big data and
cognitive computing

MDPI

Article

Mining Temporal Patterns to Discover Inter-Appliance Associations Using Smart Meter Data

Sarah Osama *, Marco Alfonse and Abdel-Badeeh M. Salem

Faculty of Computer and Information Sciences, Ain Shams University, Cairo 11566, Egypt;
marco@fcis.asu.edu.eg (M.A.); absalem@cis.asu.edu.eg (A.-B.M.S.)
* Correspondence: sarah.osama@cis.asu.edu.eg

Received: 6 March 2019; Accepted: 27 March 2019; Published: 29 March 2019

check for updates

Abstract: With the emergence of the smart grid environment, smart meters are considered one of the main key enablers for developing energy management solutions in residential home premises. Power consumption in the residential sector is affected by the behavior of home residents through using their home appliances. Respecting such behavior and preferences is essential for developing demand response programs. The main contribution of this paper is to discover the association between appliances' usage through mining temporal association rules in addition to applying the temporal clustering technique for grouping appliances with similar usage at a particular time. The proposed method is applied on a time-series dataset, which is the United Kingdom Domestic Appliance-Level Electricity (UK-DALE), and the results that are achieved discovered appliance–appliance associations that have similar usage patterns with respect to the 24 h of the day.

Keywords: hierarchical clustering; temporal association rules; smart meter; appliance–appliance association; energy management

1. Introduction

Revolutionizing the smart grid environment has been the main focus of many governments as a consequence of the expeditious rise in energy demand [1]. A smart grid is a network of power plants, utilities, substations, and smart meters for transferring electricity in a bidirectional way [2]. Smart meters are considered to be the "cornerstone of the smart grid", as they send consumption data frequently based on a time-interval basis [3]. The widespread deployment of smart meters has granted the availability of smart meter data. As a consequence, a massive amount of data is being generated. These precious data are analyzed to study and understand home residents' preferences. However, mining smart meter data is a challenging task. First, the continuous massive amount of data being generated requires mining data progressively without mining the whole database whenever new data is transmitted. Second, extracted findings are changing continuously with time, so it is important to maintain previously discovered findings in addition to discovering new findings.

Demand response (DR) programs are proposed by utilities for ensuring an efficient use of energy. DR is the change of home residents' behavior and power usage in response to the change of electricity prices [4]. The key enabler for promoting these programs is to gain home residents' trust and respect their preferences when using their home appliances. Extracting such preferences can be achieved by analyzing smart meter data to find out the patterns by which the appliances are being used.

Home residents' usage behavior follows some regular routine patterns. Appliance usage patterns can be described by appliance–time associations and appliance–appliance associations. Appliance–time association is the correlation of using an appliance at a particular time. For example, a coffee machine is usually used at 8:00, so this appliance has higher priority to be used at this time more than any other appliance. Appliance–appliance association is the correlation for using two or more appliances together. For example, home theater usage is always associated with television (TV) usage, so this preference reveals that the two appliances should be working together, since it is useless to have the home theater on without using the TV. Thereby, DR programs should be designed based on user preferences without lowering their comfort level to encourage them to use energy efficiently. Raising home residents' awareness and providing them with the needed knowledge regarding their consumption will guide them to have a better usage behavior and use energy efficiently.

In the proposed work, appliance–appliance associations are represented by using association rules and hierarchical clustering. Association rules are extracted by employing the Utility-oriented Temporal Association Rules Mining (UTARM) algorithm [5]. Agglomerative hierarchical clustering is used to group appliances with similar usage behavior together. There are two strategies of hierarchical clustering: divisive and agglomerative strategy. Divisive clustering uses a top–down approach, where all objects are initially in one cluster, and then this one cluster is split when going down. Agglomerative clustering is the opposite, where each object has its own cluster initially, and then clusters are merged when going up based on similarity or dissimilarity measures [6]. The proposed work has been applied on the United Kingdom Domestic Appliance-Level Electricity (UK-DALE) dataset, which holds consumption logs for each appliance in five dwellings [7].

The rest of this paper is organized as follows: related work is presented in Section 2. Section 3 introduces the proposed algorithm. Section 4 evaluates the results, and finally the conclusion and future work are derived in Section 5.

2. Related Work

Researchers have been focusing on extracting the correlation between appliances' usage, which is known as appliance–appliance association or inter-appliance association. Two methods of correlation are studied: frequent pattern mining and sequential pattern mining. Both methods are similar, except that sequential pattern mining is sensitive to the order by which the appliances are used.

Regarding the sequential pattern mining approach, the authors in [8] presented the StrPMiner algorithm using a batch-free approach to mine appliances' sequential patterns. The authors in [9] extracted appliances' sequential patterns using PrefixSpan on Apache Spark. Some other work present methods for extracting the relation between home activities and the appliances used. In [10], the authors developed a rule-mining algorithm using the JMeasure metric to extract appliances that are associated with activities. In [11], the authors used the Sequential PAttern Discovery using Equivalence classes (SPADE) algorithm to extract appliances' sequential patterns and then introduced its results to a proposed prediction model. In [12], the authors extracted appliances' priority based on the activity context, which may vary from one context to another. In [13], the authors developed a system that guarantees that the total power consumption will not exceed a certain limit by prioritizing the appliances based on user preference during activity context and rescheduling the unneeded ones.

Regarding the frequent pattern mining approach, the authors in [14] designed an algorithm using the sliding window technique to extract frequent usage patterns, and built up a recommendation system using the extracted patterns. The authors in [15] presented a usage notation to develop the Correlation Pattern Mining System (CPMS), extending the PrefixSpan algorithm. Then, in [16], the algorithm was modified to consider appliances' usage probability.

Since the data is being generated continuously by smart meters, it is essential to develop algorithms that mine usage patterns progressively so that they maintain the old discovered association rules in addition to extracting new ones; this was achieved in [17] and [18]. In [17], the authors enhanced CPMS, which extended the PrefixSpan algorithm to mine data progressively, and in [18], the authors developed an algorithm extending the pattern growth approach to extract appliance–appliance associations progressively.

Most of the work done using clustering techniques has focused on grouping customers with similar load profiles and paid a little attention to grouping appliances with similar behavior. The clustering of customers' load profile has been studied in [19–27]. The clustering of appliances has been studied in [28,29]. The authors in [28] proposed a proof of concept for clustering appliances' usage with respect to time—but for only one week—using hierarchical clustering and representing it using a dendogram. The authors in [29] extracted a load profile for each appliance, and then appliances with similar load profiles were grouped together.

3. Proposed Methodology

The proposed approach has extended the UTARM algorithm [5] for extracting appliances' association support values to a certain time. This time can be an hour, a day, a week, a month, or a season. Thereby, the chosen time factor is used to partition the temporal database. In this context, we have used the hour as our time factor. We have represented an appliance by its activity state as being OFF or ON, i.e., S = {0, 1}, and the value of power consumed is ignored.

The basic idea of extending the UTARM algorithm is that it calculates the association support value through taking temporal factors and utility factors into account. In our approach, the temporal factor is the hour, and it is used as a partitioning factor, while the utility factor is the probability of using an appliance at a certain hour.

Temporal association mining is the process of extracting temporal association rules from time-series data. Temporal association rules are an extension of frequent items' association rules with an aspect of time dimension. The basic idea of the time dimension is that each association rule can be valid for a period of time, which is sometimes called as an exhibition period or a lifespan [30]. Thus, they discover a group of items that frequently appear together at a specific time and last for an exhibition period.

Utility-oriented mining is the process of extracting frequent itemsets subjected to a weight or importance factor. Mining frequent itemsets assumes that all items in the itemset have the same weight, while in utility-oriented mining, each item in the itemset has a different weight that reflects users' preferences. The key for including the utility factor is to enhance the quality of the discovered association rules by considering their importance [31].

The proposed approach processes data in batches of 24 h at the end of each day. First, the raw data is preprocessed by generating a usage matrix; then, the utility value of each appliance is updated per each hour. Finally, the appliance–appliance association is discovered using association rules and hierarchical clustering. The proposed approach is illustrated in Figure 1.

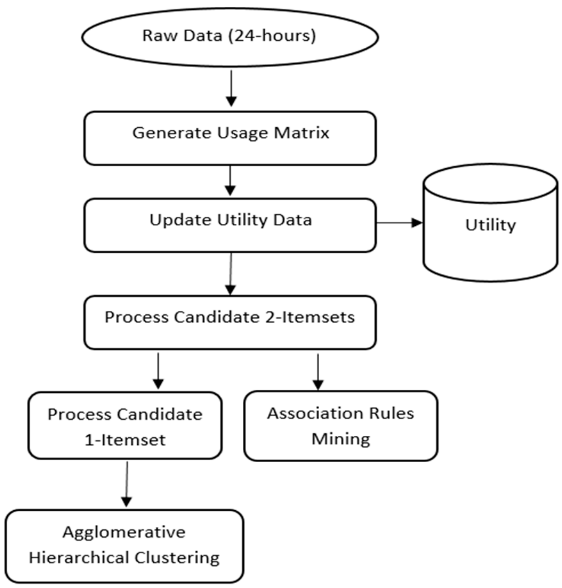

Figure 1. The Proposed Approach.

The proposed approach is achieved through three phases:

- Data Preparation.
- Calculating Appliances' Utility.
- Extracting Appliance–Appliance Association.

3.1. Data Preparation

In this work, we have used the UK-DALE dataset [7]. The dataset holds consumption logs for five houses with different durations. The raw consumption log consists of a timestamp and the power consumed in watts. The readings were logged every six seconds for each home at appliance level producing a dataset with more than 1.1 billion records. We have preprocessed our data in chunks of 24 h. For each day, we have stored only one record in the database holding the date and a generated usage matrix of size 24 * N, where N is the number of appliances. Each cell in the matrix holds a zero or one indicating the state of each appliance during the corresponding hour as an OFF or ON state, respectively. For example, house 1 holds data for 52 appliances during 4.3 years from 2012 to 2017. The total number of logs generated by this house is equal to 52 (appliances) * 4.3 (years) * 365 (days) * 24 (h) * 60 (min) * 10 (s), which is approximately around one billion records. In our approach, we have transformed the logs of each day to only one record. Thereby, the total number of logs generated are equal to 4.3 (years) * 365 (days), which is around 1570 records for the same house, achieving a much better performance and reducing the dataset size significantly.

3.2. Calculating Appliances' Utility

The aim of this phase is to study the preferences of home residents for using their appliances. These preferences are extracted by calculating the utility values for each appliance per each hour. The utility values are calculated by extending the utility association rule mining introduced in the UTARM. In our temporal database, each appliance (a) has a state (s) logged per each hour (h). The appliance state is represented by a one or zero, indicating whether the appliance is active or not, respectively. The Internal Utility (IU), External Utility (EU), and Utility (U) values are calculated for each appliance per each hour in the 24 h, since we have used the hour as our time granularity and partitioning factor. The IU is a quantitative value that measures the quantity of an item in an itemset. The quantity refers to the number of days that the appliance is logged as active. The IU is calculated as the summation of the state (s) values, since the activity state is represented by a zero or one. The EU is a value that reflects the significance of an item in an itemset. The significance is represented by the probability of having an appliance active in an hour all over the recorded days. The EU is calculated as the number of the active days divided by the total number of days (n). The U is a subjective value that reflects the weight of an item in an itemset. It is expressed by a function in terms of IU and EU. The U of an appliance is equal to the IU value multiplied by the EU value. The IU, EU, and U values are calculated using Equations (1)–(3), respectively.

$$\mathrm{IU}_a^h = \sum_{i=1}^{n} s_i^h \tag{1}$$

where n is the total number of days, and s is the state of the appliance at hour h in day i.

$$\mathrm{EU}_a^h = \frac{\text{number of active days}}{\text{total number of days}} \tag{2}$$

$$\mathrm{U}_a^h = \mathrm{IU}_a^h * \mathrm{EU}_a^h \tag{3}$$

In the next step, Transaction Weighted Utility (TWU) is the utility value determined per each partition. It is calculated by multiplying the maximum IU value and the maximum EU value generated by any appliance for each hour (h). It is described in Equation (4):

$$\mathrm{TWU}^h = \max(\mathrm{IU}^h) * \max(\mathrm{EU}^h) \tag{4}$$

where h is the hour.

3.3. Extracting Appliance–Appliance Associations

The aim of this phase is to extract appliance–appliance association rules considering utility values computed from the previous phase. The first step in the algorithm is to generate candidate 2-itemsets of home appliances, i.e., a1 and a2 represent a single candidate. For each candidate, the utility and temporal values are calculated per hour. The Frequency (FU) is the number of days for which a1 and a2 are active at a certain hour. The utility value (U) of candidate 2-itemsets, which are described in Equation (5), is calculated as the summation of the appliances' utility values. The Frequent Temporal Utility (FTU), which indicates the candidate support value and is described in Equation (6), is a function of the candidate 2-itemset's utility value and TWU:

$$\mathrm{U}^{\mathrm{h}}_{\mathrm{a1,a2}} = (\mathrm{IU}^{\mathrm{h}}_{\mathrm{a1}} * \mathrm{EU}^{\mathrm{h}}_{\mathrm{a1}}) * (\mathrm{IU}^{\mathrm{h}}_{\mathrm{a2}} * \mathrm{EU}^{\mathrm{h}}_{\mathrm{a2}}) \tag{5}$$

$$\mathrm{FTU}^{\mathrm{h}}_{\mathrm{a1,a2}} = \frac{\mathrm{FU}^{\mathrm{h}} * \mathrm{U}^{\mathrm{h}}_{\mathrm{a1,a2}}}{\mathrm{L} * \mathrm{TWU}^{\mathrm{h}}} \tag{6}$$

where a1 and a2 are appliances in the candidate 2-itemset, and L is the number of appliances in an itemset. Algorithm 1 outlines the steps used for the proposed approach by extending the UTARM algorithm. Algorithm 1 requires a minimum support (minsup) value, which is a threshold value for eliminating the infrequent patterns.

Algorithm 1 Extracting Appliances Temporal Utility

Require: minimum support *minsup*
Ensure: Utility-Oriented Temporal Association s
candidate2 ← *generate candidate 2-itemsets*
for each *item(a1, a2)* ∈ *candidate2* **do**
  for each *hour(h)* ∈ *24-h* **do**
    activeDays ← *days having item active in hour*
    length ← *length(activeDays)*
    if *length* > *0* **then**
      item.startAt^h ← *date of the first active day*
      item.endAt^h ← *date of the last active day*
      item.FU^h ← *length*
      item.U^h ← *calculated using Equation (5)*
      item.FTU^h ← *calculated using Equation (6)*
      item.selected^h ← *item.FTU*^{h} > *minsup*
      if *item.selected*^{h} **then**
        a1.FTU^h ← *calculated using Equation (6)*
        a2.FTU^h ← *calculated using Equation (6)*
      end if
    end if
  end for
end for

The next step of the algorithm is to generate association rules for discovering appliance–appliance associations. The appliance–appliance association is the extraction of appliances that are preferred to be used together: for example, the washing machine and the dryer. Association rules are expressions in the form of X $\Longrightarrow$ Y [32], indicating that the usage of appliance Y is associated with the usage of appliance X at hour h for an exhibition period. For each discovered association rule, support and confidence values are calculated. The support value, which is calculated using Equation (6), indicates the frequency of using the two appliances together at an hour. The confidence value, which is calculated using Equation (7), indicates the frequency of using appliance Y in the case of using appliance X at an hour h:

$$\mathrm{conf}(\mathrm{X} \Longrightarrow \mathrm{Y})^{\mathrm{h}} = \frac{\mathrm{ftu}(\mathrm{XUY}\ )^{\mathrm{h}}}{\mathrm{ftu}(\mathrm{X})^{\mathrm{h}}} \tag{7}$$

Algorithm (2) outlines the steps used for generating the appliances' association rules. Algorithm 2 requires a minimum support (minsup) value and a minimum confidence (minconf) value. The minsup is a threshold value for processing only the frequent patterns. The minconf is a threshold value for eliminating the insignificant association rules.

Algorithm 2: Generating Appliances' Association Rules

Require: minimum support *minsup*, minimum confidence *minconf*, candidate 2-itemsets *candidate2*
Ensure: Appliances Association Rules
for each *item(a1, a2)* $\in$ *candidate2* **do**
  for each *hour(h)* $\in$ *24-h* **do**
    if *a1.FTU*^{h} > *minsup* **then**
      $conf(a1 \Longrightarrow a2)^h \leftarrow$ *calculated using equation* 7
      $conf(a1 \Longrightarrow a2)^h.selected = conf > minconf$
    end if
    if *a2.FTU*^{h} > *minsup* **then**
      $conf(a2 \Longrightarrow a1)^h \leftarrow$ *calculated using equation* 7
      $conf(a2 \Longrightarrow a1)^h.selected = conf > minconf$
    end if
  end for
end for

Finally, hierarchical clustering is applied on the candidate 1-itemset using the FTU support value, which indicates the frequency of using an appliance at each hour. Hierarchical clustering groups appliances with similar usage behavior with respect to time.

4. Evaluation and Results

In this section, a comprehensive analysis was conducted to explain our results. To the best of our knowledge, our proposed approach is the first to consider appliances' utility with respect to temporal mining. The architecture of the proposed approach has succeeded to mine smart meters' data progressively without mining the whole database whenever new data is transmitted. The progressive approach is achieved by utilizing the computed utility data in addition to mining the newly generated data only at the end of the day.

Clustering analysis has been represented using a dendogram. The horizontal axis of the dendrogram represents the similarity or dissimilarity distance between clusters, and the vertical axis represents appliances clustered by their similar usage, as represented in Figures 2–6.

The results of house 1 are represented in Figure 2.

Some appliances are associated together at the same hours: for example, the usage of the samsung_charger and bedroom_chargers represents a cluster. Also, the usage of the amp_livingroom and subwoofer_livingroom represent another cluster having similar support values.

The results of house 2 are represented in Figure 3.

The usage of a cooker, rice_cooker, represents a cluster that might reveal their usage together during the cooking activity, and this observation proves that home activities can be identified from appliances' usage.

The results of house 3 are represented in Figure 4.

If the similarity or dissimilarity threshold is set to one, then three clusters will be extracted, which are {electric_heater, kettle}, {projector}, and {laptop}. If the similarity or dissimilarity threshold is set to two, then two clusters will be extracted, which are the {electric_heater, kettle, projector} and {laptop}.

The results of house 4 are represented in Figure 5.

It is noted that the freezer and the gas_boiler are represented in one cluster, which makes sense, since these appliances have a thermostat component making them active during the 24 hours so they have similar activity usage.

The results of house 5 are represented in Figure 6.

We can observe that appliances such as the toaster, kettle, stream_iron, and nespresso_pixie are grouped together having no support values, which means that these appliances have no frequent usage patterns.

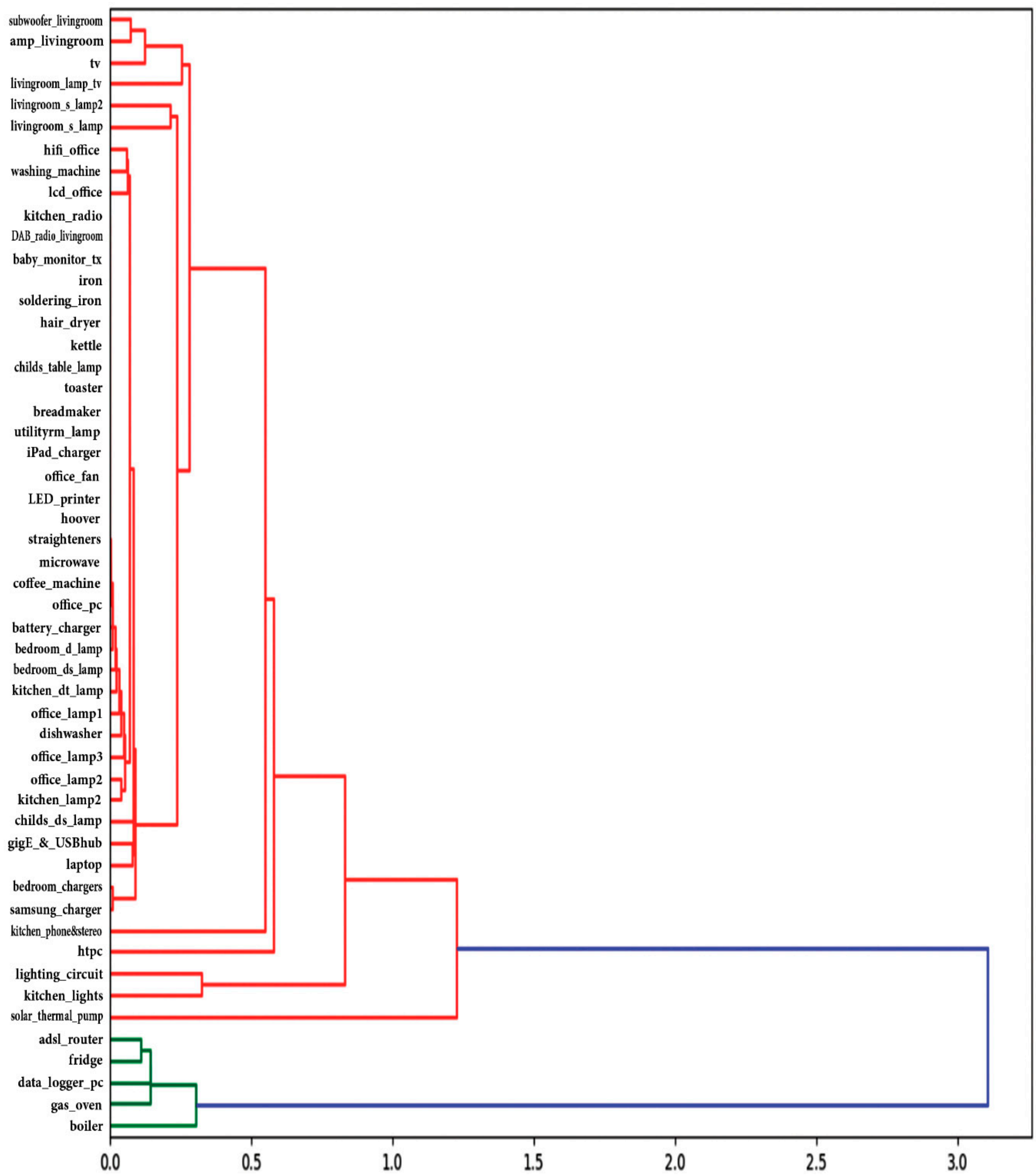

Figure 2. House 1 Appliance–Appliance Associations.

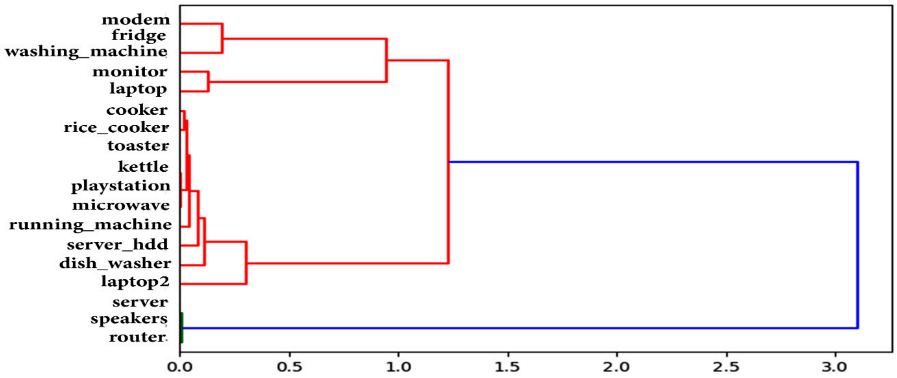

Figure 3. House 2 Appliance–Appliance Associations.

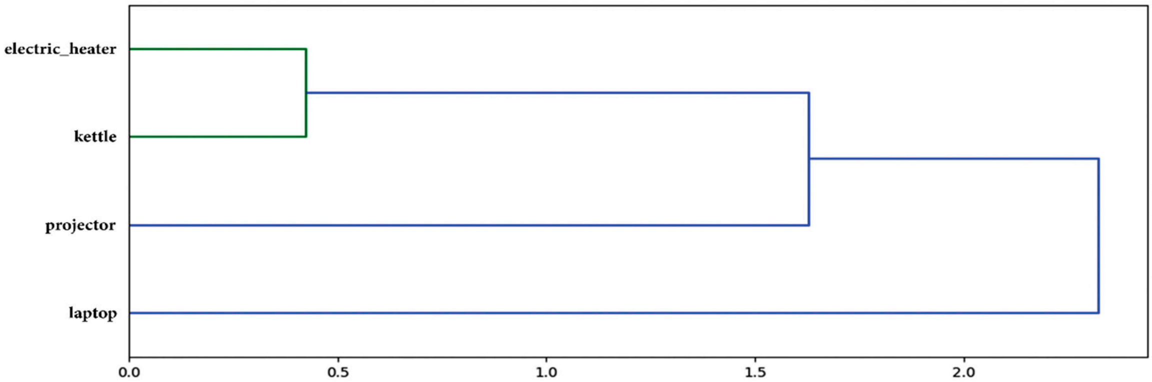

Figure 4. House 3 Appliance–Appliance Associations.

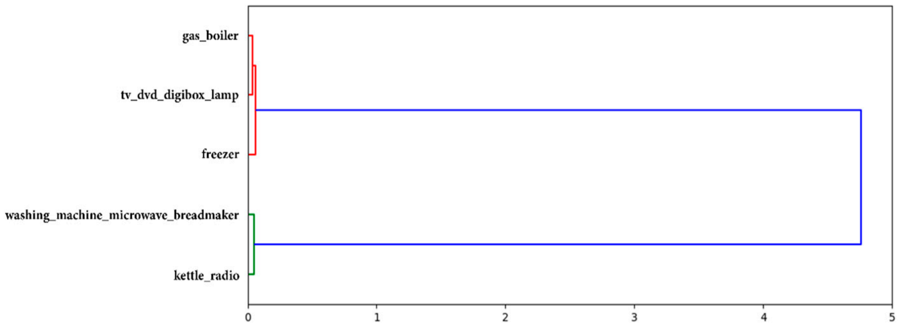

Figure 5. House 4 Appliance–Appliance Associations.

Regarding the discovered association rules, we can observe that the priority of appliances' usage differs with respect to time; also, residents' behavior changes over the time, so some findings may expire with time. Thus, each association rule is associated with a certain hour and has an exhibition period indicating the validity of the discovered association rule. Moreover, it is noted that appliances that are always active during the 24 h; for example, the fridge results in associations with all appliances at any time. In our work, we have set the minimum confidence value to be 75% to limit the number of the rules discovered.

The duration of the logged in data of house 1 is around 4.3 years. The number of association rules discovered is 1280 rules. Table 1 represents a sample of the associated discovered rules of house 1.

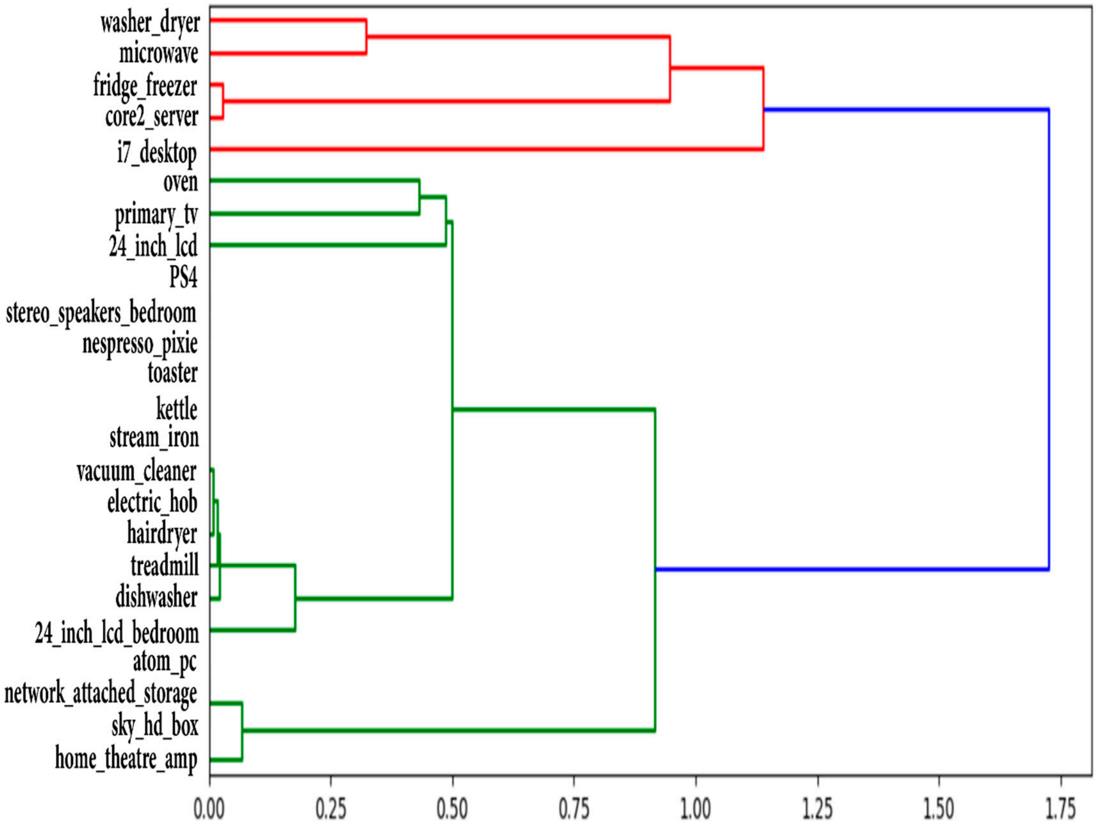

Figure 6. House 5 Appliance–Appliance Associations.

Table 1. Sample of House 1 Appliance–Appliance Associations.

Association Rule	Confidence	Hour	From (dd-mm-yy)	To (dd-mm-yy)
kitchen_lights $\Longrightarrow$ tv	96%	22	2-1-2013	22-4-2017
kitchen_lights $\Longrightarrow$ tv	96%	23	9-1-2013	25-4-2017
tv $\Longrightarrow$ amp_livingroom	100%	0	1-1-2013	22-4-2017
tv $\Longrightarrow$ amp_livingroom	100%	20	1-1-2013	25-4-2017
tv $\Longrightarrow$ amp_livingroom	100%	21	1-1-2013	25-4-2017
tv $\Longrightarrow$ amp_livingroom	100%	22	1-1-2013	25-4-2017
tv $\Longrightarrow$ amp_livingroom	100%	23	1-1-2013	25-4-2017
amp_livingroom $\Longrightarrow$ tv	92%	0	1-1-2013	22-4-2017
amp_livingroom $\Longrightarrow$ tv	91%	21	1-1-2013	25-4-2017
amp_livingroom $\Longrightarrow$ tv	97%	22	1-1-2013	25-4-2017
amp_livingroom $\Longrightarrow$ tv	96%	23	1-1-2013	25-4-2017
tv $\Longrightarrow$ subwoofer_livingroom	91%	0	1-1-2013	22-4-2017
tv $\Longrightarrow$ subwoofer_livingroom	100%	20	1-1-2013	25-4-2017
tv $\Longrightarrow$ subwoofer_livingroom	98%	21	1-1-2013	25-4-2017
tv $\Longrightarrow$ subwoofer_livingroom	95%	22	1-1-2013	25-4-2017
tv $\Longrightarrow$ subwoofer_livingroom	94%	23	1-1-2013	25-4-2017
subwoofer_livingroom $\Longrightarrow$ tv	90%	0	13-3-2013	22-4-2017
subwoofer_livingroom $\Longrightarrow$ tv	91%	21	12-3-2013	25-4-2017
subwoofer_livingroom $\Longrightarrow$ tv	99%	22	12-3-2013	25-4-2017
subwoofer_livingroom $\Longrightarrow$ tv	97%	23	12-3-2013	25-4-2017

The rules reveal that the subwoofer_livingroom, tv, and amp_livingroom are associated together by being active at hours 0, 21, 22, and 23. Also, the tv and kitchen_lights are associated together at hours 22 and 23.

The duration of the logged in data of house 2 is around seven months. It is found that the speakers, server, and router are highly associated with each other, and are always active in the background during the 24 h. Table 2 represents a sample of the discovered association rules of house 2, excluding the speakers, server, and router appliances, which are always active in the background.

Table 2. Sample of House 2 Appliance–Appliance Associations.

Association Rule	Confidence	Hour	From (dd-mm-yy)	To (dd-mm-yy)
laptop $\Longrightarrow$ monitor	90–96%	0-23	17-2-2013	10-10-2013
monitor $\Longrightarrow$ laptop	98–100%	0-23	17-2-2013	10-10-2013
laptop2 $\Longrightarrow$ laptop	95%	21	17-4-2013	9-10-2013
laptop2 $\Longrightarrow$ laptop	89%	22	17-4-2013	9-10-2013
laptop2 $\Longrightarrow$ laptop	100%	23	17-4-2013	9-10-2013
laptop $\Longrightarrow$ modem	89%	9	22-3-2013	6-10-2013
laptop $\Longrightarrow$ modem	78%	10	1-3-2013	6-10-2013
modem $\Longrightarrow$ laptop	79%	22	21-5-2013	9-10-2013
modem $\Longrightarrow$ laptop	84%	23	20-5-2013	9-10-2013
laptop2 $\Longrightarrow$ monitor	89%	21	17-4-2013	9-10-2013
laptop2 $\Longrightarrow$ monitor	83%	22	17-4-2013	9-10-2013
laptop2 $\Longrightarrow$ monitor	100%	23	16-4-2013	8-10-2013
monitor $\Longrightarrow$ modem	92%	9	22-3-2013	27-9-2013
monitor $\Longrightarrow$ modem	81%	10	18-3-2013	1-10-2013
modem $\Longrightarrow$ monitor	77%	23	20-5-2013	9-10-2013

It is observed that the monitor and the laptop are associated together. However, the confidence of the laptop usage during the monitor usage has a higher value than the usage of the monitor during the laptop usage. The duration of the logged in data of house 3 is 37 days. Table 3 represents a sample of the associated discovered rules of house 3.

Table 3. Sample of House 3 Appliance–Appliance Associations.

Association Rule	Confidence	Hour	From (dd-mm-yy)	To (dd-mm-yy)
electric_heater $\Longrightarrow$ laptop	100%	2	12-3-2013	26-3-2013
electric_heater $\Longrightarrow$ laptop	71.80%	9	12-3-2013	3-4-2013

We can observe that the usage of the laptop is associated with the usage of the electric_heater but with different confidence values based on the hour. In hour 2, the confidence was 100%, while it was 71.8% in hour 9. This observation is because the electric_heater was always active in the background, and the laptop was associated with hours 2 and 9. Thereby, an association between those two appliances is extracted at hours 2 and 9.

The duration of the logged in data of house 4 is around five months. Table 4 represents a sample of these background associations in house 4.

Table 4. Sample of House 4 Appliance–Appliance Associations.

Association Rule	Confidence	Hour	From (dd-mm-yy)	To (dd-mm-yy)
tv_dvd_digibox_lamp $\Longrightarrow$ gas_boiler	100.0%	0–24	9-3-2013	1-10-2013
gas_boiler $\Longrightarrow$ tv_dvd_digibox_lamp	100.0%	0–24	9-3-2013	1-10-2013
tv_dvd_digibox_lamp $\Longrightarrow$ freezer	98–100%	0–24	9-3-2013	1-10-2013
freezer $\Longrightarrow$ tv_dvd_digibox_lamp	99–100%	0–24	9-3-2013	1-10-2013
gas_boiler $\Longrightarrow$ freezer	97–100%	0–24	10-3-2013	1-10-2013

It is noted that appliances such as the tv_dvd_digibox_lamp, gas_boiler, and freezer are always working in the background.

The duration of the logged data of house 5 is around 4.5 months. Table 5 represents a sample of the association rules discovered for house 5, excluding appliances that are always active in the background.

Table 5. Sample of House 5 Appliance–Appliance Associations.

Association Rule	Confidence	Hour	From (dd-mm-yy)	To (dd-mm-yy)
primary_tv $\Longrightarrow$ i7_desktop	100%	21	29-6-2014	11-11-2014
primary_tv $\Longrightarrow$ i7_desktop	99%	22	29-6-2014	12-11-2014
primary_tv $\Longrightarrow$ i7_desktop	92%	23	29-6-2014	12-11-2014
24_inch_lcd $\Longrightarrow$ i7_desktop	100%	11	30-6-2014	6-9-2014
24_inch_lcd $\Longrightarrow$ i7_desktop	100%	12	30-6-2014	7-9-2014
24_inch_lcd $\Longrightarrow$ i7_desktop	100%	13	30-6-2014	7-9-2014
24_inch_lcd $\Longrightarrow$ i7_desktop	100%	14	30-6-2014	7-9-2014
24_inch_lcd $\Longrightarrow$ i7_desktop	100%	15	30-6-2014	7-9-2014
24_inch_lcd $\Longrightarrow$ i7_desktop	100%	16	30-6-2014	6-9-2014
24_inch_lcd $\Longrightarrow$ i7_desktop	100%	17	30-6-2014	6-9-2014
24_inch_lcd $\Longrightarrow$ i7_desktop	100%	18	29-6-2014	6-9-2014
24_inch_lcd $\Longrightarrow$ i7_desktop	100%	19	29-6-2014	6-9-2014
24_inch_lcd $\Longrightarrow$ i7_desktop	100%	20	30-6-2014	6-9-2014
oven $\Longrightarrow$ i7_desktop	100%	19	8-7-2014	13-11-2014
oven $\Longrightarrow$ i7_desktop	100%	20	29-6-2014	10-11-2014
oven $\Longrightarrow$ i7_desktop	100%	21	29-6-2014	11-11-2014
oven $\Longrightarrow$ i7_desktop	100%	22	30-6-2014	12-11-2014
oven $\Longrightarrow$ i7_desktop	100%	23	30-6-2014	12-11-2014

It is observed that the appliance i7_desktop has a high confidence value for being used during the usage of the primary_tv, 24_inch_lcd, and oven. This is because the appliance i7_desktop is highly associated with hours from 11:00 to 00:00; thus, any appliance that is associated with these hours will result in an association with the i7_desktop.

The conducted results succeeded in extracting appliances' associations using hierarchical clustering and association rules mining. Table 6 shows a comparison between both methods.

Table 6. Comparison between Hierarchical Clustering and Association Rules.

	Association Rules	Hierarchical Clustering
Associations Discovered	Based on the behavior for each hour per day.	Based on the behavior across the 24 h.
Exhibition Period	Identifies the lifespan for the discovered associations.	Identifies the associations across all the recorded days.
History	Behavioral history can be obtained.	No history can be obtained.

From this comparison, hierarchical clustering can be described as a generic approach, and association rules can be described as a specific approach for extracting appliance–appliance associations.

The proposed approach is developed using Python and MongoDB on an Intel(R) Core(TM) i7-6500U CPU and a RAM of 8.00 GB. The proposed approach was evaluated by comparing its conducted results with the pattern-growth approach, since that most of the previous work extends it. It succeeded in extracting associations that have a higher weight in addition to achieving a better runtime performance. Figure 7 shows the runtime analysis for mining the generated data for only one day. The x-axis represents the number of appliances, and the y-axis represents the execution time in seconds.

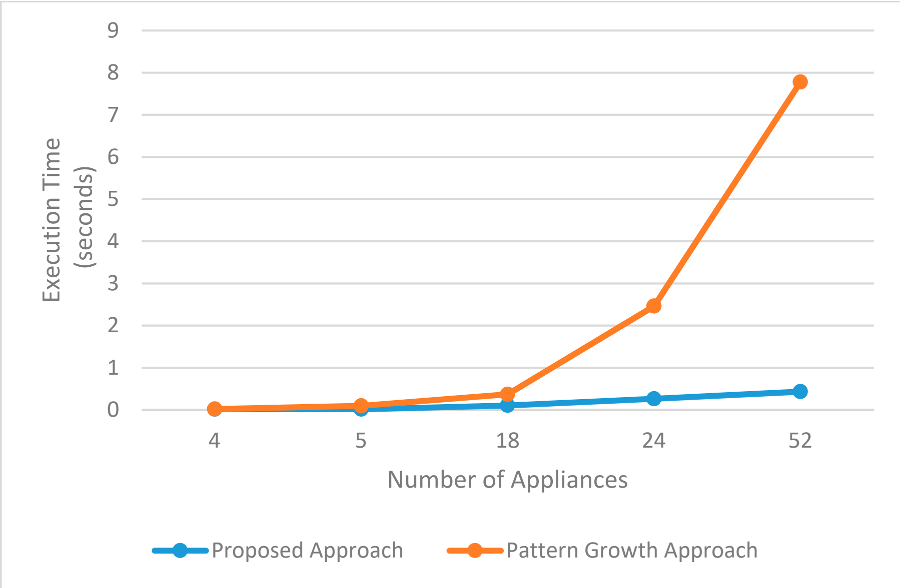

Figure 7. Runtime Analysis.

The experiment is performed on smart meter data generated in one day, since the data is processed at the end of each day. It is observed that the proposed approach and the pattern-growth approach have similar runtimes for a small number of appliances. However, the proposed approach has a better performance as the number of appliances increases. That's because the cost of building the frequent-pattern tree and pruning the infrequent patterns is high when the number of appliances is increased.

The conducted results can integrate with the DR management techniques developed in [33,34] through home energy management systems (HEMS). Thereby, they respond to DR programs and reschedule home appliances while keeping in consideration the extracted preferences of home residents. One example is if there are two clusters of appliances that are associated together, and the first cluster has higher confidence values than the second one. Then, at peak hours, HEMS should keep appliance associations that have higher confidence values active together and reschedule the others to another time.

5. Conclusions and Future Work

Developing DR programs has become an interest for saving energy in the residential sector. Energy is wasted by home residents due to their lack of knowledge about their consumption. Raising their awareness will guide them toward an efficient use of energy. Preserving home residents' comfort level is a key factor for motivating them to respond to DR programs. Thus, a lot of research is presented in order to mine smart meter data for extracting the preferences of home residents. The conducted results can be integrated with home energy management systems to respond to DR programs.

In this work, we have extended the UTARM algorithm to discover associations between appliances. The basic idea of using UTARM is that an association is measured based on two factors: the temporal factor, which was the hour, and the utility factor, which was the weight of using an appliance at the hour. Our work mine data progressively at the end of each day in chunks of 24 h. Initially, the utility values are updated; then, FTU support values are calculated, revealing the appliance association level to an hour. Then, hierarchical agglomerative clustering is applied using FTU support values to group appliances with similar usage together. The results achieved are represented using a dendogram.

The hierarchical clustering and UTARM algorithm succeeded in discovering appliance–appliance associations. However, the UTARM algorithm identifies the validity of the association rule through its exhibition period, as some rules may expire as residents' behavior changes.

In the future work, our proposed approach can be extended to extract appliance–appliance associations that take into account the value of power consumed and the hour of use.

Author Contributions: Conceptualization, S.O. and M.A.; Methodology, S.O.; Software, S.O.; Validation, S.O., M.A. and A.-B.M.S.; Formal Analysis, S.O. and M.A.; Investigation, S.O.; Resources, S.O.; Data Curation, S.O.; Writing-Original Draft Preparation, S.O.; Writing-Review and Editing, S.O., M.A. and A.-B.M.S.; Supervision, M.A. and A.-B.M.S.

Funding: This research received no external funding.

Conflicts of Interest: There are no conflicts of interest.

References

1. Zhou, S.; Brown, M.A. Smart meter deployment in Europe: A comparative case study on the impacts of national policy schemes. *J. Clean. Prod.* **2017**, *144*, 22–32. [CrossRef]
2. Bekara, C. Security issues and challenges for the IoT-based smart grid. *Procedia Comput. Sci.* **2014**, *34*, 532–537. [CrossRef]
3. Palacios-García, E.J.; Guan, Y.; Savaghebi, M.; Vásquez, J.C.; Guerrero, J.M.; Moreno-Munoz, A.; Ipsen, B.S. Smart metering system for microgrids. In Proceedings of the 41st Annual Conference of the IEEE on Industrial Electronics Society (IECON), Yokohama, Japan, 9–12 November 2015; pp. 003289–003294.
4. Albadi, M.H.; El-Saadany, E.F. A summary of demand response in electricity markets. *Electr. Power Syst. Res.* **2008**, *78*, 1989–1996. [CrossRef]
5. Maragatham, G.; Lakshmi, M. UTARM: An efficient algorithm for mining of utility-oriented temporal association rules. *Int. J. Knowl. Eng. Data Min.* **2015**, *3*, 208–237. [CrossRef]
6. Aghabozorgi, S.; Shirkhorshidi, A.S.; Wah, T.Y. Time-series clustering—A decade review. *Inf. Syst.* **2015**, *53*, 16–38. [CrossRef]
7. Kelly, J.; Knottenbelt, W. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Sci. Data* **2015**, *2*, 150007. [CrossRef]
8. Hassani, M.; Beecks, C.; Töws, D.; Seidl, T. Mining Sequential Patterns of Event Streams in a Smart Home Application. In Proceedings of the Lerner Wissen Adaption (LWA) Conference, Trier, Germany, 7–9 October 2015; pp. 159–170.
9. Honarvar, A.R.; Sami, A. Extracting usage patterns from power usage data of homes' appliances in smart home using big data platform. *Int. J. Inf. Technol. Web Eng.* **2016**, *11*, 39–50. [CrossRef]
10. Rollins, S.; Banerjee, N. Using rule mining to understand appliance energy consumption patterns. In Proceedings of the 2014 IEEE International Conference on Pervasive Computing and Communications (PerCom), Budapest, Hungary, 24–28 March 2014; pp. 29–37.
11. Ding, Y.; Borges, J.; Neumann, M.A.; Beigl, M. Sequential pattern mining—A study to understand daily activity patterns for load forecasting enhancement. In Proceedings of the 2015 IEEE First International Smart Cities Conference (ISC2), Guadalajara, Mexico, 25–28 October 2015; pp. 1–6.
12. Zhang, X.; Kato, T.; Matsuyama, T. Learning a context-aware personal model of appliance usage patterns in smart home. In Proceedings of the 2014 IEEE Innovative Smart Grid Technologies-Asia (ISGT ASIA), Kuala Lumpur, Malaysia, 20–23 May 2014; pp. 73–78.
13. Liao, Y.S.; Liao, H.Y.; Liu, D.R.; Fan, W.T.; Omar, H. Intelligent Power Resource Allocation by Context-Based Usage Mining. In Proceedings of the 2015 IIAI 4th International Congress on Advanced Applied Informatics, Okayama, Japan, 12–16 July 2015; pp. 546–550.
14. Schweizer, D.; Zehnder, M.; Wache, H.; Witschel, H.F.; Zanatta, D.; Rodriguez, M. Using consumer behavior data to reduce energy consumption in smart homes: Applying machine learning to save energy without lowering comfort of inhabitants. In Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 9–11 December 2015; pp. 1123–1129.
15. Chen, Y.C.; Peng, W.C.; Lee, W.C. A novel system for extracting useful correlation in smart home environment. In Proceedings of the 2013 IEEE 13th International Conference on Data Mining Workshops, Dallas, TX, USA, 7–10 December 2013; pp. 357–364.

16. Chen, Y.C.; Chen, C.C.; Peng, W.C.; Lee, W.C. Mining correlation patterns among appliances in smart home environment. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Tainan, Taiwan, 13–16 May 2014; pp. 222–233.
17. Chen, Y.C.; Hung, H.C.; Chiang, B.Y.; Peng, S.Y.; Chen, P.J. Incrementally mining usage correlations among appliances in smart homes. In Proceedings of the 2015 18th International Conference on Network-Based Information Systems, Taipei, Taiwan, 2–4 September 2015; pp. 273–279.
18. Singh, S.; Yassine, A. Mining energy consumption behavior patterns for households in smart grid. *IEEE Trans. Emerg. Top. Comput.* **2017**. [CrossRef]
19. McLoughlin, F.; Duffy, A.; Conlon, M. A clustering approach to domestic electricity load profile characterisation using smart metering data. *Appl. Energy* **2015**, *141*, 190–199. [CrossRef]
20. Haben, S.; Singleton, C.; Grindrod, P. Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE Trans. Smart Grid* **2016**, *7*, 136–144. [CrossRef]
21. Flath, C.; Nicolay, D.; Conte, T.; van Dinther, C.; Filipova-Neumann, L. Cluster analysis of smart metering data. *Bus. Inf. Syst. Eng.* **2012**, *4*, 31–39. [CrossRef]
22. Benítez, I.; Quijano, A.; Díez, J.L.; Delgado, I. Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers. *Int. J. Electr. Power Energy Syst.* **2014**, *55*, 437–448. [CrossRef]
23. Kwac, J.; Flora, J.; Rajagopal, R. Household energy consumption segmentation using hourly data. *IEEE Trans. Smart Grid* **2014**, *5*, 420–430. [CrossRef]
24. Cao, H.Â.; Beckel, C.; Staake, T. Are domestic load profiles stable over time? An attempt to identify target households for demand side management campaigns. In Proceedings of the 2013 39th Annual Conference of the IEEE Industrial Electronics Society (IECON), Vienna, Austria, 10–13 November 2013; pp. 4733–4738.
25. Chelmis, C.; Kolte, J.; Prasanna, V.K. Big data analytics for demand response: Clustering over space and time. In Proceedings of the 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 29 October–1 November 2015; pp. 2223–2232.
26. Quilumba, F.L.; Lee, W.J.; Huang, H.; Wang, D.Y.; Szabados, R.L. Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities. *IEEE Trans. Smart Grid* **2015**, *6*, 911–918. [CrossRef]
27. Al-Otaibi, R.; Jin, N.; Wilcox, T.; Flach, P. Feature construction and calibration for clustering daily load curves from smart-meter data. *IEEE Trans. Ind. Inform.* **2016**, *12*, 645–654. [CrossRef]
28. Gajowniczek, K.; Ząbkowski, T. Data mining techniques for detecting household characteristics based on smart meter data. *Energies* **2015**, *8*, 7407–7427. [CrossRef]
29. Adika, C.O.; Wang, L. Autonomous appliance scheduling based on time of use probabilities and load clustering. In Proceedings of the 2012 10th International Power & Energy Conference (IPEC), Ho Chi Minh City, Vietnam, 12–14 December 2012; pp. 42–47.
30. Ale, J.M.; Rossi, G.H. An approach to discovering temporal association rules. In Proceedings of the 2000 ACM Symposium on Applied Computing-Volume 1, Como, Italy, March 19–21 2000; pp. 294–300.
31. Zida, S.; Fournier-Viger, P.; Lin, J.C.; Wu, C.W.; Tseng, V.S. EFIM: A highly efficient algorithm for high-utility itemset mining. In Proceedings of the Mexican International Conference on Artificial Intelligence, Cuernavaca, Mexico, 25–31 October 2015; pp. 530–546.
32. Srikant, R.; Vu, Q.; Agrawal, R. Mining association rules with item constraints. In Proceedings of the Third Conference on Knowledge Discovery in Databases and Data Mining (KDD), Newport Beach, CA, USA, 14–17 August 1997; pp. 67–73.
33. Chai, B.; Chen, J.; Yang, Z.; Zhang, Y. Demand response management with multiple utility companies: A two-level game approach. *IEEE Trans. Smart Grid* **2014**, *5*, 722–731. [CrossRef]
34. Apostolopoulos, P.A.; Tsiropoulou, E.E.; Papavassiliou, S. Demand response management in smart grid networks: A two-stage game-theoretic learning-based approach. *Mob. Netw. Appl.* **2018**, 1–14. [CrossRef]