Comparing Electricity Consumer Categories Based on Load Pattern Clustering with Their Natural Types

Conference Paper · August 2017

DOI: 10.1007/978-3-319-65482-9_51

Zigui Jiang
Sun Yat-Sen University

Rongheng Lin
Beijing University of Posts and Telecommunications

Zhihan Liu
Beijing University of Posts and Telecommunications

Comparing Electricity Consumer Categories Based on Load Pattern Clustering with Their Natural Types

Zigui Jiang[1(✉)], Rongheng Lin[1], Fangchun Yang[1], Zhihan Liu[1], and Qiqi Zhang[2]

[1] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China
{ziguijiang,rhlin,fcyang,zhihan}@bupt.edu.cn
[2] State Grid Shanghai Municipal Electric Power Company, Shanghai, China

Abstract. As one aspect of smart city, smart gird has similar situation such as big data issue. Data analysis of daily load data generated by smart meters can benefit both electricity suppliers and end consumers. Electricity consumer categorization based on load pattern clustering is one of research subjects. This paper aims to achieve a better understanding of electricity consumer categorization by detecting the relationships among consumer categories and their natural types. A two-stage clustering based on multi-level 1D discrete wavelet transform and K-means algorithm is applied to perform daily load curve clustering and load pattern clustering. Additionally, to obtain distinct consumer categories, method of category identification based on association rule mining and characteristic similarity is also proposed in this paper. Experiment is conducted on data set of 24-value daily load data with labels of consumer types. Based on the comparison of experimental results, both relationships and differences exist among consumer categories and consumer types but consumer types cannot determine consumer categories.

Keywords: Smart Grid · Consumer category · Consumer type · Load pattern · Clustering

1 Introduction

As information and communication technologies (ICTs) and Internet of thongs (IoT) technologies are widely applied with the development of smart cities, big data are increasingly produced in every part of the cities. Such big data Analysis is beneficial for understanding, monitoring, regulating and planing the cities [1]. This situation also exits in smart grid which is one aspect of smart city. For instance, smart meters in consumer side record power consumption of electricity consumers in a high frequency. Based on the analysis of these detailed measurements, electricity suppliers can enable their operations such as energy control, demand side management and flexible pricing schemes [2,3]. Diverse power consumption behaviors refer to distinctive characteristics of electricity consumers,

S. Ibrahim et al. (Eds.): ICA3PP 2017, LNCS 10393, pp. 658–667, 2017.
DOI: 10.1007/978-3-319-65482-9_51

which drive consumer categorization based on their load pattern similarity. End consumers can choose suitable payment programs offered by electricity suppliers specifically for their categories. Furthermore, electricity consumers have their own characteristics such as natural type and location. Do these characteristics have relationships with power consumption behaviors? Thus, this paper focuses on sophisticated data analysis of daily load data generated by smart meters to detect the relationships among consumer categories and their natural types.

Every electricity consumer belongs to a type which refers to the sort that the consumer naturally is. For example, school, restaurant, hotel or normal resident. On the other hand, consumers also can be categorized into different groups based on the similarity of their electricity power consumption behaviors. Such groups are called consumer categories. Without further analysis, it cannot be ensure that electricity consumers in the same types have similar consumption behaviors or consumers in different types belong to different categories. Since consumer types are usually apparent, achieving consumer categorization is the primary task in this paper.

Following an analysis of the relevant literature dealing with consumer categorization, it is noted that there are two main research aspects, which are categorization algorithm and multi-stage categorization framework. Clustering algorithms such as self-organizing maps (SOM) and K-means are widely adopted in consumer categorization [3–7]. Moreover, consumer categorization framework generally contains a two-stage clustering [2,8–11]. The first stage is to extract load patterns for each individual consumer by daily load curves clustering. Then the next stage conducts a second clustering based on the selected representative load patterns. Panapakidis et al. [9] used K-means algorithm in the two-stage clustering. In their approach, load pattern with the largest cluster, the maximum daily energy or the peak load can be selected as the representative load pattern of each consumer for the second clustering. This simplifies consumer categorization so that the final clustering result does generate the consumer categories directly. However, those unselected load patterns lead to information loss. Mets et al. [2] adopted fast wavelet transformation and g-means algorithm for the two-stage clustering. They also mentioned the limitation of load pattern selection. Therefore, all load patterns rather than one representative are employed for the second clustering. However, this also leads to another problem which is indistinct consumer categories. The final clustering result may show that a consumer belongs to several categories.

According to this review of the literature, this paper enhance the detailed analysis of clustering results to identify the distinct consumer categories. An approach including a two-stage clustering and category identification is proposed. Power consumption characteristics of different consumer types are also be identified based on the same two-stage clustering. Additionally, this paper compares the characteristics of consumer categories and consumer types, and finds out relationships and differences among them. According to the comparison and findings, it can achieve a better understanding of power consumption characteristics which are helpful for improving consumer categorization and conducting new consumer classification.

The rest of this paper is structured as follows. Section 2 explains the proposed approach and describes the algorithms adopted in the approach. Section 3 presents the experimental results with comparison and evaluation. Finally, the paper is concluded in Sect. 4.

2 Approach

In order to compare consumer category with consumer type, it is essential to obtain the representative load patterns of every consumer category and consumer type, which present its typical electricity power consumption characteristics. As mentioned in Sect. 1, load pattern extraction and consumer categorization are based on daily load curve clustering and load pattern clustering, respectively. Thus, this section explains the main approach to achieve the two-stage clustering and consumer category identification.

The procedure of the proposed approach is described in Fig. 1. First, daily load curve clustering is applied to every individual consumer to extract load patterns. Second, another clustering is applied to the overall load patterns of all consumers to achieve fuzzy categories. Finally, consumer categories are identified based on the fuzzy categories. Characteristics are the representative load patterns of consumer categories.

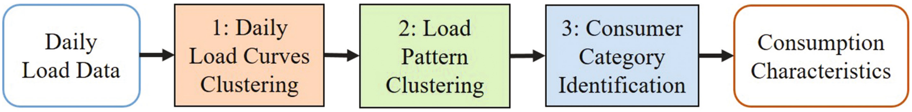


Fig. 1. The procedure of electricity consumer categorization based on a two-stage clustering and category identification.

2.1 Daily Load Curve Clustering

The load data of one day can be drawn as a curve so-called daily load curve. A daily load curve presents the power consumption of an electricity consumer in one day. It is supposed that every consumer has her/his typical consumption behaviors which can be presented by several load patterns. In general, load patterns are extracted by clustering of daily load curves in a certain period. The two-stage clustering in this paper applies a fused load curve clustering algorithm based on multi-level 1D discrete wavelet transform (DWT) and K-means. This clustering algorithm is specially designed for load curve clustering and proposed in a previous paper which is under review now. In the previous paper, it is proved that this algorithm improves curve clustering performance with less information loss of dimensionality reduction.

The curve clustering algorithm has two main steps. The first step is to reduce the dimensions of daily load curves by multi-level 1D DWT. Two types of output, approximation signals and detail signals, are produced in the first step. In the second step, two types of signals are processed separately. Taking into consideration the different properties of two signals, approximation signals are normalized with z-score to ignore the distance difference while detail signals are used directly. Then, normalized approximation signals and original detail signals are clustered separately by K-means to produce two groups of clusters, which are fused into one group of clusters finally.

Since Haar wavelet is simple and can compress a discrete signal into half, it is adopted in the multi-level 1D DWT. The mother wavelet function of Haar wavelet is described as follows:

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2, \\ -1 & 1/2 \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Furthermore, the optimal Ks of K-means for two signals are determined by the Simplified Silhouette Width Criterion (SSWC) [12].

2.2 Load Pattern Clustering

After the daily load curve clustering, each consumer has several load patterns that can present her/his typical power consumption behaviors. All load patterns of consumers are employed in the load pattern clustering in order to keep as much information as possible. Since load patterns are also load curves, the clustering algorithm adopted in this stage is the same as the one adopted in daily load curve clustering.

As mentioned in the former section, the result of load pattern clustering is indistinct in terms of consumer categories. This means that several load patterns of a consumer belong to various clusters that refer to different consumer categories. Thus, the consumer categories obtained in this stage are called fuzzy consumer categories, which are required to be further analysis.

2.3 Consumer Category Identification

Based on the fuzzy consumer categories from the two-stage clustering, consumer category identification aims to obtain distinct consumer categories with diverse consumption characteristics. Generally, consumers who have several load patterns belonging to the same groups of fuzzy consumer categories are in the same consumer categories. Therefore, it is a problem of finding association rules in fuzzy consumer categories. Regarding each load pattern as an item, consumer category identification is to find all frequent itemsets using minimum support [13,14]. Apriori algorithm is adopted in this stage because it is the key algorithm for an extraction of association rules [15].

Let a couple (A, B) be an association rule, where $A \neq \emptyset$, $B \neq \emptyset$ and $A \cap B = \emptyset$, then this rule is noted as: $A \rightarrow B$. The support of an association rule $Sup(A \rightarrow B)$ is defined as the support of the itemset $A \cup B$, which refers to the percentage of transactions containing both A and B. The definition fellows the equation below:

$$Sup(A \rightarrow B) = Sup(A \cup B) = \frac{|t(A \cup B)|}{t(A)}. \tag{2}$$

For Apriori, it is to find the items with a $support \geq minsup$. In this paper, $minsup = 0$ is set in order to find out all frequent itemsets of load patterns. Moreover, all frequent itemsets are then combined based on their similarity.

For electricity consumers with n-dimensional daily load curves, let $X = \{x_1, x_2, \cdots, x_m\}$ and $Y = \{y_1, y_2, \cdots, y_r\}$ be two frequent itemsets of load patterns, where $x_i = \langle x_{i1}, x_{i2}, \cdots, x_{in} \rangle$, $1 \leq i \leq m$ and $y_j = \langle y_{j1}, y_{j2}, \cdots, y_{jn} \rangle$, $1 \leq j \leq r$. The most common way to calculate similarity is based on distance measure. As Euclidean distance is the default distance measure of K-means and it is also adopted in the former clustering algorithm in the proposed approach, the similarity calculation in the stage also adopts it. However, in order to ignore the distance difference and achieve curve similarity of shape variation, z-score normalization is applied to each $x_i \in X$ and $y_j \in Y$ before the calculation. Let x_i' and y_j' be the normalized x_i and y_j, respectively. Then, the similarity of one load pattern x_i and the frequent itemset Y is calculated as follows:

$$S_{x_i,Y} = \min_{y_j \in C} \{\mathrm{dist}(x_i', y_j')\}, \tag{3}$$

where $\mathrm{dist}(x_i, y_j)$ is the distance between x_i' and y_j'. The similarity of X and Y is calculated as the average of S_{x_i} over $i = 1, 2, \cdots, m$:

$$Sim(X, Y) = \frac{1}{m} \sum_{i=1}^{m} S_{x_i}. \tag{4}$$

A parameter *mindis* is set to determine whether combine two frequent itemsets or not. X and Y are combined into one set when $Sim(X, Y) \leq mindis$. After the combination, the remaining frequent itemsets refer to the characteristics of consumer categories.

3 Results and Evaluation

This section presents the experimental results of the proposed approach to consumer categories and consumer types with detailed comparison and discussion.

3.1 Data Set

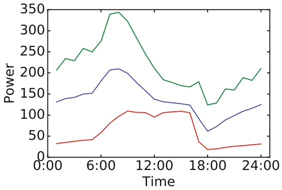
The data set used in the experiment contains 24-value daily load data of 657 electricity consumers in a one-year period. It records the electricity power consumption at every 1 h so that the daily load data have 24 values from 1:00 to 24:00. Moreover, these consumers are labeled with nine consumer types which are full service restaurant, large hotel, small hotel, hospital, outpatient, midrise apartment, primary school, super market and warehouse.

3.2 Results

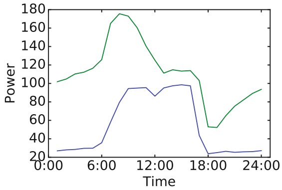
Experimental results include load patterns of individual consumer and the representative load patterns, also called consumption characteristics, of both consumer types and consumer categories. The characteristics of consumer categories are obtained by the proposed approach while the characteristics of consumer types are obtained by a similar two-stage clustering using the same clustering algorithm.

Load Patterns. In the first stage, daily load curve clustering is performed for every individual consumer to obtain their load patterns. Examples are shown in Fig. 2 to present the similarity and difference among consumers in the same type. The consumers are randomly selected from warehouses. In Fig. 2, each warehouse has at least one or several similar load patterns in terms of shape variation. Comparing all load patterns of 657 consumers, it is found that consumers in the same types do have similar load patterns while certain differences also exit among them.

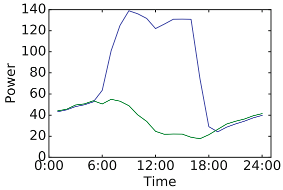
Consumer Type. The consumption characteristics of consumer types are also obtained by a two-stage clustering. The first clustering is the same as the one in daily load curve clustering. After load pattern extraction for each consumer, the load patterns are grouped based on the labeled types of consumers. Then the second clustering is performed separately for different groups. Simply, the clustering results shown in Fig. 3 are consumption characteristics of different consumer types. Although the characteristics of the same types seems unique, part of load patterns in different characteristics also have similarity which may lead to different results of consumer categories.



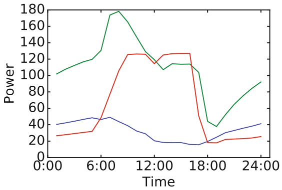
(a) Warehouse 1



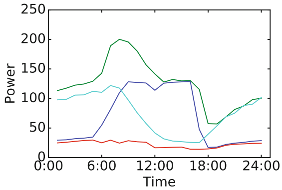
(b) Warehouse 2



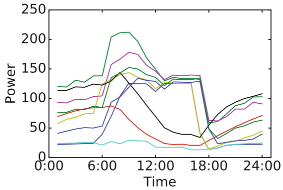
(c) Warehouse 3



(d) Warehouse 4

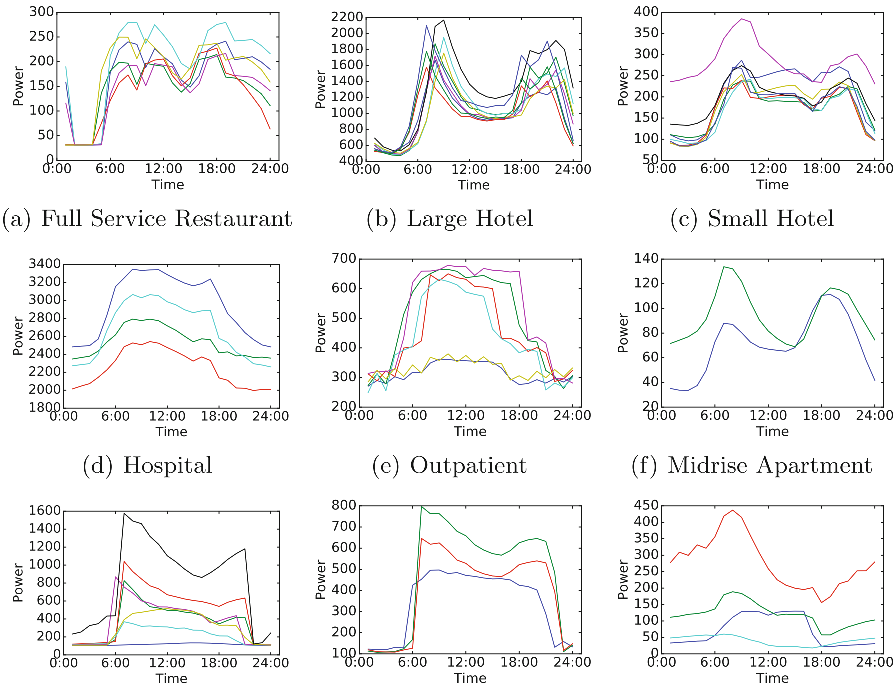


(e) Warehouse 5



(f) Warehouse 6

Fig. 2. Load patterns of six consumers who are randomly selected from warehouses.



(a) Full Service Restaurant (b) Large Hotel (c) Small Hotel

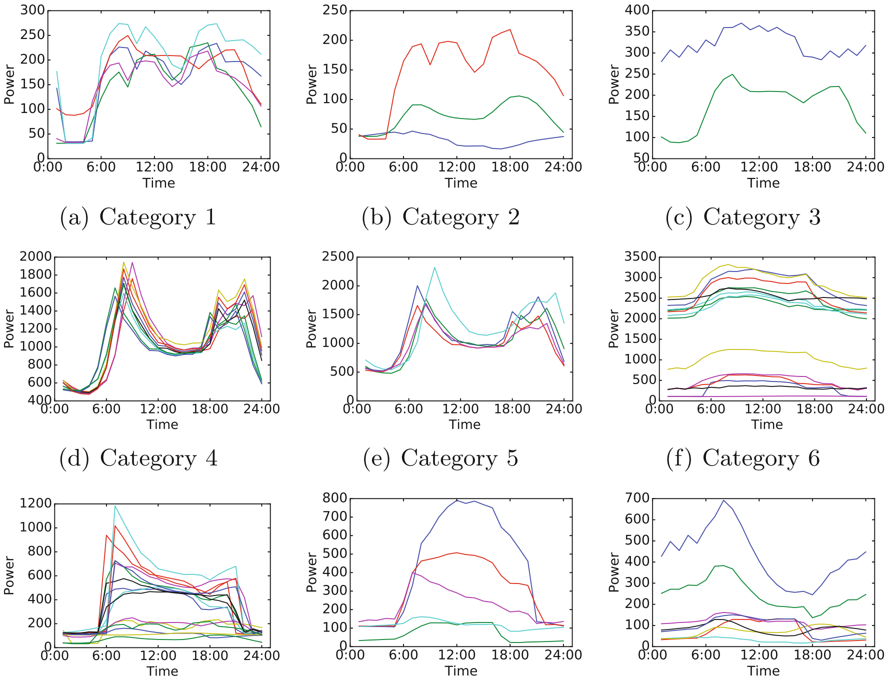
(d) Hospital (e) Outpatient (f) Midrise Apartment

(g) Primary School (h) Super Market (i) Warehouse

Fig. 3. Electricity power consumption characteristics of nine consumer types, which are extracted based on the load pattern clustering of consumer types.

Consumer Category. Based on the proposed approach, a group of consumer categories are identified from the daily load data of 657 electricity consumers. Figure 4 shows the consumption characteristics of nine identified consumer categories when $mindis = 2$. Each curve in Fig. 4 denotes a representative load pattern of a consumer category. It is noticed that most of characteristics show unique load patterns except for Category 4 and Category 5. Load patterns with similar shape variation but diverse power degrees are grouped in the same categories such as Category 6. This result basically meets the requirement of consumer categorization. Furthermore, these consumption characteristics can be regarded as labels and training samples to classify new electricity consumers. In that case, unsupervised clustering problem becomes supervised classification which is easier to be conducted and evaluated.

Comparing the characteristics shown in Figs. 3 and 4, some consumer types are grouped into same categories due to the similar shape variation of their load patterns. It is concluded that the natural types of electricity consumers cannot determine the categorization that based on load pattern similarity. It is highly



(a) Category 1 (b) Category 2 (c) Category 3

(d) Category 4 (e) Category 5 (f) Category 6

(g) Category 7 (h) Category 8 (i) Category 9

Fig. 4. Nine identified consumer categories when $mindis = 2$. Each subfigure denotes the electricity power consumption characteristic of one consumer category.

possible that grouping electricity consumers based on their natural types leads to excessively meticulous division result. On the contrary, consumer categorization can preferably describe their shared power consumption characteristics.

3.3 Parameter Estimation

Due to the normalization before similarity calculation, the value of parameter $mindis$ is small. To estimate parameter $mindis$, the experiment conducts category identification with the value of parameter $mindis$ from 0 to 5 with a step size of 0.1. Figure 5 indicates the curve of numbers of consumer categories based on parameter $mindis$. According to Fig. 5, it can be noted that the number of consumer categories decreases with the raise of parameter $mindis$ value. The decrease of the curve tends to change smoothly after around $mindis = 3$.

In the former experiment, $mindis = 2$ is set for category identification. The selection of parameter $mindis$ is based on the observation of identified categories and their characteristics. Similar load patterns should be grouped into one consumer category. On the other hand, the number of consumer categories should

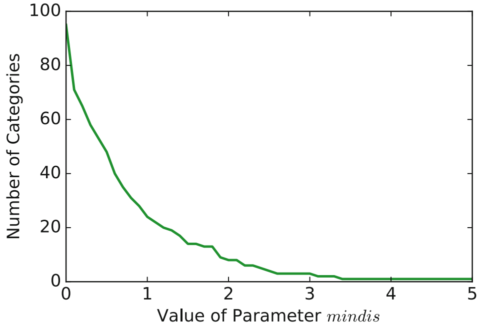


Fig. 5. Number of consumer categories vs. parameter *mindis* from 0 to 5 with a step size of 0.1.

be appropriate. Based on these two ideas, $mindis = 2$ is selected so that nine categories are identified. Actually, it would be better that parameter *mindis* is set depending on the accuracy if the characteristics are used for new consumer classification. However, classification is not included in this paper due to the limitation of time and paper length.

4 Conclusion

This paper presents a two-stage clustering that contains daily load curve clustering and load pattern clustering, and a proposed method of distinct consumer identification based on association rule mining and characteristic similarity. The approach is implemented on the data set of 24-value daily load data of 657 electricity consumers with nine labeled consumer types. Comparing the power consumption characteristics of consumer categories with those of consumer types, the natural types of electricity consumers cannot fully determine the consumer categorization that based on load pattern similarity. Additionally, consumers can be labelled once the categorization is achieved. Thus, regarding the consumption characteristics of consumer categories as training sample, it is simple to perform new consumer classification.

Due to the limitation of time and paper length, this paper does not present a sophisticated work. As a result, the future work contains improvement of the proposed approach, experiments on various data sets and new consumer classification.

Acknowledgments. This work is supported by the National High Technology Research and Development Program (863 Program) of China (2015AA050203) and Beijing Natural Science Foundation (4174099).

References

1. Kitchin, R.: The real-time city? big data and smart urbanism. GeoJournal **79**(1), 1–14 (2014)
2. Mets, K., Depuydt, F., Develder, C.: Two-stage load pattern clustering using fast wavelet transformation. IEEE Trans. Smart Grid **7**(5), 2250–2259 (2016)
3. Kwac, J., Flora, J., Rajagopal, R.: Household energy consumption segmentation using hourly data. IEEE Trans. Smart Grid **5**(1), 420–430 (2014)
4. Figueiredo, V., Rodrigues, F., Vale, Z., Gouveia, J.B.: An electric energy consumer characterization framework based on data mining techniques. IEEE Trans. Power Syst. **20**(2), 596–602 (2005)
5. Albert, A., Rajagopal, R.: Smart meter driven segmentation: what your consumption says about you. IEEE Trans. Power Syst. **28**(4), 4019–4030 (2013)
6. Alahakoon, D., Yu, X.: Smart electricity meter data intelligence for future energy systems: a survey. IEEE Trans. Ind. Inform. **12**(1), 425–436 (2016)
7. Haben, S., Singleton, C., Grindrod, P.: Analysis and clustering of residential customers energy behavioral demand using smart meter data. IEEE Trans. Smart Grid **7**(1), 136–144 (2016)
8. Chicco, G.: Overview and performance assessment of the clustering methods for electrical load pattern grouping. Energy **42**(1), 68–80 (2012)
9. Panapakidis, I.P., Alexiadis, M.C., Papagiannis, G.K.: Electricity customer characterization based on different representative load curves. In: 2012 9th International Conference on the European Energy Market, pp. 1–8. IEEE (2012)
10. Khumchoo, K.Y., Kongprawechnon, W.: Cluster analysis for primary feeder identification using metering data. In: 2015 6th International Conference of Information and Communication Technology for Embedded Systems, pp. 1–6. IEEE (2015)
11. Wang, Y., Chen, Q., Kang, C., Zhang, M., Wang, K., Zhao, Y.: Load profiling and its application to demand response: a review. Tsinghua Sci. Technol. **20**(2), 117–129 (2015)
12. Vendramin, L., Campello, R.J., Hruschka, E.R.: On the comparison of relative clustering validity criteria. In: SDM, pp. 73–744. SIAM (2009)
13. Rohit, S.: Association rule mining algorithms: survey. Int. Res. J. Eng. Technol. **3**(10), 500–505 (2016)
14. Rathod, R.R., Garg, R.D.: Regional electricity consumption analysis for consumers using data mining techniques and consumer meter reading data. Int. J. Electr. Power Energy Syst. **78**, 368–374 (2016)
15. Addi, A.M., Tarik, A., Fatima, G.: Comparative survey of association rule mining algorithms based on multiple-criteria decision analysis approach. In: 2015 3rd International Conference on Control, Engineering & Information Technology, pp. 1–6. IEEE (2015)