Suite (FCPS) [17] is

TABLE I
CLUSTERING RESULTS FOR IRISH CER DATA SET

Cluster label	Number of samples	Number of decisions for subnetwork labeled:		Number of correct decisions	Number of wrong decisions	Percentage of correct decisions
		Res.[*1]	*SME*			
Res.[*1]	3639	3589	50	3589	50	98.63%
SME	427	159	268	268	159	62.76%
ALL	**4066**	**3748**	**318**	**3857**	**209**	**94.86%**

[*1] *Res.* stands for *Residential*

considered. The main clustering problem in this benchmark set is the occurrence of 2 poorly separable and touching clusters as shown in Fig. 2a. After some experimentation, the following values of control parameters were selected: $win_{min} = 2$, $win_{max} = 4$, and $d_{coef} = 4$. Fig. 2 illustrates the operation of our clustering technique. Fig. 2a, as already said, represents the data, Figs. 2b through 2f show the evolution of the tree-like structure of our generalized S-ONN in data set at different stages of the learning process, and finally Figs. 2g and 2h present the plots of the number of neurons and the number of subnetworks (equal to the number of detected clusters), respectively, versus learning epoch number. Our approach automatically adjust the number of neurons in the network (starting from the initial number of 2 neurons) and detects the correct number of data clusters (i.e., 2 clusters) in the considered data set by disconnecting the tree-like structure of the generalized S-ONN into 2 subnetworks. Sets of neurons in particular subnetworks represent multi-point prototypes of particular clusters.

IV. APPLICATION TO IRISH CER ELECTRICITY CONSUMPTION DATA CLUSTERING AND COMPARATIVE ANALYSIS WITH ALTERNATIVE APPROACHES

The performance of our clustering technique based on generalized S-ONNs will now be verified in the real-life, complex Irish CER data clustering problem. Due to a very high dimensionality of the considered data and after some experimentations, the parameters win_{min} and win_{max} which control the range of changes in the number of neurons, were set to 70 and 90, respectively. The third control parameter d_{coef} remains the same (i.e., $d_{coef} = 4$) as in low-dimensional benchmark data clustering in previous section. Fig. 3 illustrates the progress of the learning and thus, the clustering process of the considered data. Our approach automatically adjusts both the overall number of neurons in the network (starting from the initial number of 2 neurons) - see Fig. 3a - and the number of disconnected subnetworks (equal to the number of detected clusters) - see Fig. 3b. Finally, 2 clusters are found in the considered data set. Each of them is represented by a separate subnetwork of 3 and 41 neurons, respectively. The neuron subsets define multi-point prototypes of particular clusters. After the network calibration (see [14] for comments), the 3-neuron subnetwork is labeled *SME* and the 41-neuron subnetwork is labeled *Residential*, according to the labels of clusters they represent.

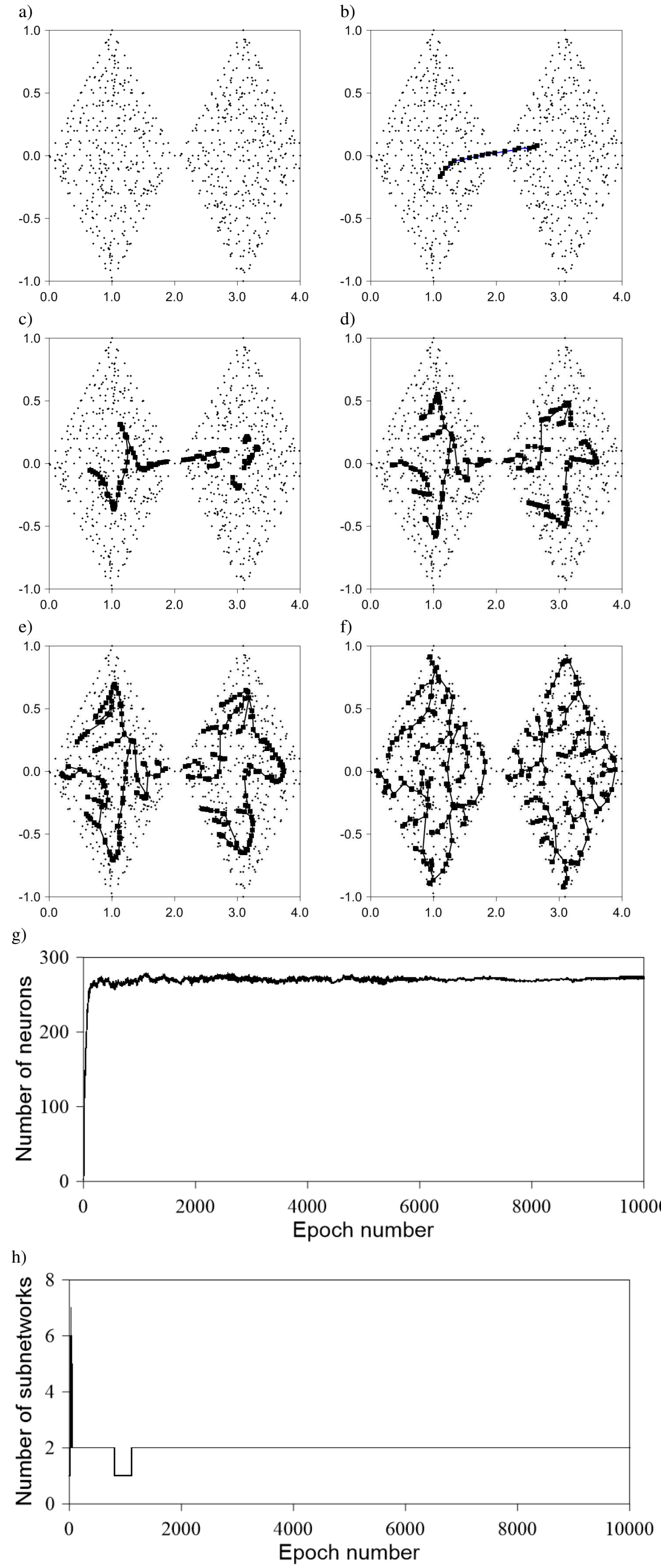

Fig. 2. *TwoDiamonds* data set (a) and the evolution of the structure of the generalized S-ONN in it in learning epochs: b) No. 5, c) No. 50, d) No. 100, e) No. 500, and f) No. 10 000 (end of learning), as well as plots of the number of neurons (g) and the number of subnetworks (clusters) (h) vs. epoch number

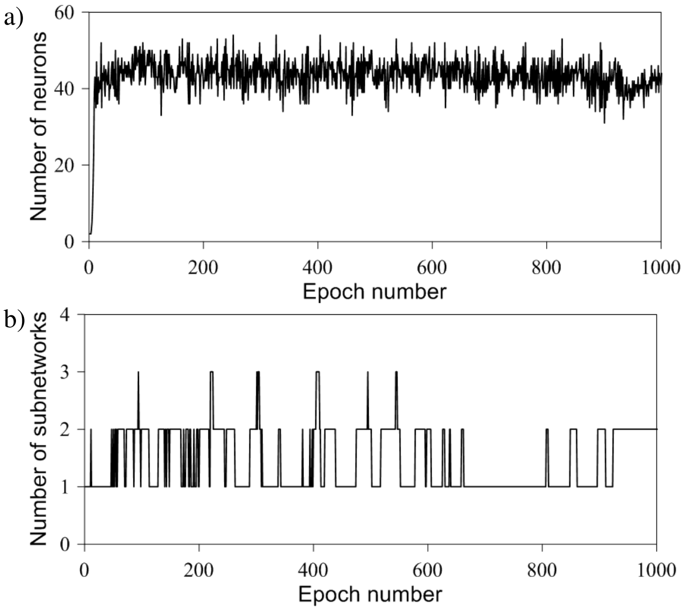

Fig. 3. Plots of the number of neurons (a) and the number of subnetworks (clusters) (b) vs. epoch number for Irish CER data set

Since the number of clusters and the assignments of particular data samples to clusters is known in the original Irish CER data set (let's emphasize again that this knowledge is **not used** by our clustering method), a direct verification of the obtained results is also possible as shown in Table I. It can be seen that the percentage of correct decisions for *Residential* customers is very high (98.63%) but the overall clustering accuracy (94.86%) is negatively affected by a relatively low accuracy (i.e., 62.76%) for *SME* customers. In order to investigate this problem, some details regarding the clustering process for the electricity consumption data in two selected weeks: one during summer (July 20-26, 2009) and one during winter (December 7-13, 2009) are presented in Figs. 4 and 5, respectively.

Fig. 4a shows multi-point prototypes for particular clusters. *SME*-cluster prototype consists of 3 sub-profiles and *Residential*-cluster prototype - of 41 sub-profiles describing, in detail, customer sub-groups within particular clusters. For better readability, the plots for the first and last day of the considered week are presented in Figs. 4b and 4c, respectively. The electricity consumption plots for *Residential* customers correctly and incorrectly classified are shown in Figs. 4d and 4e, respectively. Similarly, Figs. 4f and 4g show the electricity consumption plots for *SME* customers correctly and incorrectly classified. Fig. 5 presents analogous information for the selected winter week.

Based only on a visual inspection of Figs. 4 and 5, we can find that the incorrectly classified *SME* customers (see Figs. 4g and 5g) have completely different electricity consumption profiles than the correctly classified *SME* customers (see Figs. 4f and 5f). Therefore, we may suspect that some *SME* customers are much closer to *Residential* customers in terms of their electricity consumption profiles than to other *SME* customers. Following that, we made the following experiment.

TABLE II

CLUSTERING RESULTS FOR MODIFIED IRISH CER DATA SET

Cluster label	Number of samples	Number of decisions for subnetwork labeled:		Number of correct decisions	Number of wrong decisions	Percentage of correct decisions
		*Res.-m.**1	*SME-m.**1			
*Res.-m.**1	3639+106	3589+106	50	3589+106	50	98.66%
*SME-m.**1	427-106	159-106	268	268	159-106	83.49%
ALL	**4066**	**3748**	**318**	**3963**	**103**	**97.47%**

*1 *Res.-m.* and *SME-m.* stand for modified *Residential* and modified *SME* customer groups

TABLE III

COMPARATIVE ANALYSIS WITH ALTERNATIVE CLUSTERING TECHNIQUES FOR IRISH CER DATA SET

Method	Cluster label	Number of samples	Number of decisions for subnetwork labeled:		Number of correct decisions	Number of wrong decisions	Percentage of correct decisions
			*Res.**1	*SME*			
Density-based clustering	*Res.**1	3639	3626	13	3626	13	99.64%
	SME	427	200	227	227	200	53.16%
	ALL	4066	3826	240	3853	213	94.76%
k-means	*Res.**1	3639	3639	0	3639	0	100.0%
	SME	427	325	102	102	325	23.89%
	ALL	4066	3964	102	3741	325	92.01%
Farthest First	*Res.**1	3639	3639	0	3639	0	100.0%
	SME	427	416	11	11	416	2.58%
	ALL	4066	4055	11	3650	416	89.77%
EM	*Res.**1	3639	3522	117	3522	117	96.78%
	SME	427	140	287	287	140	67.21%
	ALL	4066	3662	404	3809	257	93.68%
Our approach	*Res.**1	3639	3589	50	3589	50	98.63%
	SME	427	159	268	268	159	62.76%
	ALL	**4066**	**3748**	**318**	**3857**	**209**	**94.86%**

*1 *Res.* stands for *Residential*

Based on surveys attached to data records (unfortunately, not all surveys are available), we found that 106 out of 155 incorrectly classified *SME* customers are enterprises employing 5 or less workers. We changed, experimentally, their labels from *SME* to *Residential*. The new clustering results - presented in Table II - show a significant increase in the percentage of correct decisions for the modified *SME* group. It confirms our hypothesis that a significant number of *SME* customers have electricity consumption profiles much closer to *Residential* ones than to the remaining part of the *SME* group.

Using WEKA software [18] - a widely known open source framework for data mining algorithms - we also perform a comparative analysis of our approach (the results of Table I are considered) and 4 alternative methods including Density-based, *k*-means, Farthest-first, and Expectation-maximization (EM) clustering algorithms. The alternative methods require, however, the specification of the number of clusters to be found in data. Therefore, they are favored over our approach which does not need such a specification. Despite of that, our approach - as shown in Table III - outperforms all of them in terms of the overall clustering accuracy and - together with EM approach - give the best balanced accuracies for both clusters.

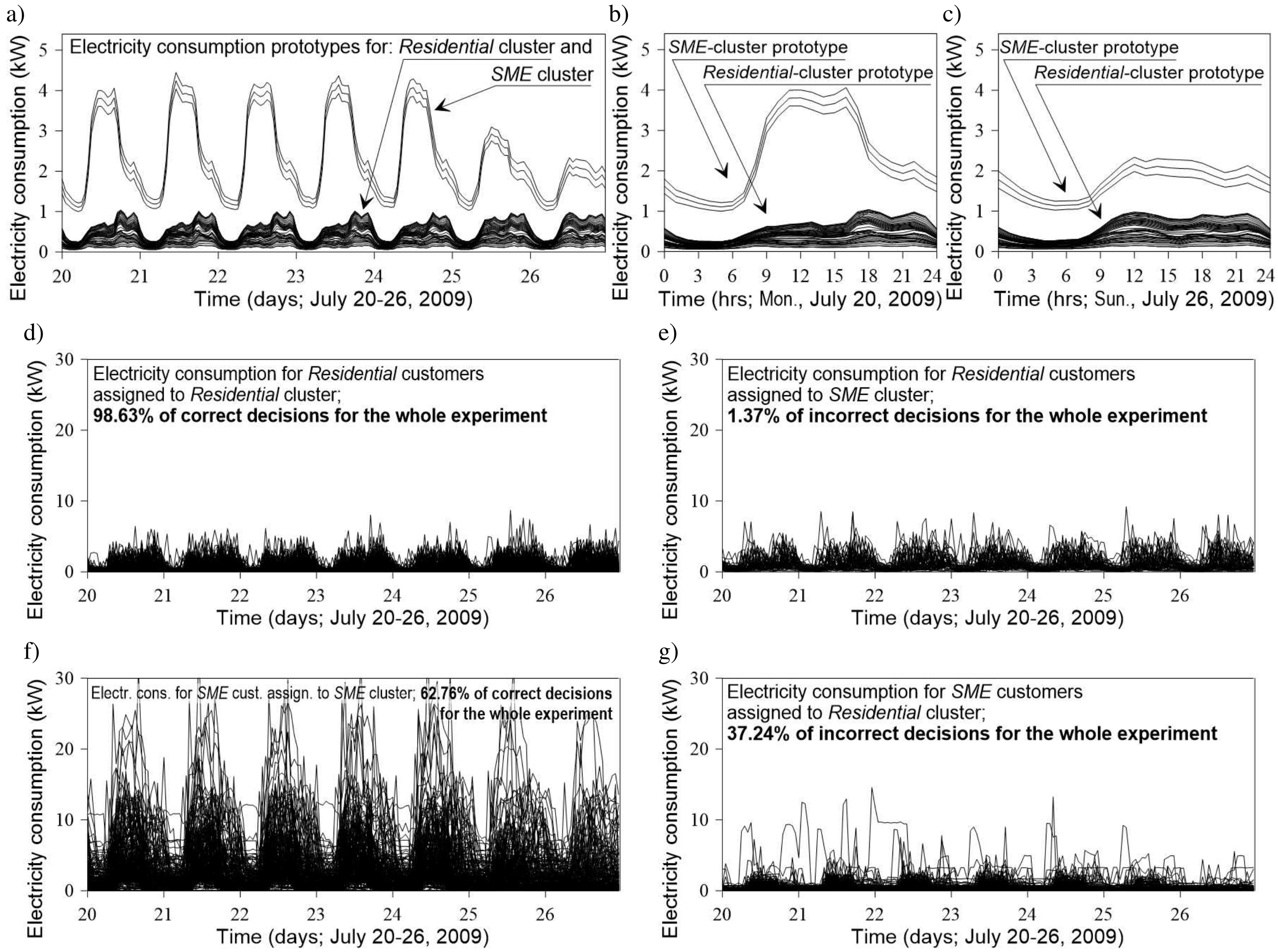

Fig. 4. Prototypes of electricity consumption for *Residential* cluster and *SME* cluster in exemplary week of July, 2009 (a), on Monday (b), and Sunday (c) of that week; plots of electricity consumption for *Residential* customers correctly (d) and incorrectly classified (e) as well as plots of electricity consumption for *SME* customers correctly (f) and incorrectly classified (g)

V. Conclusions

In this paper, we present the application of our clustering technique based on proposed by us generalized S-ONNs with evolving tree-like splitting-merging structures to the clustering of Irish CER electricity consumption data collected as a part of a smart metering pilot study conducted by that institution. First, the Irish CER data are briefly characterized. Then, the operation of our clustering technique is outlined and illustrated using a benchmark data set. In turn, the application of our approach to the Irish CER data clustering is presented, evaluated, and discussed as well as a comparative analysis with several alternative approaches is performed.

References

[1] “Navigant Research. Smart meters global market analysis and forecasts,” http://www.navigantresearch.com/research/smart-meters, 2014, [Online; accessed 07.07.14].

[2] A. Mutanen, M. Ruska, S. Repo, and P. Jarventausta, “Customer classification and load profiling method for distribution systems,” *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1755–1763, July 2011.

[3] M. B. Gorzałczany, F. Rudziński, and J. Piekoszewski, “Gene expression data clustering using tree-like SOMs with evolving splitting-merging structures,” in *Proc. 2016 IEEE Int. Conf. Neural Networks (IJCNN), 25-29 July 2016*, Vancouver, Canada, 2016, pp. 3666–3673.

[4] M. B. Gorzałczany, J. Piekoszewski, and F. Rudziński, “Generalized tree-like self-organizing neural networks with dynamically defined neighborhood for cluster analysis,” in *Artificial Intelligence and Soft Computing - ICAISC 2014*, ser. Lecture Notes in Computer Science, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Żurada, Eds. Berlin: Springer-Verlag, 2014, vol. 8468, pp. 725–737.

[5] M. B. Gorzałczany, J. Piekoszewski, and F. Rudziński, “Microarray leukemia gene data clustering by means of generalized self-organizing neural networks with evolving tree-like structures,” in *Artificial Intelligence and Soft Computing - ICAISC 2015*, ser. Lecture Notes in Computer Science, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Żurada, Eds. Springer-Verlag, 2015, vol. 9119, pp. 15–25.

[6] M. B. Gorzałczany, F. Rudziński, and J. Piekoszewski, “Generalized SOMs with splitting-merging tree-like structures for WWW-document clustering,” in *2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT-15)*, J. M. Alonso, H. Bustince, and M. Reformat, Eds., vol. 89. Gijón, Spain: Atlantis Press, Jun. 30 – Jul. 3 2015, pp. 186–193.

[7] “Irish social science data archive. CER smart metering project,” https:

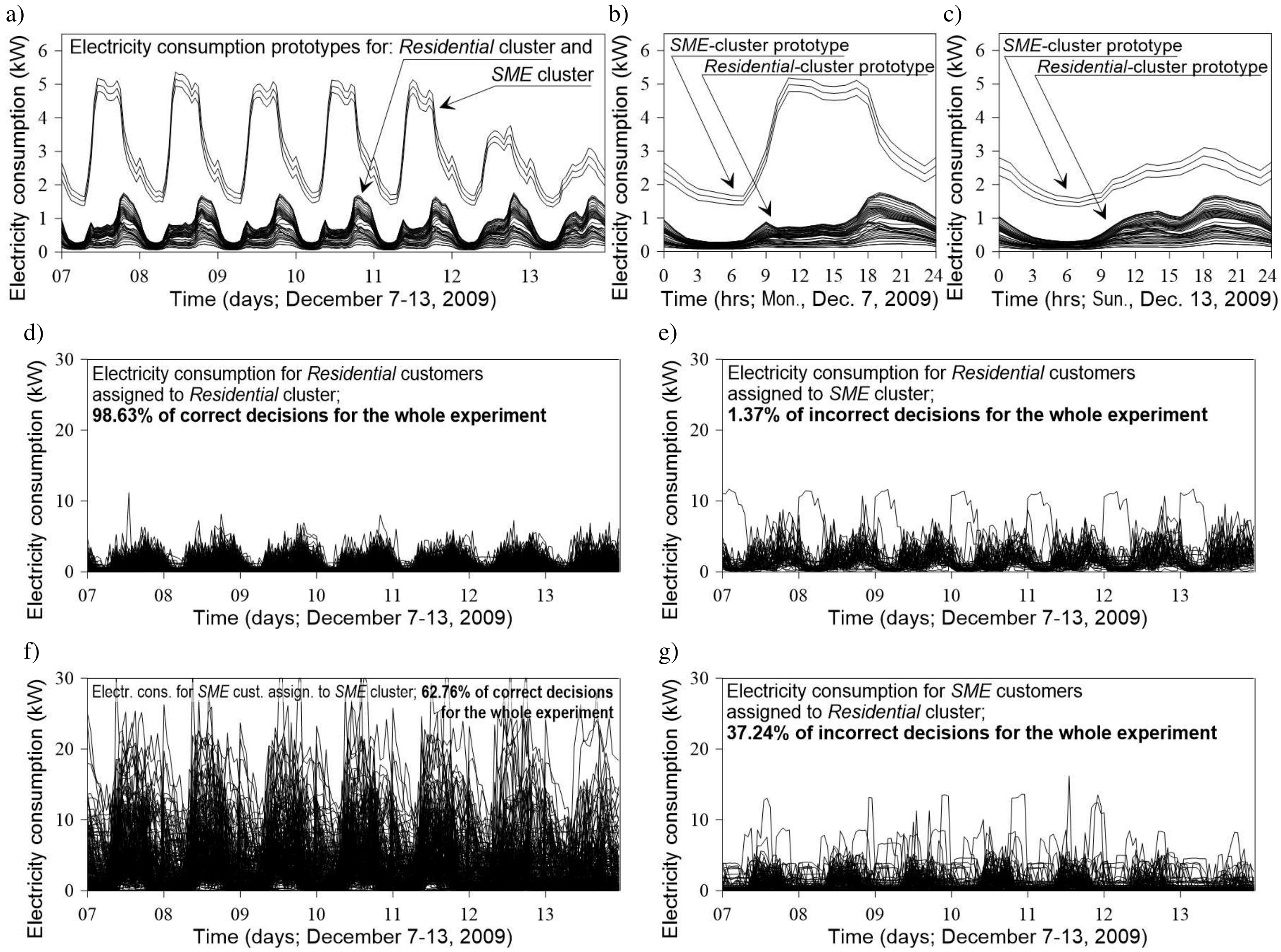

Fig. 5. Prototypes of electricity consumption for *Residential* cluster and *SME* cluster in exemplary week of December, 2009 (a), on Monday (b), and Sunday (c) of that week; plots of electricity consumption for *Residential* customers correctly (d) and incorrectly classified (e) as well as plots of electricity consumption for *SME* customers correctly (f) and incorrectly classified (g)

//www.ucd.ie/issda, 2012, [Online].

[8] J. A. Balaji, H. D. S. Ram, and B. B. Nair, *Machine Learning Approaches to Electricity Consumption Forecasting in Automated Metering Infrastructure (AMI) Systems: An Empirical Study*. Cham: Springer International Publishing, 2017, pp. 254–263.

[9] M. B. Gorzałczany and F. Rudziński, "Generalized self-organizing maps for automatic determination of the number of clusters and their multiprototypes in cluster analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. PP, no. 99, pp. 1–13, 2017.

[10] M. B. Gorzałczany and F. Rudziński, "Cluster analysis via dynamic self-organizing neural networks," in *Artificial Intelligence and Soft Computing - ICAISC 2006*, ser. Lecture Notes in Computer Science, L. Rutkowski, R. Tadeusiewicz, L. A. Zadeh, and J. M. Żurada, Eds. Berlin: Springer-Verlag, 2006, vol. 4029, pp. 593–602.

[11] M. B. Gorzałczany and F. Rudziński, "WWW-newsgroup-document clustering by means of dynamic self-organizing neural networks," in *Artificial Intelligence and Soft Computing - ICAISC 2008*, ser. Lecture Notes in Computer Science, L. Rutkowski, R. Tadeusiewicz, L. A. Zadeh, and J. M. Żurada, Eds. Berlin: Springer-Verlag, 2008, vol. 5097, pp. 40–51.

[12] M. B. Gorzałczany and F. Rudziński, "Application of genetic algorithms and Kohonen networks to cluster analysis," in *Artificial Intelligence and Soft Computing - ICAISC 2004*, ser. Lecture Notes in Computer Science, L. Rutkowski, R. Tadeusiewicz, L. A. Zadeh, and J. H. Siekmann, Eds. Berlin: Springer-Verlag, 2004, vol. 3070, pp. 556–561.

[13] M. B. Gorzałczany and F. Rudziński, "Modified Kohonen networks for complex cluster-analysis problems," in *Artificial Intelligence and Soft Computing - ICAISC 2004*, ser. Lecture Notes in Computer Science, L. Rutkowski, R. Tadeusiewicz, L. A. Zadeh, and J. H. Siekmann, Eds. Berlin: Springer-Verlag, 2004, vol. 3070, pp. 562–567.

[14] T. Kohonen, *Self-Organizing Maps*, 3rd ed. Berlin: Springer-Verlag, 2001.

[15] N. R. Pal, J. C. Bezdek, and E. C.-K. Tsao, "Generalized clustering networks and Kohonen's self-organizing scheme," *IEEE Transactions on Neural Networks*, vol. 4, no. 4, pp. 549–557, 1993.

[16] J. C. Bezdek, T. R. Reichherzer, G. S. Lim, and Y. Attikiouzel, "Multiple-prototype classifier design," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 28, no. 1, pp. 67–79, 1998.

[17] A. Ultsch, "Clustering with SOM: U*C," in *Proceedings of the Workshop on Self-Organizing Maps*, Paris, France, 2005, pp. 75–82.

[18] "Weka 3: Data mining software in java," https://www.cs.waikato.ac.nz/ml/weka, 2014, [Online].