УДК 004.9

GAS DISAGGREGATION APPROACH BASED ON CLUSTER ANALYSIS

O. Sinkevych, L. Monastyrskii, B. Sokolovskii, Ya. Boyko

Ivan Franko National University of Lviv,
50, Drahomanov St., Lviv, , 79005, Ukraine

The paper presents an approach for understanding the energy consumption of family household, where the heating system uses gas as the main source of energy. Based on the daily gas usage collected during one year, cluster analysis and feature detection have been conducted in order to separate energy time series into meaningful groups (clusters). To do that, the unsupervised learning methods – fuzzy c-means and k-means clustering are applied to finding the relevant clusters of extracted features from the annual gas time series. The validation of clustering results is done via fuzzy partition coefficients and silhouette scores. Here, positive and negative values of gas differences, calculated in sequential time points are used as parameters for the learning algorithm as well as features which describe the behavior of gas consumers. Calculated clusters provide the initial step for splitting the gas time series into different classes relevant to gas consumption purposes. Due to the fact that the behavioral model is totally defined by its features, the latter are grouped into the separate consumer classes using cluster distance metrics. Depending on seasonality, an approach to investigate and improve the annual data clustering results has been suggested. A disaggregation algorithm, which is suggested in the paper, takes into account the calculated clusters to produce consumer descriptions corresponding to different gas usage behaviors.

Keywords: *unsupervised learning, fuzzy c-means, k-means, time series, clustering, feature detection.*

Problem statement. To formalize the problem, let $\mathbf{G} = \left\{\mathbf{G}^{(j)}\right\}$ be the set of gas time series $\mathbf{G}^{(j)} = \{g_i\}^j$, where j corresponds to building number, g_i is the gas consumption converted to instance power value [3] measured at time i. Let $\mathbf{G}_s^{(j)} \subset \mathbf{G}$ be the set of time series $\mathbf{G}^{(j)}$ belongs to one of four specified seasons: 1: winter, 2: spring, 3: summer and 4: autumn, where $s \in \{1,2,3,4\}$, $\bigcup_s \mathbf{G}_s^{(j)} = \mathbf{G}$. Each provided $\mathbf{G}_s^{(j)}$ consists of time series represent actual gas consumption behavior of the chosen building i.

Based on these data, the consumption behavior can be expressed and understood via cluster analysis, i.e., each calculated cluster center c_s^k is defined by the group of features $\gamma_s^{(j,k)}$ extracted from each $\mathbf{G}_s^{(j)}$ and joined by similar pattern, $k = 1,2,\ldots$. Here, all the time series from $\mathbf{G}_s^{(j)}$ are separated daily consumption vectors, i.e., $\mathbf{G}_s^{(j)} = \{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n\}$, where $\mathbf{g}_n$ is daily time series, collected with some pre-defined time step. Also, the group of features $\boldsymbol{\gamma}_S^{(j,k)}$ can be extracted from the whole time series $\mathbf{G}^{(j)}$ and defined as $\boldsymbol{\gamma}^{(j,k)}$, which form clusters with centers c^k. The cluster analysis also can be applied to dissimilarity matrix constructed by the extracted time series features.

The clustering procedure simplifies the development of the gas disaggregation method: the investigation of identified patterns in $\mathbf{G}^{(j)}$ and $\mathbf{G}_S^{(j)}$ produces features, time intervals $\mathbf{t}^{(j)}$, $\mathbf{t}_S^{(j)}$ and threshold values $\varepsilon^{(j,k)}$, $\varepsilon_S^{(j,k)}$ to select gas usage modes and thereby, allows to investigate the consumer behavior.

The problem is defined as follows: 1) firstly, clusters $\boldsymbol{\gamma}_S^{(j,k)}$ and $\boldsymbol{\gamma}^{(j,k)}$ should be calculated via state-of-art unsupervised learning method; 2) based on given clusters $\boldsymbol{\gamma}_S^{(j,k)}$ and $\boldsymbol{\gamma}^{(j,k)}$, consumer models should be identified and used to exclude its consumption part from the aggregated gas data.

Review of recent works. The problem of energy time series clustering in order to investigate and understand its structure is of the significant interest among data scientists. For instance, in [4] authors have investigated the effect of similarity measures in the application of clustering for discovering representatives and building energy consumption patterns. In [5] it has been proposed two distinct temporal patterns of energy performance for commercial and residential buildings, characterized by energy use reductions and increases. Also here has been presented the complex picture of energy use dynamics over time when compared to previous studies.

The paper [6] proposes a time series clustering based methodology to define the segmentation of residential gas consumers. The segments of gas consumers without preliminary feature extraction are obtained through a detailed clustering analysis using smart metering data. Generally, many of state-of-art time series clustering methods are widely studied in [7].

In spite of time series clustering, energy disaggregation problem is still a challenging one. Because gas is usually consumed by the small number of consumers, its disaggregation is simpler than electric data. However, algorithms used for the electrical disaggregation can be applied to gas. Papers [8], [9] cover the methods based on feature extraction and clustering suitable for that purpose.

In paper [10] authors presents a method which uses using dynamic pattern matching of total gas consumption measurements, typical of those provided by a smart meter. The results were obtained using visual time series signature recognition.

Detailed analysis of the available sources revealed a necessity of the development of the new methods as well as the improvement of existing.

Aim of the work. The goal of this work is the development of the approach for the home gas analysis as well as the gas disaggregation algorithm to separate energy time series into consumption classes.

Presentation of the main research. To obtain the real gas metering data for our study, we have used open access database, created under REFIT Smart Home project [11]. This database consists of vast amount of different data types, collected from 20 residential buildings in UK. To simplify future work, we developed the SQLite database and architecture for storing building energy data.

At the first stage, the raw gas time series **G** were undergone data preprocessing procedure: short batches of missing values were linearly interpolated, long batches were removed from the consideration and anomaly detection algorithm [12] was applied to detect outliers.

During the next stage, the actual feature extraction and clustering are being conducted. The smallest time step between consecutive gas data points is 30 minutes length; hence the information about shorter gas consumer modes is unavailable and we can analyze only the major consumptions during the day. Knowing that we propose to use here the following features

$$\gamma_i^j = g_{i+1} - g_i,\ \forall j, \tag{1}$$

as positive and negative magnitudes in consumption changing. To complement these feature set, monotonic magnitudes, their gradients, etc., can also be extracted from the preprocessed time series.

The investigation of the whole annual time series starts with features γ_i^j being extracted to corresponding feature arrays $\boldsymbol{\gamma}_+^j = \left[\gamma_1^j, \gamma_2^j, \ldots, \gamma_n^j\right]^+$ and $\boldsymbol{\gamma}_-^j = \left[\gamma_1^j, \gamma_2^j, \ldots, \gamma_n^j\right]^-$, where each of these arrays contains positive and negative magnitudes; seasonal time series features are placed to $\boldsymbol{\gamma}_+^{j,s}$ and $\boldsymbol{\gamma}_-^{j,s}$ arrays, respectively. Latter are obtained from the daily typical averaged profiles $\mathbf{P}_s^{(j)}$ as the averaged vector sum, calculated for each season as follows

$$\mathbf{P}_s^{(j)} = \frac{1}{n}\sum\nolimits_n \mathbf{w}_n \mathbf{g}_n,\ \mathbf{g}_n \in \mathbf{G}_s^{(j)}, \forall j\,. \tag{2}$$

Here $\mathbf{w}_n$ is weight vector with component $w_n \in [0,1]$ which signifies the effect of more typical series among the all arrays.

The calculation of $\mathbf{w}_n$ is performed using the Euclidian distance between corresponding time series $\mathbf{G}_s^{(j)}$

$$d_{k,m} = \left\|\mathbf{g}_k - \mathbf{g}_m\right\|_{l_2},\ \forall k,m;\ k \neq m \tag{3}$$

$$d_{k,m} \to \mathbf{w}_k \tag{4}$$

Operator $\to$ means the inversely proportional mapping of distance (3) into weight $\mathbf{w}_k$ (4).

Each feature array $\boldsymbol{\gamma}_+^j$, $\boldsymbol{\gamma}_-^j$, $\boldsymbol{\gamma}_+^{j,s}$ and $\boldsymbol{\gamma}_-^{j,s}$ is firstly clustered via fuzzy c-means algorithm (FCM). The reason of choosing this clustering method is that some features may belong to the multiple classes and each cluster item has own membership score.

FCM is defined by minimizing function [13]

$$\Phi_f = \arg\min\nolimits_{\mathbf{c}} \mathbf{M}^l \left\|\boldsymbol{\gamma} - \mathbf{c}\right\|^2 \tag{5}$$

where $c^k, c_s^k \in \mathbf{c}$, $\boldsymbol{\gamma}_+^j, \boldsymbol{\gamma}_-^j, \boldsymbol{\gamma}_+^{j,s}, \boldsymbol{\gamma}_-^{j,s} \subseteq \boldsymbol{\gamma}$, $\mathbf{M}^l$ is membership matrix, l is parameter, which controls the cluster fuzziness.

In comparison to FCM, k-means clustering (KM) is can be applied [14]. To do that, the minimization procedure is defined as follows

$$\Phi_k = \arg\min_{\mathbf{c}} \|\boldsymbol{\gamma} - \boldsymbol{\mu}\|^2, \tag{6}$$

where $\boldsymbol{\mu}$ is the mean vector of $\mathbf{c}$.

FCM and KM algorithms require pre-defined number of clusters, which causes the problem of choosing the optimal partition number and usually highly depends on the subject area. To deal with this issue, many validation methods are recommended: silhouette score (SC), KCE, WB-index [15], etc. For the sake of simplicity, here we restrict ourselves to silhouette score (SC) and fuzzy partition coefficient (FPC) [14].

Using FCM feature sets $\boldsymbol{\gamma}_+^j$, and $\boldsymbol{\gamma}_-^j$, as well as $\boldsymbol{\gamma}_+^{j,s}$ and $\boldsymbol{\gamma}_-^{j,s}$ are parted into clusters, where magnitudes defined in (1) are grouped in the meaning of gas consumers. Generally speaking, each positive magnitude from an arbitrary cluster with centroid $_+c^k$ or $_+c_s^k$ is approximately matched to the negative magnitude from the cluster with $_-c^k$ or $_-c_s^k$. Here positive and negative clusters are approximately equal regarding to their signs; these clusters determine ON/OFF gas consumer states or transition between the corresponding modes. This process leads to the gas consumer model

$$X^{(k)} = \left\langle \gamma_+^*, \arg\min_{\gamma_-^*} \left\| \left(+\mathbf{C}_k^*\right) - \left(-\mathbf{C}^*\right) \right\|, \mathbf{t}, \mathbf{å} \right\rangle \tag{7}$$

where $\gamma_+^* = mean\left(\gamma \in +\mathbf{C}_k^*\right)$ is average magnitude from cluster $_+\mathbf{C}^k$ with centroid $_+c^k$ or $_+c_s^k$, notation $(*)$ identifies membership to cluster with centroid $_+c^k$ or $_+c_s^k$, $\gamma_-^* \in -\mathbf{C}^*$, $-\mathbf{C}^*$ is set of all negative clusters, $\mathbf{t}$ are time intervals when gas was consumed by $X^{(k)}$, $\varepsilon^{(j,k)}, \varepsilon_s^{(j,k)} \in \mathbf{å}$, $\|\cdot\|$ is clustering distance [16].

Model (6) with preceding data cleansing and clustering describes the gas consumer.

To program and test our method, a simple framework has been implemented using Python 3. Firstly, to store the raw gas data provided in CSV format, SQLite database was created and corresponding Python modules were written to process time series. In our experiments we have used the annual energy data sampled during 2014 year, which covered four distinctive seasonal usage patterns.

Each of sample buildings has the unique usage profile; nevertheless all building profiles share common time-dependent and behavioral features.

After data pre-processing, we have created the feature extraction (1) and shape processing (2) units. Cleaned data were split into whole annual and seasonal arrays; next, FCM and KM feature clustering were applied to both datasets. Using the validation metrics FPC and SC respectively, the best cluster numbers were calculated. From the obtained partitions based on the validation results, corresponding positive and negative magnitudes were analyzed in order to being grouped using the optimization procedure in (7).

The identification of parameters $\mathbf{t}$ and $\boldsymbol{\varepsilon}$ can be done using the visual analysis or automatically by sliding window [17].

The presentation of the results of our numerical experiments is started with plotting the seasonal daily profiles. Because of space limit we did not plot the annual consumption data. To simplify the visualization, results obtained for the arbitrary chosen building is given.

Fig. 1 shows typical averaged daily winter profile $\mathbf{P}_1^{(1)}$ calculated using (2). We used 30 minutes time intervals per day; the total day numbers during this season is 90 including weekends. Fig. 2 represents summer daily profile $\mathbf{P}_3^{(1)}$ calculated by (2) for 92 days.

Results of the FCM and KM magnitude clustering done for $\boldsymbol{\gamma}_+^1$ and $\boldsymbol{\gamma}_-^1$ for the whole annual gas data are shown in fig. 3 (2 pre-defined centroids) and 4 (3 pre-defined centroids), respectively. Here δ is general notation for positive and negative magnitudes.

FPC and SC for each clustering are $\mathrm{FPC}^+ = \{2:0.94, 3:0.93\}$, $\mathrm{SC}^+ = \{2:0.82, 3::0.81\}$, $\mathrm{FPC}^- = \{2:0.96, 3:0.94\}$, $\mathrm{SC}^- = \{2:0.85, 3:0.82\}$, where superscripts +,– correspond to $\boldsymbol{\gamma}_+^1$, $\boldsymbol{\gamma}_-^1$, values in curly braces are $\{\text{cluster\_number}:\text{validation score}\}$.

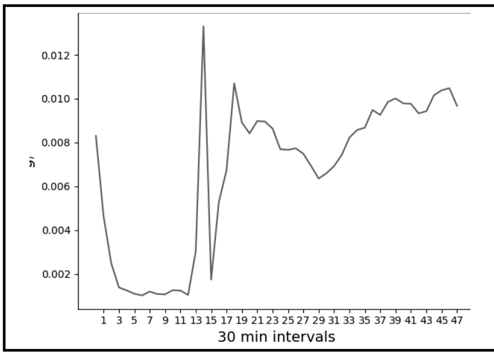

Fig. 1. Winter gas usage profile

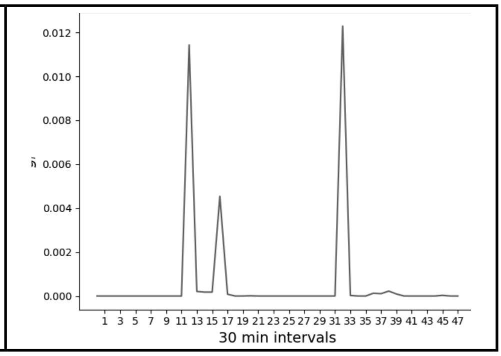

Fig. 2. Summer gas usage profile

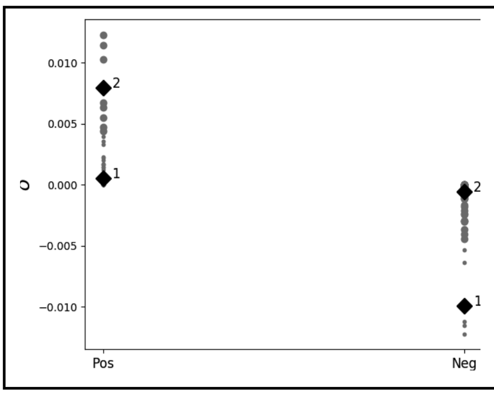

Fig. 3. FCM clustering given by 2 centroids

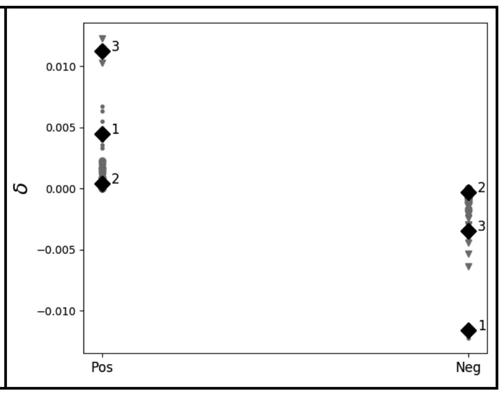

Fig. 4. FCM clustering given by 3 centroids

The clustering of magnitudes from sets $\left\{\gamma_+^{1,1},\gamma_-^{1,1}\right\}$ (fig. 5) and $\left\{\gamma_+^{1,3},\gamma_-^{1,3}\right\}$ (fig. 6) has been done accordingly to the FPC and SC values.

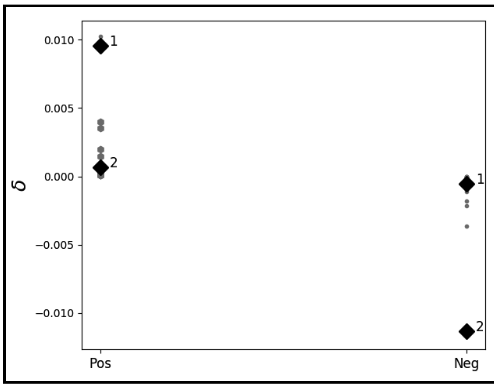

Fig. 5. Winter magnitude clusters

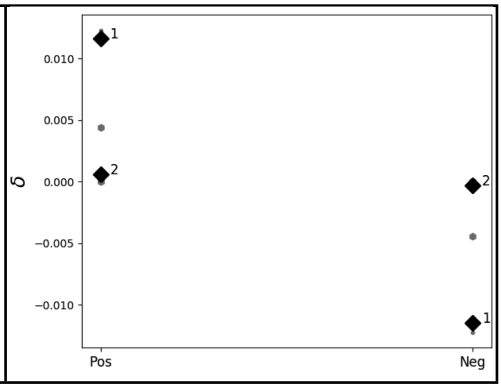

Fig. 6. Summer magnitude clusters

Here fuzzy partition scores are $\mathrm{FPC}\left(\gamma_+^{1,1},\gamma_-^{1,1}\right)=2:\left(0.96;0.98\right)$, $\mathrm{SC}\left(\gamma_+^{1,1},\gamma_-^{1,1}\right)=2:\left(0.84;0.88\right)$, $\mathrm{FPC}\left(\gamma_+^{1,3},\gamma_-^{1,3}\right)=2:\left(0.95;0.96\right)$, $\mathrm{SC}\left(\gamma_+^{1,1},\gamma_-^{1,1}\right)=2:\left(0.84;0.89\right)$. All of these results share the same cluster numbers. Otherwise, for example, clustering of magnitudes from $\left\{\gamma_+^{1,2},\gamma_-^{1,2}\right\}$ which corresponds to spring season (fig. 7) has been resulted in different cluster numbers based on the FPC and SC values $\mathrm{FPC}\left(\gamma_+^{1,2}\right)=3:0.92$, $\mathrm{SC}\left(\gamma_+^{1,2}\right)=3:0.75$, $\mathrm{FPC}\left(\gamma_-^{1,2}\right)=2:0.94$, $\mathrm{SC}\left(\gamma_-^{1,2}\right)=2:0.8$.

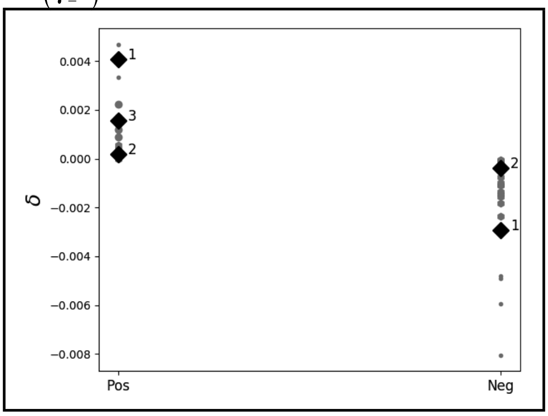

Fig. 7. Spring magnitude clusters

As it is seen in fig. 1, which depicts the daily winter gas profile $\mathbf{P}_1^{(1)}$, night hours are characterized by the absence of heating, whereas early morning hours demonstrate high gas usage peak can be related to some domestic morning activities (cooking, taking a shower, etc.). During the rest of the day we observe the constant heating process with overlapped domestic activities.

Daily summer profile $\mathbf{P}_3^{(1)}$ shows (fig. 2) two prominent peaks in morning and evening hours. This pattern can be described by the presence of high gas usage associated with cooking and water procedures.

Hence, here it arises the problem of splitting the observed patterns into meaningful groups related to gas consumers: water outlets, heater, gas stove, etc. Because of the non-availability of labeled dataset, i.e., when each gas consumer is linked to its unique power consumption, we can use only the unsupervised techniques and visual analysis. To simplify our analysis we consider two general consumers – those using heating $X^{(1)}$ (which values may aggregate short-term usage of other consumers) and cooking/shower consumer $X^{(2)}$ (related to morning and evening periods).

The results of clustering of positive and negative features γ_i^j extracted from the whole year using FCM with two and three centers are shown in fig 3, 4. The FPC and SC validation scores calculated for two and three centroids (which are associated with two and three gas consumers) are close enough; therefore additional information is required to distinguish heaters from other consumers. Nevertheless due to the visual analysis three centroids (consumers) can be more relevant to the reality: $X^{(2)}$ relates to high gas usage (positive cluster 3 and negative cluster 1), $X^{(1)}$ relates to average gas usage (positive cluster 1 and negative cluster 3) and new consumer $X^{(3)}$ related to small short-term user (e.g., hand washing, heating food, etc.).

The clustering of seasonal features $\left\{\gamma_+^{1,1}, \gamma_-^{1,1}\right\}$ and $\left\{\gamma_+^{1,3}, \gamma_-^{1,3}\right\}$ using FCM and KM methods identifies two distinctive consumers $X^{(1)}$ and $X^{(2)}$. Summer profile $\mathbf{P}_3^{(1)}$ almost consists of $X^{(2)}$ pattern; consumer $X^{(1)}$ represents minor gas usage. Winter profile $\mathbf{P}_1^{(1)}$ shows the presence of the same consumer $X^{(2)}$, but consumer $X^{(1)}$ here can be split into two groups regarding to heating and small short-term users $X^{(3)}$.

Unfortunately, spring $\mathbf{P}_2^{(1)}$ and autumn $\mathbf{P}_4^{(1)}$ profiles are more complicated for the investigation. For example, clustering of features $\left\{\gamma_+^{1,2}, \gamma_-^{1,2}\right\}$ results in three consumers, where $X^{(2)}$ has lower magnitude values than during summer and winter seasons; different numbers of corresponding positive and negative clusters have been obtained.

Parameters **t** and **å** can be determined by matching features γ_i^j from calculated clusters with profiles $\mathbf{P}_1^{(1)} \ldots \mathbf{P}_4^{(1)}$, i.e., index *i* corresponds to time interval, values of $\boldsymbol{\varepsilon}$ can be defined as min-max or average values of the set of features γ_i^j.

Conclusions. This paper deals with the analysis of domestic gas consumption based on the disaggregation of the available experimental data. Here we have used the unsupervised learning, which is very important due to the fact that supervised learning requires labeled datasets which are not always available. A short description of the results is given as followed:

- conducted research of features for domestic annual and seasonal gas consumption data using fuzzy c-means method;
- seasonal daily profiles are calculated using suggested weighted average method;
- developed a model for gas consumer based on extracted and clustered features: clusters corresponds consumer determined by distance metric between positive and negative feature clusters;
- revealed relationships between extracted features and consumer models.

The practical meaning of obtained results is that suggested low-complexity approach can be used for the predictive analysis of gas consumption and its optimization. Also, it can be valuable for the developing smart energy control systems using single board computers (Raspberry Pi3) embedded into the smart home architecture.

Prospects for further research are:

- conducting the consistency between cluster (unsupervised) and supervised machine learning approaches;
- expanding feature set from the given gas time series using gradients, shape extraction, etc.

EFERENCES

1. Jouppi N., Borchers A., Boyle R. et al. (2017). In-Datacenter Performance Analysis of a Tensor Processing Unit, *The 44th Annual International Symposium on Computer Architecture*, Toronto, Canada, 24-28 April. 2017: proceedings, IEEE, pp. 112. (in English)
2. Ahmed E, Rehmani M. (2017). Mobile Edge Computing: Opportunities, solutions, and challenges, *Future Generation Computer Systems*, No. 70, pp. 5963. (in English)
3. Kane T., Firth S., Hassan T., Dimitriou V. (2017), Heating behavior in English homes: An assessment of indirect calculation methods, *Energy and Buildings*, No. 148, pp. 89105. (in English)
4. Iglesias F., Kastner W. (2013). Analysis of Similarity Measures in Time Series Clustering for the Discovery of Building Energy Patterns, *Energies*, Vol. 6, No. 2, pp. 579597. DOI: 10.3390/en6020579. (in English)
5. Popadopoulos S., Bonczak B., Kontokosta C. (2018). Pattern recognition in building energy performance over time using energy benchmarking data, *Applied Energy*, Vol. 221, pp. 576586. DOI: 10.1016/j.apenergy.2018.03.079. (in English)
6. Fernandes M. P., Viegas J. L., Vieira S. M., Sousa J. M. C. (2017). Segmentation of Residential Gas Consumers using Clustering Analysis, *Energies*, Vol. 10, No. 12, pp. 20472074. DOI: 10.3390/en10122047. (in English)
7. Aghabozorgi S., Shirkhorshidi A. S., Wah T. J. (2015). Time-series clustering – A decade review, *Information Systems*, Vol. 53, pp. 1638. DOI: 10.1016/j.is.2015.04.007. (in English)
8. Räsänen T., Kolehmainen M. (2009). Feature-Based Clustering for Electricity Use Time Series Data, *International Conference on Adaptive and Natural Computing Algorithms*, Warsaw, Poland, 11-14 April 2009: proceedings. Springer, Berlin, Vol. 5495, pp. 401412. DOI: 10.1007/978-3-642-04921-7_41. (in English)
9. Parson O., Ghosh S., Weal M., Rogers A. (2014). An unsupervised training method for non-intrusive appliance load monitoring, *Artificial Intelligence*, Vol. 217, pp. 119. DOI: 10.1016/j.artint.2014.07.010. (in English)
10. Alzaatreh A., Mahdjoubi L., Gething B., Sierra F. (2018). Disaggregating High-resolution Gas Metering Data Using Pattern Recognition, *Energy & Buildings*, Vol. 176, pp. 1732.DOI: 10.1016/j.enbuild.2018.07.011. (in English)
11. REFIT – REFIT: Smart Homes and Energy Demand Reduction [Electronic resource]. Access mode: http://www.refitsmarthomes.org/.(in English)
12. Baragona R., Battaglia F. (2007). Outliers Detection in Multivariate Time Series by Independent Component Analysis, *Neural Computation*, Vol. 19, No. 7, pp. 19621984. DOI: 10.1162/neco.2007.19.7.1962. (in English)

13. Pal N.R., Bezdek J.C. (1995). On cluster validity for the fuzzy c-means model, *IEEE Transactions on Fuzzy Systems*, Vol. 3, No. 3, pp. 370379. DOI: 10.1109/91.413225. (in English)
14. Kodinariya T. M., Makwana P. R. (2013). Review on determining number of Cluster in K-Means Clustering, *International Journal of Advance Research in Computer Science and Management Studies*, Vol. 1, No. 6, pp. 9095. (in English)
15. Hämäläinen J., Jauhiainen S., Kärkkäinen T. (2017). Comparison of Internal Clustering Validation Indices for Prototype-Based Clustering, *Algorithms*, Vol. 10, No. 3, pp. 105119. DOI: 10.3390/a10030105. (in English)
16. Pandit S., Gupta S. (2011). A comparative study on distance measuring approaches for clustering, *International Journal of Research in Computer Science*, Vol. 2, No. 1, pp. 2931. (in English)
17. Gallardo P., Lazo C., Mardones J. (2015). An initial energy disaggregation technique for smart metering and detection of elevators under the vision of the Internet of Things, *2015 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, Santiago, Chile, 28-30 Oct. 2015: proceedings, IEEE, pp. 917922. DOI: 10.1109/Chilecon.2015.7404682. (in English)

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. In-datacenter performance analysis of a tensor processing unit / [N. Jouppi, A. Borchers, R. Boyle et al.] // The 44th Annual International Symposium on Computer Architecture (ISCA), Toronto, Canada, 24-28 Apr. 2017: proceedings. IEEE 2017. – P. 1–12.
2. Ahmed E, Mobile Edge Computing: Opportunities, solutions, and challenges / E. Ahmed, M. Rehmani // Future Generation Computer Systems. –2017.– №. 70, P. 5963.
3. Heating behavior in English homes: An assessment of indirect calculation methods / [T. Kane, S. Firth, T. Hassan, et al. // Energy and Buildings. – 2017. – №. 148.– P. 89105.
4. Iglesias F. Analysis of Similarity Measures in Time Series Clustering for the Discovery of Building Energy Patterns / F. Iglesias, W Kastner // Energies. – 2013. – Vol. 6, №. 2.– P. 579597. DOI: 10.3390/en6020579.
5. Popadopoulos S. Pattern recognition in building energy performance over time using energy benchmarking data / S. Popadopoulos, B. Bonczak, C. Kontokosta // Applied Energy. – 2018. – Vol. 221. P. 576586. DOI: 10.1016/j.apenergy.2018.03.079.
6. Segmentation of Residential Gas Consumers using Clustering Analysis / [M. P. Fernandes, J. L. Viegas, S. M. Vieira et al.] // Energies. – 2017. – Vol. 10, №. 12. P. 20472074. DOI: 10.3390/en10122047.
7. Aghabozorgi S. Time-series clustering – A decade review / S. Aghabozorgi, A. S. Shirkhorshidi, T. J Wah // Information Systems. – 2015. – Vol. 53. – P. 1638. DOI: 10.1016/j.is.2015.04.007.
8. Räsänen T. Feature-Based Clustering for Electricity Use Time Series Data / T. Räsänen, M. Kolehmainen // International Conference on Adaptive and Natural Computing Algorithms, Warsaw, Poland, 11-14 April 2009: proceedings. – Springer, Berlin, 2009. – Vol. 5495. – P. 401412. DOI: 10.1007/978-3-642-04921-7_41.
9. An unsupervised training method for non-intrusive appliance load monitoring / [O. Parson, S. Ghosh, M. Weal et al.] // Artificial Intelligence. – 2014. – Vol. 217. – P. 119. DOI: 10.1016/j.artint.2014.07.010.

10. Disaggregating High-resolution Gas Metering Data Using Pattern Recognition / [A. Alzaatreh, L. Mahdjoubi, B. Gething et al. // Energy & Buildings. – 2018. – Vol. 176. – P. 1732. DOI: 10.1016/j.enbuild.2018.07.011.
11. REFIT – REFIT: Smart Homes and Energy Demand Reduction [Electronic resource]. – Access mode: http://www.refitsmarthomes.org/.
12. Baragona R. Outliers Detection in Multivariate Time Series by Independent Component Analysis / R. Baragona, F. Battaglia // Neural Computation. – 2007. – Vol. 19, №. 7. – P. 19621984. DOI: 10.1162/neco.2007.19.7.1962.
13. Pal N.R. On cluster validity for the fuzzy c-means model / N.R. Pal, J.C Bezdek // IEEE Transactions on Fuzzy Systems. – 1995 .– Vol. 3, №.3. – P. 370379. DOI: 10.1109/91.413225.
14. Kodinariya T. M. Review on determining number of Cluster in K-Means Clustering / T. M. Kodinariya, P. R. Makwana // International Journal of Advance Research in Computer Science and Management Studies. – 2013. – Vol. 1, №. 6. – P. 9095.
15. Hämäläinen J. Comparison of Internal Clustering Validation Indices for Prototype-Based Clustering / J. Hämäläinen, S. Jauhiainen, T. Kärkkäinen // Algorithms. – 2017. – Vol. 10, № 3. – P. 105119. DOI: 10.3390/a10030105.
16. Pandit S. A comparative study on distance measuring approaches for clustering / S. Pandit, S. Gupta International Journal of Research in Computer Science. –2011. – Vol. 2, № 1. P. 2931.
17. Gallardo P. An initial energy disaggregation technique for smart metering and detection of elevators under the vision of the Internet of Things / P. Gallardo, C. Lazo, J. Mardones // Proceedings of 2015 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Santiago, Chile, 28-30 October 2015: proceedings. – IEEE, 2015. – P. 917922. DOI: 10.1109/Chilecon.2015.7404682.

DOI 10.32403/2411-9210-2019-1-41-23-33

КЛАСТЕРНИЙ МЕТОД ДЕЗАГРЕГАЦІЇ ДАНИХ СПОЖИВАННЯ ГАЗУ

О. Сінькевич, Л. Монастирський, Б. Соколовський, Я. Бойко

Львівський Національний Університет імені Івана Франка,
вул. Драгоманова, 50, Львів, 79005, Україна
oleh.sinkevych@lnu.edu.ua

У статті запропоновано підхід до аналізу енергоспоживання жилого будинку, в якому система опалення використовує газ як основне джерело енергії. На основі щоденного використання газу, зібраного протягом одного року, було проведено кластерний аналіз і виявлення ознак для того, щоб розділити часові ряди споживання газу на значущі групи (кластери). Такі кластери визначають початковий етап для задачі дезагрегації виокремлення з часових рядів споживання газу складові, що відповідають різним типам споживання. Валідація результатів кластеризації здійснюється за допомогою нечітких коефіцієнтів розбиття та силуетних коефіцієнтів. Тут в якості параметрів *для навчання моделей використовуються як додатні, так і від'ємні значення різниці споживання газу, розраховані в послідовних часових точках, а також характеристики, що описують поведінку споживачів газу. Залежно від сезонності запропоновано підхід до дослідження та покращення результатів кластеризації даних на річному числовому інтервалі. Параметри, які визначають модель споживача групуються в окремі класи відповідно до відстаней між кластерами, які обчислюються за заданими метриками. Алгоритм дезагрегації, який пропонується в роботі, враховує розраховані кластери для отримання параметрів споживачів відповідно до параметрів споживання газу.*

Запропонований алгоритм навчання без учителя був реалізований з використанням Python 3 і відкритих бібліотек під ОС Ubuntu 18. Цей алгоритм був застосований до реальних даних споживання газу, тобто, щорічні та сезонні часові ряди пройшли процедуру виявлення ознак та кластерний аналіз з подальшим аналізом та інтерпретацією споживачів газу. Розроблена методика дозволяє використовувати сезонні кластери ознак для побудови точних математичних моделей споживачів у відповідності до часових вікон споживання газу.

Ключові слова: *навчання без учителя, нечітка кластеризація, метод к–середніх, часові ряди, кластеризація, виділяння ознак.*

Стаття надійшла до редакції 12.02.2019
Received 12.02.2019