Available online at www.sciencedirect.com
ScienceDirect
Applied Energy 85 (2008) 830–840
APPLIED ENERGY
www.elsevier.com/locate/apenergy

Reducing energy consumption by using self-organizing maps to create more personalized electricity use information

Teemu Räsänen *, Juhani Ruuskanen, Mikko Kolehmainen

Research Group of Environmental Informatics, Department of Environmental Sciences, University of Kuopio, P.O. Box 1627, FIN-70211 Kuopio, Finland

Received 5 September 2007; received in revised form 17 October 2007; accepted 22 October 2007
Available online 14 February 2008

Abstract

Identification of electricity use is one of the key elements to motivate customers to promote activities leading more efficient use of energy. Furthermore, electricity use comparisons with other similar customers give more interesting and concrete point of view to examine own consumption habits. In future, electricity providers and retailers are willing and probably forced by legislation to provide such information by the means of energy conservation and efficiency improvement. On the other hand, high number of customers set challenges to handle electricity use data and to create proper comparison information. In this study we present efficient and highly automated way to create comparison groups based on customers building characteristics. The main advantages of the data-based approach are that customer location is noticed, comparison groups are created using concrete building information, data processing is highly automated and also method is computationally efficient. Additionally, presented method provide tool to target and to create customer specific electricity saving guidance. The performance of suggested approach was tested using data set which contained electricity use and building information concerning almost 8000 customers.

Keywords: Electricity use; Energy efficiency; Clustering; Self-organizing map; *K*-means

1. Introduction

Increasing concern about global climate change has led to greater interest in energy consumption. It is well known that changes in consumption habits, leading to more efficient use of energy sources, are essential. Recent European Union legislation reflects this concern and more stringent laws concerning efficient use of energy are in the pipeline. This means that energy distributors and retail energy sales companies selling electricity or other types of energy should provide information about past and current consumption, load profiles, customer segmentation, geographical location of customers, comparisons of the consumer's current energy consumption with consumption for the same period in the previous year in graphical form and comparisons

* Corresponding author. Tel.: +358 17 162373; fax: +358 17 163191.
E-mail address: teemu.rasanen@uku.fi (T. Räsänen).

0306-2619/$ - see front matter
doi:10.1016/j.apenergy.2007.10.012

with an average normalized or benchmarked user of energy of the same category [1]. Such information is expected to motivate consumers to promote activities leading to energy conservation and efficiency improvement [2,3].

Energy conservation involves reduced energy consumption through e.g. lower heating levels or consumption limits on appliances. It is often influenced by regulations, consumer behavior and lifestyle changes. On the other hand, energy efficiency means getting the most out of every unit of energy and it is a by-product of other social goals such as productivity, comfort, monetary savings or fuel competition [4]. An informative bill is one of the generally used ways to give individual consumers analyzed and understandable information [5]. Visual graphics and tables can be used to help consumers to understand their electricity consumption and perhaps change their consumption behavior. Internet-based information services are also useful and low cost, especially for providing detailed real-time electricity use information [3]. Moreover, it has been shown that enabling consumers to compare their consumption with that of similar customers is one way to affect customer's behavior and consumption habits.

In previous studies, comparison groups were created using one of four criteria: meter read route, street, billing cycle, and house characteristics. The house characteristics used were floor area, housing type, and heating fuel [2]. The overall conclusion was that multiple house characteristic data, street name or meter book are suitable for the creation of statistically valid comparison groups. On the other hand, customers have been classified into groups based on electricity use load-patterns. In these approaches, the classification is done on the basis of their "electrical behavior", which is a key factor for setting up new tariff offers, leading to tariff structures more closely related to actual costs of electricity provision in different time periods [6]. However, the methods used to group customers should be improved so that each customer's location is included in the clustering process. The environmental characteristics of consumers' location, including climate conditions, have a significant effect on electricity consumption. In addition, providing information about the consumption of similar customers in the neighborhood makes electricity use information more concrete and understandable. Neighborhood comparisons may encourage communication among customers regarding the methods of energy conservation [2]. It should also be possible for energy companies to cluster large numbers of customers automatically, without time-consuming manual work.

The purpose of this study was to develop new methods for handling increasing amounts of customer and electricity use data in order to create new services leading to more efficient use of electricity. In this paper, we present a highly automated and efficient data-based method for creating more personalized information about consumers electricity use containing realistic comparison information. Overall, this study presents tools for energy distributors and electricity retailers to meet tightening legislation concerning energy efficiency set by the European Commission.

2. Materials and methods

2.1. Data used

In this study, we use data describing almost 8000 customer's characteristics and annual electricity use. The customers were located near Salo which is town in southwestern Finland. The city of Salo has about 25,000 inhabitants and the largest sector of the economy is services at 47.8% (including community services 19.1%), followed by manufacturing and refining at 46.5%. Primary production is low at 1.5%. The largest industries are electronics, machinery and other engineered metal and plastic products.

The climate in Southern Finland is a northern temperate climate. The main factor influencing Finland's climate is the country's geographical position between the 60th and 70th northern parallels in the Eurasian continent's coastal zone, which shows characteristics of both a maritime and a continental climate, depending on the direction of air flow. Furthermore, Finland is near enough to the Atlantic Ocean to be continuously warmed by the Gulf Stream, which explains the unusually warm climate considering the absolute latitude. Annual mean temperature in southern Finland was about 7° below zero in the year 2006 [7].

In the test group some of the customers were living in the city, densely populated area, and others were located in the rural areas nearby. Annual electricity use information was provided by Finnish energy company Fortum Ltd., and building data was provided by Population Register Centre. The electricity use data contained information from 1991 to 2004. The building data contained variables describing characteristics of buildings and variables used in clustering are illustrated in Table 1. These data sets, coming from different organizations, were linked together using customers and buildings street addresses. The building data contained also coordinates to indicate buildings location which was used in creating actual comparison groups.

Data quality is important element in creation of successful data-based approaches. The main goal of data quality evaluation and assessment is to find out if the data is usable for its planned purpose. In this case, data quality assessment was carried out by data providers using their best practices in order to maintain data reliable and useful.

2.2. Data processing chain

In this study, automated data processing method were developed to convert large amount of customer related data into more useful comparing group information. The main elements of the data processing chain are summarized in Fig. 1 and theory of used computational methods is shortly introduced in the following sections.

The proposed approach contains eight main phases, starting with data preprocessing where data sets were connected using address information and variable specific modifications were carried out. Finally, after the several data processing phases' actual comparison information for the each customer was illustrated using simple graphical figures.

2.3. Self-organizing map

The self-organizing map is effective software tool for the visualization and modeling of high-dimensional data and also it is one of the best known unsupervised neural learning algorithms. The goal of the SOM

Table 1
The variables used in customer clustering

Variable	Type of variable
Buildings construction year	Nominal
Number of flats in building	Nominal
Number of floors in building	Nominal
Number of rooms in flat	Nominal
Residential area of flat or building	Nominal (m^2)
Property features; electricity	Class (0 or 1)
Property features; gas	Class (0 or 1)
Property features; sewer	Class (0 or 1)
Property features; water	Class (0 or 1)
Property features; warm water	Class (0 or 1)
Property features; solar panel	Class (0 or 1)
Property features; elevator	Class (0 or 1)
Property features; air conditioning	Class (0 or 1)
Property features; sauna	Nominal
Property features; normal kitchen	Class (0 or 1)
Property features; kitchenette	Class (0 or 1)
Property features; small kitchen	Class (0 or 1)
Property features; large kitchen	Class (0 or 1)
Frame; concrete	Class (0 or 1)
Frame; brick	Class (0 or 1)
Frame; wooden	Class (0 or 1)
Frame; other material	Class (0 or 1)
Heating; circulating water	Class (0 or 1)
Heating; air	Class (0 or 1)
Heating, electricity	Class (0 or 1)
Heating; wood	Class (0 or 1)
Heating; non	Class (0 or 1)

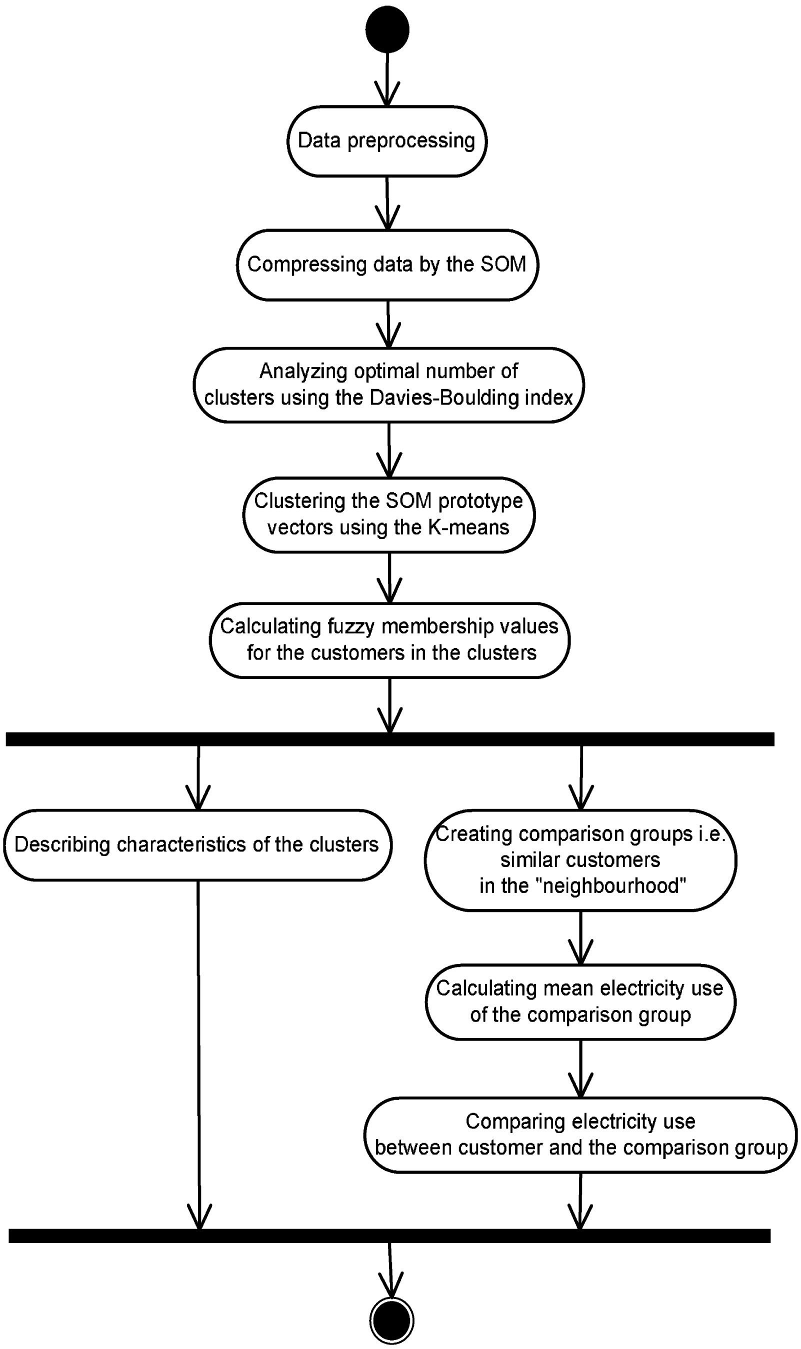

Fig. 1. The main elements of data processing chain.

algorithm is to find prototype vectors that represent the input data set and at the same time realize a continuous mapping from input space to a lattice. This lattice consists of a defined number of 'neurons' and may, for example, be a two-dimensional map that is easily visualized [8].

The basic principle behind the SOM algorithm is that the weight vectors of neurons which are first initialized randomly, come to present a number of original measurement vectors during an iterative data input process. With each training pattern the winning neuron, called best-matching unit (BMU), is first found by comparing the input and weight vectors of neurons by Euclidean distance metrics. The weights of the winning neuron (BMU) and its neighbors are then moved towards the input vector according to a learning rate factor which decreases monotonically towards the end of learning.

So, each data row, which describes in this case characteristics of customer, is classified into neuron which contains most similar customers. Neurons prototype vector presents characteristics of all the customers in each neuron. In this case, the training data for SOM contained 27 variance scaled inputs. The SOM having 100 neurons in a 10×10 hexagonal arrangement was constructed. The linear initialization and batch training algorithms were used in the training of the map and map was taught with 10 rounds. Gaussian function was used as the neighborhood function and the initial neighborhood had the value of 3.

2.4. K-means clustering

The number of clusters in the case specific application may not be known a priori. In the *K*-means algorithm the number of clusters has to be predefined. It is common that the algorithm is applied with different number of clusters and then the best solution among them is selected using a validity index. The Davies–Bouldin (DB) index [9] is calculated as follows:

$$\mathrm{DB} = \frac{1}{N} \sum_{i=1}^{N} \max_{j, j \neq i} \frac{S_i + S_j}{d_{ij}} \tag{1}$$

where N is the number of clusters. The within (S_i) and between (d_{ij}) cluster distances are calculated using the cluster centroids as follows:

$$S_i = \frac{1}{|C_i|} \sum_{x \in C_i} \|x - m_j\| \tag{2}$$

$$d_{ij} = \|m_i - m_j\| \tag{3}$$

where m_i is the centre of cluster C_i, with $|C_i|$ the number of points belonging to cluster C_i. The objective is to find the set of clusters that minimizes Eq. (3).

The Davies–Bouldin index was used to solve optimal number of clusters. Index was calculated for SOM prototype vectors. The DB index varies slightly between calculations because initial starting point is set randomly. In this case, indexes was calculated 20 times and mean value of the index using different numbers of clusters was used when optimal number of clusters was selected. After that *K*-means algorithm was used to cluster SOM prototype vectors in order to create reasonable number of comparing groups.

The *K*-means algorithm was applied to the clustering of the self-organizing map prototype vectors. The method is a well-known non-hierarchical cluster algorithm [10]. The basic version begins by choosing number of clusters and randomly picking *K* cluster centers. After that each point is assigned to the cluster whose mean is closest in a Euclidean distances sense. Finally, the mean vectors of the points assigned to each cluster are computed, and those are used as new centers in an iterative approach until convergence criterion is met.

2.5. Fuzzy membership values

In the clustering every data point is forced to some cluster. An ideal case happens when all the data point are identical inside the each cluster. However, in most of the applications clusters have several data points which diverge from points located near cluster center. It is obvious that if the characteristics of cluster center are used for decisions, such decisions would not be so reliable concerning data points which are located in the same cluster but far away from cluster center. Fuzzy logic is mathematical way to handle uncertainty [11] and for reasons above, fuzzy membership values was calculated between each data point (customer) and cluster center. These values were used type of dissimilarity measures, where 0 means that data point is really far from the group and 1 means that data point belongs perfectly to the group. The membership values were calculated using group-average linkage method and second-order exponential membership function. The group-average linkage method first evaluates the average point for each group and calculates the distances between these group-averages and all data records.

2.6. Creating comparison groups

After the clustering, the characteristics of clusters were analyzed using graphical visualizations and numerical summaries. The short descriptions for clusters were created. The descriptions contained main features of each cluster and differences in comparison with other clusters. Actual comparing group was constructed for the each customer by picking up only customers who belong to the same cluster and locates in the customers 'neighborhood'. The building data contained spatial coordinates for each building and using that information neighborhood was defined. Size of neighborhood is case specific depending on areas population density and changeable climate circumstances. In this case, neighborhood buffer radius was set to one kilometer around the building.

After the creation of actual comparing group, mean value of group's annual electricity use was calculated. Finally, each customer's and its comparing group's electricity use in proportion to apartment's floor area was compared using numerical analyses and graphs. Furthermore, we used data of last three years in comparison and analyzed how customer's and group's electricity use has been developed during this period. The characteristics of clusters were also used to create customer specific electricity saving guidance.

3. Results

With the aim of creating more personalized information concerning customer's electricity use, multi-phased data processing was constructed. Presented approach was applied using data describing almost 8000 customer's annual electricity use and building characteristics. As a result of clusterization, customers were clustered into 12 clusters. Number of customers in each cluster is illustrated in Table 2 and short textual descriptions are presented in Table 3. In order to identify the clusters in more detailed level, statistics for the clusters were calculated. Mean values of the cluster centers are given in Table 4 and standard deviations are shown in Table 5. Furthermore, Table 6 presents differences between clusters according to electricity use.

Actual comparison groups for each customer were created so that only similar customers in the "neighborhood" were grouped. Example for creation of comparison group is presented in Fig. 2, where customer belonging cluster 2 is marked using star. The comparison group for this customer contained 87 customers who had similar characteristics and located less than kilometer from the customer. The customers which belong to the comparison group are marked with triangle and other customers are marked with dot.

Table 2
Number of customers in each cluster

	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12
Number of customers	1196	1017	579	86	637	518	1358	1063	232	475	544	145

Table 3
Characterization of the clusters by textual descriptions

Cluster	Description
c1	Old three room apartments in wood framed terraced house, sauna, circulating water heating
c2	Two or three room flats in four floor block of flat, sauna, elevator, air conditioning, circulating water heating
c3	Two or three room flats in three floor block of flat , circulating water heating
c4	Three room flats in two floor small block of flat, sauna, large kitchen, no heating, unknown frame material
c5	Large wood framed detached houses, sauna, circulating water heating
c6	Three room apartments in concrete or wood framed terraced house, gas, sauna, circulating water heating
c7	Four room apartments in wood framed terraced house, electric heating
c8	Small one or two room flats in concrete framed block of flats, air conditioning, circulating water heating
c9	Newish small one or two room flats in concrete framed block of flats, sauna, air condit., circ. water heating
c10	Old small wood framed detached houses, no warm water, sauna, heated using electric combined with wood
c11	Large two or three room flats in brick framed small block of flat, circulating water heating
c12	Old small two or three room flats in concrete framed small block of flats, circulating water heating

Table 4
Descriptions of the clusters c1–c12 by mean values of variables

Variable	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12
Buildings construction year	1962	1986	1971	1980	1974	1989	1986	1977	1995	1948	1970	1962
Number of flats in building	2.20	23.97	22.99	12.78	1.18	6.45	1.98	20.59	19.36	1.40	13.87	12.86
Number of floors in building	1.44	4.01	3.29	2.17	1.30	1.32	1.13	3.16	2.66	1.47	2.62	2.57
Number of rooms in flat	3.35	2.55	2.41	2.90	4.99	2.70	3.98	1.72	1.58	2.87	2.58	2.55
Residential area of flat	87.99	68.56	63.01	82.51	140.07	74.85	107.77	42.79	41.97	76.04	80.78	65.00
Property features; electricity	1.00	1.00	1.00	0.17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Property features; gas	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00
Property features; sewer	1.00	1.00	1.00	0.84	1.00	1.00	1.00	1.00	1.00	0.90	1.00	1.00
Property features; water	1.00	1.00	1.00	0.84	1.00	1.00	1.00	1.00	1.00	0.87	1.00	1.00
Property features; warm water	1.00	1.00	1.00	0.16	0.99	1.00	1.00	1.00	1.00	0.12	0.97	1.00
Property features; solar panel	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
Property features; elevator	0.01	0.82	0.08	0.00	0.02	0.01	0.11	0.43	0.71	0.01	0.17	0.00
Property features; air conditioning	0.00	0.91	0.36	0.02	0.18	0.79	0.39	0.65	0.87	0.01	0.37	0.01
Property features; sauna	0.84	0.71	0.01	0.35	0.99	0.91	0.96	0.35	0.75	0.63	0.27	0.09
Property features; normal kitchen	0.99	1.00	1.00	0.53	0.98	1.00	0.99	0.00	0.00	0.96	0.56	1.00
Property features; kitchenette	0.01	0.00	0.00	0.01	0.02	0.00	0.01	1.00	0.00	0.02	0.17	0.00
Property features; small kitchen	0.00	0.00	0.00	0.10	0.01	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Property features; large kitchen	0.00	0.00	0.00	0.35	0.00	0.00	0.00	0.00	0.00	0.01	0.27	0.00
Frame; concrete	0.00	1.00	1.00	0.00	0.03	0.25	0.03	0.83	0.80	0.02	0.15	0.88
Frame; brick	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.74	0.00
Frame; wood	1.00	0.00	0.00	0.19	0.97	0.75	0.97	0.17	0.20	0.98	0.11	0.12
Frame; other material	0.00	0.00	0.00	0.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Heating; circulating water	1.00	1.00	1.00	0.15	0.92	1.00	0.00	0.99	0.93	0.04	0.87	1.00
Heating; air	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Heating, electricity	0.00	0.00	0.00	0.06	0.00	0.00	1.00	0.01	0.07	0.44	0.13	0.00
Heating; wood	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.52	0.00	0.00
Heating; non	0.00	0.00	0.00	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

The fuzzy membership value was calculated for the chosen customer and value 0.61 indicates that customer belongs clearly to the comparison group but differs from actual cluster center. The mean annual electricity use in proportion of floor area was then calculated using customers belonging to the comparison group. Afterwards, customer's and comparison groups' annual consumption information was visualized using bar graph, which is presented in Fig. 3.

Moreover, proposed method was used to find out customers consuming more electricity than average which was proper basis to target and to create personal electricity saving guidance. The characteristics of comparison groups were used to create customer specific tips how to save electricity. For example, guidance for customers in c2 cluster, where most of the customers lives in a block of flats, contained information as follows [13]:

– Switch on TV completely, don't leave it on standby.
– Use task lighting to target work and leisure activities.
– Use energy saving bulbs in lightning.
– Defrost your fridge and freezer on a regular basis or whenever necessary.
– Use electrical appliances which are rated most energy efficient appliances according to EU Energy Labeling.

On the other hand, electricity saving guidance for customers in cluster c10 should focus more on renovation and efficient heating of old house, checking and improving insulation of walls and roofs or even changing warming system to more efficient one.

Furthermore, as a result of research project, presented data processing and clustering method was constructed in the form of pilot application running under the Matlab computing environment. The application contained customer search, multiple visualizations describing characteristics of each comparison group, and electricity use comparisons.

Table 5
Descriptions of the clusters c1–c12 by standard deviation of variables

Variable	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12
Buildings construction year	14	12	9	17	15	9	18	14	7	19	13	11
Number of flats in building	2.69	8.33	6.04	12.08	0.99	2.50	1.89	9.66	9.72	1.33	11.79	5.67
Number of floors in building	0.50	1.52	0.64	0.72	0.46	0.49	0.35	1.36	1.16	0.50	1.44	0.75
Number of rooms in flat	0.89	0.75	0.72	1.13	0.92	0.99	1.18	0.71	0.63	1.19	1.46	1.11
Residential area of flat	21.96	16.77	12.46	37.77	34.66	23.16	39.08	13.78	11.33	31.36	49.41	22.80
Property features; Electricity	0.00	0.00	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00
Property features; gas	0.00	0.09	0.00	0.00	0.04	0.12	0.04	0.08	0.11	0.00	0.00	0.00
Property features; sewer	0.00	0.00	0.00	0.37	0.00	0.00	0.00	0.00	0.00	0.30	0.00	0.00
Property features; water	0.00	0.00	0.00	0.37	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00
Property features; warm water	0.00	0.00	0.00	0.37	0.10	0.00	0.04	0.00	0.00	0.33	0.16	0.00
Property features; solar panel	0.00	0.12	0.00	0.00	0.04	0.00	0.00	0.50	0.13	0.00	0.00	0.00
Property features; elevator	0.08	0.38	0.27	0.00	0.15	0.09	0.32	0.48	0.46	0.08	0.38	0.00
Property features; air conditioning	0.05	0.28	0.48	0.15	0.38	0.41	0.49	0.48	0.34	0.09	0.48	0.12
Property features; sauna	0.37	0.45	0.10	0.48	0.12	0.29	0.20	0.00	0.43	0.48	0.44	0.29
Property features; normal kitchen	0.12	0.06	0.00	0.50	0.15	0.00	0.12	0.00	0.00	0.19	0.50	0.00
Property features; kitchenette	0.12	0.06	0.00	0.11	0.12	0.00	0.12	0.00	0.00	0.15	0.37	0.00
Property features; small kitchen	0.00	0.00	0.00	0.31	0.09	0.00	0.00	0.00	0.00	0.06	0.04	0.00
Property features; large kitchen	0.00	0.00	0.00	0.48	0.00	0.00	0.00	0.00	0.00	0.09	0.44	0.00
Frame; concrete	0.00	0.00	0.00	0.00	0.17	0.44	0.18	0.38	0.40	0.14	0.36	0.33
Frame; brick	0.00	0.00	0.00	0.11	0.07	0.00	0.03	0.00	0.00	0.00	0.44	0.00
Frame; wood	0.00	0.00	0.00	0.39	0.18	0.44	0.18	0.38	0.40	0.14	0.31	0.33
Frame; other material	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Heating; circulating water	0.00	0.00	0.06	0.36	0.28	0.00	0.00	0.10	0.26	0.20	0.34	0.00
Heating; air	0.00	0.00	0.00	0.00	0.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Heating, electricity	0.00	0.00	0.06	0.24	0.00	0.00	0.04	0.10	0.26	0.50	0.33	0.00
Heating; wood	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.50	0.06	0.00
Heating; non	0.00	0.00	0.00	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6
Characterization of the clusters c1–c12 by electricity use in proportion of floor area (kWh/m^2)

Year	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12
2001	48	30	27	66	47	41	74	33	36	104	41	30
2002	51	32	29	75	49	44	83	36	45	114	45	33
2003	45	30	28	67	43	41	76	33	45	100	41	36

4. Discussion

The aim of the study was to develop an advanced data-based method for creating more personalized information about consumer energy use. In this case, personalized information means realistic comparison information about similar customers. The results of this study indicate that customer building characteristics, geographical location, and annual electricity use information can be used to create of usable comparison groups. Comparing the annual electricity use in proportion to floor area of customers apartment and comparison groups with graphs was clear and easily understandable way to present new information.

The presented method was based on applying a self-organizing map (SOM) and *K*-means algorithm in data processing and clustering. Using these computational methods, multi-dimensional data was transformed into the form of comparison groups automatically, without time-consuming manual work. In more detailed level, data preprocessing, determining of optimal cluster amount, clustering, creation of actual comparison information and visualization of customer specific consumption bars was carried out automatically. Customer geographical location was used to identify all the similar customers in the neighborhood and to create comparison group using only these customers. The final location based grouping ensures that the characteristics of the living environment which affects energy consumption are also take into consideration.

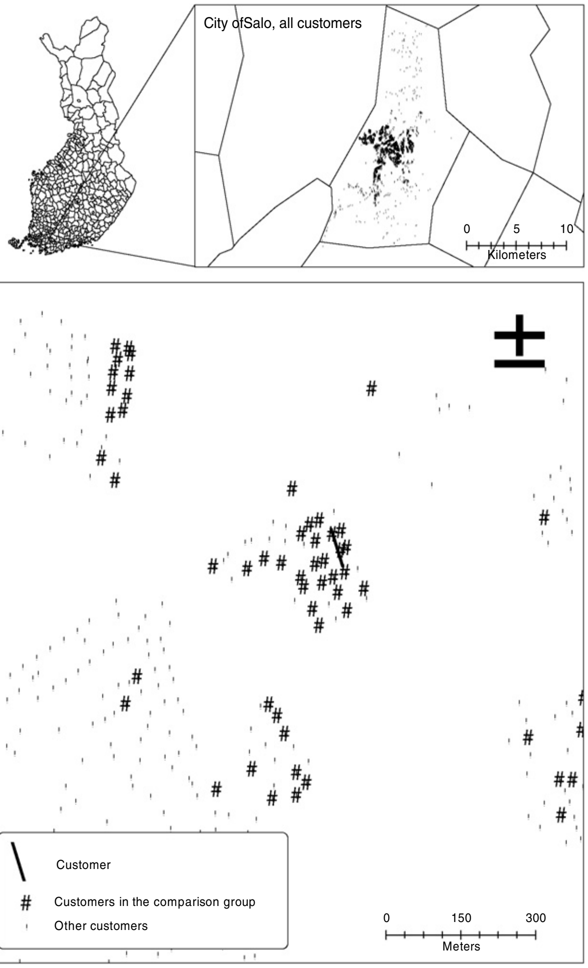

Fig. 2. Illustration of location of chosen customer (star) and its comparison group i.e. similar customers in neighborhood (triangle).

In this case, the data processing method was tested in practise using almost 8000 customers, clustered into 12 groups based on 27 variables describing their building characteristics. The number of groups was determined using the Davies–Bouldin index. For this data set, 12 was the optimal number of clusters. Furthermore, each cluster was identified by creating a short description, based on the main differences in characteristics between groups. For example old detached houses with combined electric and wood heating and no municipal warm water were clustered into one group. Clusterization created was validated together with energy company experts and the results of process, indicated that the suggested customer groups were reasonable and suitable for the creation of more personalized comparison information.

The self-organizing map combined with *K*-means clustering was found to be an efficient method to cluster large amount of such data. The Davies–Bouldin index is one practical method to optimise number of clusters. However, the optimal number of clusters is always compromise and all customers will not belong to the actual cluster center. Consequently, there is always some error in comparisons between comparison groups and customers who differ from the cluster center. We identified customers in each cluster using a fuzzy membership value, which indicated the distance between the customers characteristics and the actual cluster center. Such index was used as a validity measure for comparison information.

This study presents a computationally efficient and cost-effective way to provide customers with more personal and concrete information about their electricity use compared with that of others, which gives a good basis to evaluate their behavior as electricity consumers. In contrast to energy consumption systems based on

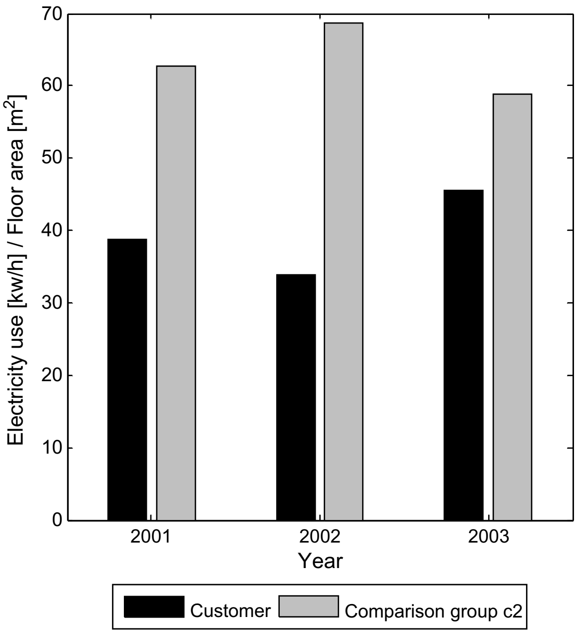

Fig. 3. A comparison of the annual electricity use in proportion of floor area between chosen customer and it's comparison group. Actual comparison group contains only customers of cluster c2 which are within 1 km from the chosen customer.

household monitoring systems [14], the consumption and building data used in this approach is already collected for the another purposes and therefore minor extra costs or investments are caused to the energy companies. Moreover, finding customers consuming more than average customers, living in a similar apartment, gives possibility to target and to create customer specific electricity saving guidance. Intelligent data processing methods were proven suitable for creating such information, which can be delivered to customers using web applications or informative bills. The overall aim of providing this information is to motivate customers to engage in activities leading to energy conservation, e.g. structural changes to the buildings or changes in energy management, and also such information provides a means for evaluating those changes afterwards [2]. In addition, customer clustering gives a comprehensive and deeper understanding of the current clientele which is useful for marketing, customer management and many other business activities [12].

In future investigations, it might be possible to use more accurate electricity use information. In this study, we used annual electricity use which provides only approximate information about customers' behavior. In Finland, hourly measured electricity use data is now available, which should be utilized in next generation of energy services. As a result of this development, the amount of data is increasing and there is need for more efficient computational data management systems. On the other hand, more frequent measuring gives a whole new view of and accuracy in analyzing customers' consumption behavior. Thus, energy companies could provide more detailed information to customers so that they can decrease electricity use and improve energy efficiency.

5. Conclusions

Electricity use in households is one of the main concerns by the means of energy use reduction. Peoples awareness about the problem, consequences and the own consumption behavior is key element to promote energy conservation activities. The purpose of the study was to develop advanced data-based method for creating more personalized information about consumer's energy use. In this case, personalized information means realistic comparison information between similar customers. Presented method is suitable to handle large amounts of customer's characteristics, including location, and electricity use data automatically. This kind of tool provides energy companies possibility to give more detailed information to customers, to target electricity saving guidance and promote actions to improve energy efficiency. Furthermore, presented approach offers also one way to respond to the coming legislation challenges set by European Commission demanding more actions to improve energy efficiency.

Acknowledgements

This research was part of the "Intelligent Environmental Services" research project which was funded by the Finnish Funding Agency Technology and Innovation and Enfo Ltd. We would like to thank Mr. Juhani Helander (Enfo Ltd.), Mrs. Maiju Vuorinen (Fortum Markets Ltd.), Mr. Timo Liiri (Fortum Martkets Ltd.) and Mr. Hannu Santala (Fortum Markets Ltd.) for providing experimental data and important guidance during research project.

References

[1] The European Parliament and the Council of the European Union. Directive 2006/32/EC of the European Parliament and of the Council on energy end-use efficiency and energy services and repealing Council Directive 93/76/EEC. 2006.
[2] Iyer M, Kempton W, Payne C. Comparison groups on bills: automated, personalized energy information. Energy Build 2006;38:988–96.
[3] Isamu M. The effects of information on residential demand for electricity. The Energy J 2004;25:1–17.
[4] Herring H. Energy efficiency – a critical view. Energy 2006;31:10–20.
[5] Wilhite H, Ling R. Measured energy savings from a more informative energy bill. Energy Build 1995;22:145–55.
[6] Chicco C, Napoli R, Piglione F, Postolache P, Scutariu M, Toader C. Load pattern-based classification of electricity customers. IEEE Trans Power Syst 2004;19:1232–9.
[7] Finnish Meteorological Institute. Finland's Climate. Retrieved August 17, 2007. Available from http://www.fmi.fi/weather/climate.html.
[8] Kohonen T. Self-organizing maps. 2nd ed. Berlin: Springer-Verlag; 1997.
[9] Davies D, Bouldin D. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1979;2:224–7.
[10] MacQueen J. Some methods for classification and analysis of multivariate observations. In: Le Cam LM, Neyman J, editors. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1. Berkeley: University of California Press; 1967. p. 281–97.
[11] Zimmermann HJ. Fuzzy set theory and its applications. 2nd ed. Boston, MA: Kluwer Academic Publishers; 1991.
[12] Prince S. Customers: The "new" strategic asset for utilities. The Electricity J 2005;18:85–90.
[13] Europes Energy Portal. Saving energy. Retrieved October 15, 2007. Available from the European Commission's website. http://energy.eu/.
[14] Ueno T, Inada R, Saeki O, Tsuji K. Effectiveness of an energy-consumption information system for residential buildings. Appl Energy 2006;83:868–83.