Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 6 (2020) 155–160
www.elsevier.com/locate/egyr

7th International Conference on Energy and Environment Research, ICEER 2020, 14–18 September, ISEP, Porto, Portugal

Determination of the typical load profile of industry tasks using fuzzy C-Means

Rúben Barreto[a,b], Pedro Faria[a,b], Zita Vale[b,*]

[a] GECAD - Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Rua Dr. António Bernardino de Almeida 431, 4249-015 Porto, Portugal

[b] Polytechnic of Porto, R. Dr. António Bernardino de Almeida 431, 4249-015 Porto, Portugal

Received 23 October 2020; accepted 11 November 2020

Abstract

This paper aims to promote the importance and advantages that the clustering method brings to the world of industry, making it possible to increase production efficiency and to manage the energy resources available better. The purpose of this paper is to group the consumption profiles of a task, in order to be able to determine which is the typical load profile of the task through the Fuzzy C-Means clustering method. The case study of this paper focuses on a task performed by three machines that make up a textile production line that makes several products. Each product, when going through a task performed by a specific machine, has a specific consumption and duration. Thus, by machine, it is determined which is the typical profile of ideal consumption to perform the designated task. In the same way, the general consumption profile of the task is highlighted, that is, the possible consumption profile to be expected when executing this task on one of the three machines.

Peer-review under responsibility of the scientific committee of the 7th International Conference on Energy and Environment Research, ICEER, 2020.

Keywords: Clustering; Data mining; Fuzzy C-Means; Typical load profile; Unsupervised learning

1. Introduction

Nowadays, in the world of industry, there is a high demand for methods that allow us to efficiently optimize the management of energy resources without interfering with production. However, in order to make efficient management, it is essential to acquire specific data, in order to be able to carry out the analysis correctly and, subsequently, the planning of how to act in the face of the problem. Thus, the typical load profile (TLP) concept, becomes crucial to facilitate and improve the management of energy resources.

Typically, the TLP, according to [1] and [2], represents the average daily energy consumption of a group that has consumption profiles with similar characteristics. As a rule, this grouping contains different consumption profiles for a given consumer subject to certain conditions. However, taking into account the context of this paper, instead

* Corresponding author.
E-mail address: zav@isep.ipp.pt (Z. Vale).

https://doi.org/10.1016/j.egyr.2020.11.094

Peer-review under responsibility of the scientific committee of the 7th International Conference on Energy and Environment Research, ICEER, 2020.

of consumer profiles, this grouping contains a diversity of consumption profiles of machines performing various tasks. In order for the TLP of tasks to be as realistic and accurate as possible, it is necessary to implement clustering algorithms.

In this context, clustering is a method that can group a variety of consumption profiles that have similarities at the pattern level. Clustering is an unsupervised method that has enormous relevance in machine learning, data mining, pattern recognition, and more [3,4]. In clustering, there are a variety of methods, with partitioning clustering being one of the most well-known [5]. In general, it allows, through an objective function, to partition an input data set by the desired number of clusters, where each of these clusters contains the input data that share similar characteristics [6]. Partitioning clustering encompasses popular algorithms, such as K-Means and Fuzzy C-Means (FCM), which are classified, respectively, as hard and soft clustering [7,8]. Of these two, FCM stands out for its ability to provide more information about the data treated, where it allows the ambiguity of each of these data to associate with more than one cluster, unlike the hard clustering algorithm. In other words, the FCM algorithm allows reducing the limitations of the K-Means algorithm, making it possible to obtain better results, as can be seen in [9,10].

Thus, within the scope of this paper and energy, the FCM method demonstrated in [11], is used, where it allows the creation of the intended TLPs, which later facilitate the management of energy resources. Thus, this paper is structured as follows. Section 1 depicts the introduction to this paper. Section 2 shows the methodology used through a diagram, and also describes the different phases of the same. Section 3 highlights the case study of this paper, illustrating the different scenarios studied. Section 4 demonstrates the results and, finally, Section 5 presents the conclusion.

2. Approach

This section explains the methodology in detail, supported by Fig. 1. Initially, Data acquisition takes place, which consists of obtaining the own consumptions of the machine, in periods of 5 min and the information of the tasks, regarding the beginning and the end of each task.

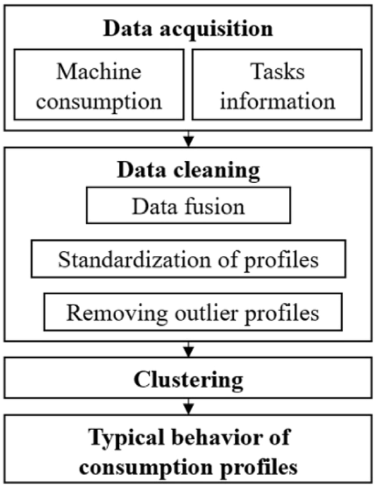

Fig. 1. Proposed Methodology.

The Data cleaning step, in an initial phase, consists of the data fusion, between the information from the tasks with the consumption of the machine in order to determine the consumptions, the durations, and the number of times that each task was performed. Subsequently, the set of consumption profiles existing in a month is brought together by the task. These profiles do not always have the same duration, leading to the need to standardize the periods of these consumption profiles first. Where in this process, after analyzing the different consumption profiles of the respective machine task, an X and Y value is determined, which represents, respectively, the minimum and maximum value of periods that each profile must-have. Those profiles that have a number of periods less than X are eliminated, while in profiles where a number of periods are more significant than Y, in these, the excess is simply removed. Subsequently, of the profiles that respected the previous procedure, those that did not reach the Y limit, the remaining periods up to that value are added, starting from the second half of the values of the respective consumption profile. That is, in order to avoid the initial peak consumption of the task, which corresponds to the warming-up of the machine to perform the task, the second half of the consumption profile is used in order to continue the profile. In order to reach the Y limit, it may be necessary to use the values of the second half of the profile more than once, wherein the end, if the limit is exceeded, the surpluses are removed.

Then, by having all consumption profiles with the same duration, those that do not fit with the others (outlier profiles) are removed. In this way, the Clustering of the consumption profiles of each task becomes possible. As for the third step, it consists of grouping the different consumption profiles by consumption pattern, through the Fuzzy C-Means (FCM) method. In the context of this paper and in general, this method consists of assigning each of the consumption profiles to a respective cluster. In order to obtain the ideal cluster for a consumption profile, for each K cluster, it is determined and assigned a degree of membership for each existing consumption profile, where it corresponds to a value that varies between 0 and 1, and indicates the degree of how much the profile belongs to that cluster. The sum of all the membership values attributed to the K clusters by consumption profile must be equivalent to 1. Thus, each consumption profile is associated with the different K clusters by the membership value.

After forming the K clusters, the respective centroids are determined. The centroid corresponds to the TLP of the machine performing the respective task. In the same way, it is intended to determine the general TLP of the task, that is, a profile in which the task can be represented, seen from the perspective of the production line.

3. Case study

The case study of this paper focuses on a task performed by three machines (A, B, and C) that make up a textile production line that makes several products. Fig. 2 aims to demonstrate the different consumption profiles of the three machines performing the task before the treatment explained in the second step in the proposed methodology, wherein total there are 59 profiles with periods of 5 min. This paper presents four scenarios (A, B, C, and D), where the first three illustrate the different consumption profiles used (after treatment), per machine, with the respective centroids of the generated clusters. In this way, it is possible to determine the best TLP for each machine. The last scenario consists of generating different clusters with all the consumption profiles of the three machines, making it possible to visualize how the general TLP of the task in question would be.

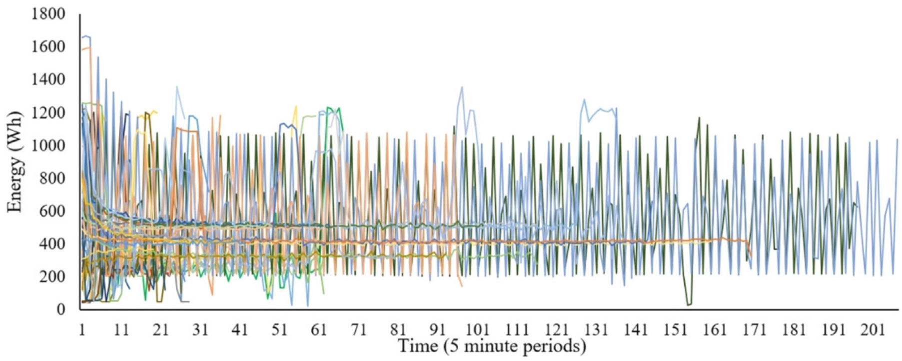

Fig. 2. All consumption profiles before being treated.

4. Results

Table 1 illustrates the resulting partitions for each of the proposed scenarios, highlighting, by scenario, the number of elements in each cluster. In each scenario, the profile of Cluster 1 and 2 is presented, with the exception of scenario D, where there are six clusters. In scenario D, it was considered that the ideal would be to place all consumption profiles with a total number of 40 periods.

Concerning scenario A, illustrated in Fig. 3, it shows the machine that performs the task in question less often. However, of the three machines, it is in this that it is verified that the consumption profiles share a very similar

Table 1. Number of profiles in each Cluster.

Cluster	Scenario A (Machine A)	Scenario B (Machine B)	Scenario C (Machine C)	Scenario D (All Machines)
1	2	9	14	8
2	1	7	3	5
3	–	–	–	5
4	–	–	–	7
5	–	–	–	1
6	–	–	–	10

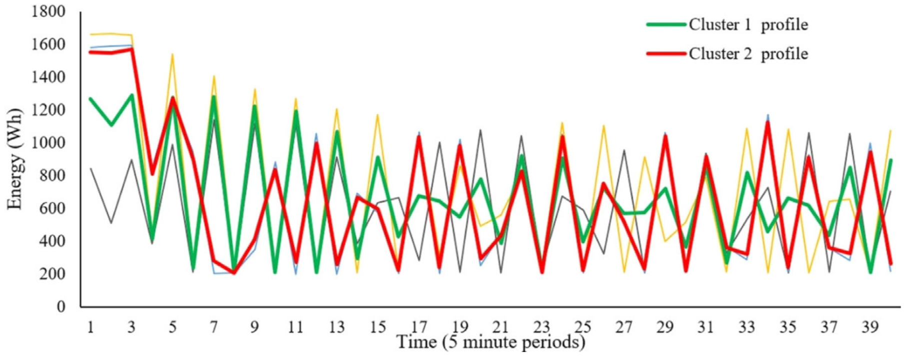

Fig. 3. Profiles of the generated Clusters and the consumption profiles of machine A. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

consumption pattern. Cluster 1 is made up of elements that have very similar consumption patterns, with the profile in blue being the excluded one, since it stands out from the others. So, this cluster is the TLP of this machine.

As for scenario B, not shown in detail due to limitations of space in this paper, it includes 16 consumption profiles of a machine, where in the first moments, it appears that there is, in most of the profiles, a peak that represents the machine to be prepared to perform the assignment. As the task is performed, the behavior of the profiles becomes more uniform, with little variation. The profile of Cluster 1 corresponds to the TLP of this machine.

Regarding scenario C, not shown in detail due to limitations of space in this paper, this represents the largest variety of consumption patterns, with a total of 17 profiles. Where the profile of Cluster 2, depends solely on the three profiles that have higher consumption, making them stand out from the rest. Thus, like most profiles, it is found in the set below, making the TLP correspond to the profile of Cluster 1. Finally, in order to compare the clusters of scenario D with each of the clusters generated in the previous scenarios, Fig. 4 is illustrated, where scenarios A, B, and C are represented, respectively, in graphs (a), (b), and (c). The legend of the genre “Cluster s.c profile” is used, where the elements s and c indicate, respectively, the scenario and the cluster.

In scenario D, 6 clusters (k = 6) are generated to obtain at least 2 clusters that are similar to the 2 clusters generated in scenarios A, B, and C. In this way, it is easier to obtain the general TLP of the task in question. In an initial phase, it also appears that two of the profiles of the clusters in scenario D, D.2 and D.6, are almost identical. From a general perspective, this figure shows that, despite having few consumption profiles, the machine in scenario A is the one that requires more energy to perform the task in question, while the machine in scenario C is the one that requires less. Concerning the machine A scenario, we can see that Cluster D.5 has many similarities with Clusters A.1 and A.2, both in terms of peak occurrences and in the magnitude of energy consumed. Regarding the remaining clusters in scenario D, they have no similarity with the clusters in scenario A. Concerning graph (b), in this case, it is already verified that at least the profile of Cluster D.1 has many similarities about the magnitude of consumption with Cluster B.1. In this way, due to the similarity between them, the profile of Cluster D.1 can be seen as the general TLP of this task. In the case of cluster D.4, this is included among the clusters generated in scenario B, however, in a final phase, it presents some similarities with cluster B.1, in terms of consumption, being slightly smaller. Cluster D.3, has many similarities with Cluster B.2 in relation to energy consumption. This cluster has energy peaks that Cluster B.2 does not. Finally, in scenario C, the profile of Cluster D.3 has some similarities

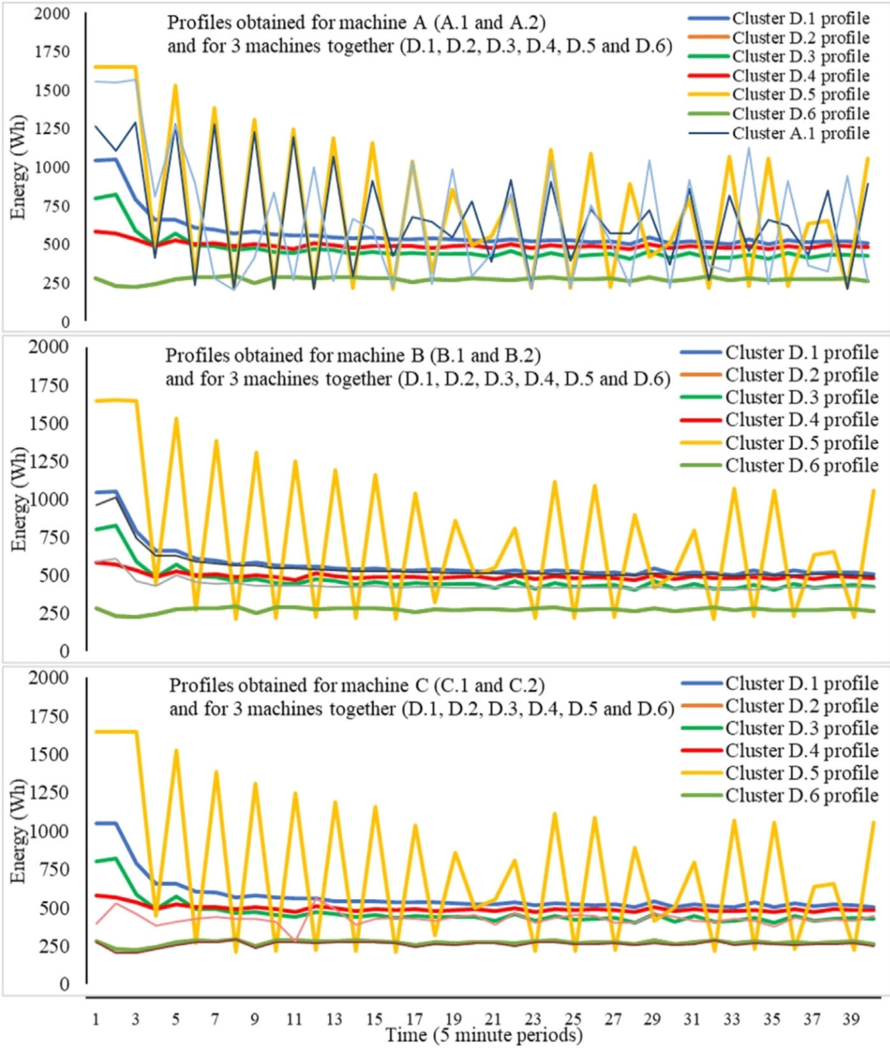

Fig. 4. Comparison of Scenario D with Scenarios A, B, and C.

with Cluster C.2. As for the profile of Clusters D.6, it presents a pattern very similar to the profile of Cluster C.1. However, they have slightly higher consumption than the profile of cluster C.1.

5. Conclusion

This paper details a method that helps make the management of an industry's energy resources more efficient. This method, through FCM, allows creating the TLP of a task executed by a machine, where it stands out when a power peak occurs, the magnitude of that peak as well as the number of times it occurs. In this way, for a given product made on a production line, which goes through different tasks until finalized, it is possible to estimate the necessary energy resources through these TLPs, thus facilitating management planning. Within the scope of this paper, a possible improvement would be to consider different modes of operation of a machine for a given task. So, instead of just taking into account one task's TLP, per machine, we look for several. In this way, it would enable an increase in the accuracy of the estimation of the energy resources necessary for the manufacture of a given product.

CRediT authorship contribution statement

Rúben Barreto: Data curation, Investigation, Formal analysis, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Pedro Faria:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Zita Vale:** Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing - original draft, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work has received funding from Portugal 2020 under SPEAR project (NORTE-01-0247-FEDER-040224) and from FEDER Funds through COMPETE program and from National Funds through (FCT), Portugal under the project UIDB/00760/2020, and CEECIND/02887/2017.

References

[1] Azad SA, Ali ABMS, Wolfs P. Identification of typical load profiles using K-means clustering algorithm. In: Asia-pacific world congr. comput. sci. eng. 2014, http://dx.doi.org/10.1109/APWCCSE.2014.7053855.

[2] Natale N, et al. Assessment of typical residential customers load profiles by using clustering techniques. In: 2017 AEIT int. annu. conf. infrastructures energy ICT oppor. foster. innov., vol. 2017-Janua. 2017, p. 1–6. http://dx.doi.org/10.23919/AEIT.2017.8240518.

[3] Jain LC, Behera HS, Mandal JK, Mohapatra DP. Computational intelligence in data mining – vol. 2: Proceedings of the international conference on CIDM. Smart innov. syst. technol, vol. 32, 2015, http://dx.doi.org/10.1007/978-81-322-2208-8.

[4] Figueiredo V, Rodrigues F, Vale Z, Gouveia JB. An electric energy consumer characterization framework based on data mining techniques. IEEE Trans Power Syst 2005;20(2):596–602. http://dx.doi.org/10.1109/TPWRS.2005.846234.

[5] Fuchs C, Spolaor S, Nobile MS, Kaymak U. A swarm intelligence approach to avoid local optima in fuzzy C-means clustering. In: IEEE int. conf. fuzzy syst., vol. 2019-June. 2019, p. 1–6. http://dx.doi.org/10.1109/FUZZ-IEEE.2019.8858940.

[6] Pimentel BA, De Souza RMCR. A multivariate fuzzy c-means method. Appl Soft Comput J 2013;13(4):1592–607. http://dx.doi.org/10.1016/j.asoc.2012.12.024.

[7] Silva C, Faria P, Vale Z. Assessment of distributed generation units remuneration using different clustering methods for aggregation. In: 2018 IEEE international conference on communications, control, and computing technologies for smart grids, vol. 217. 2018, p. 1–6. http://dx.doi.org/10.1109/SmartGridComm.2018.8587438.

[8] Silva C, Faria P, Vale Z. Discussing different clustering methods for the aggregation of demand response and distributed generation. In: 2018 IEEE symposium series on computational intelligence. 2018, p. 1645–50. http://dx.doi.org/10.1109/SSCI.2018.8628781.

[9] Lu Y, Zhang T, Zeng Z. Adaptive weighted fuzzy clustering algorithm for load profiling of smart grid customers. In: 2016 IEEE/CIC int. conf. commun. china. 2016, http://dx.doi.org/10.1109/ICCChina.2016.7636874.

[10] Lu Y, Ma T, Yin C, Xie X, Tian W, Zhong SM. Implementation of the fuzzy C-means clustering algorithm in meteorological data. Int J Database Theory Appl 2013;6(6):1–18. http://dx.doi.org/10.14257/ijdta.2013.6.6.01.

[11] Bezdek JC. Pattern recognition with fuzzy objective function algorithms. 1st ed.. Boston, MA: Springer US; 1981.