IEEE TRANSACTIONS ON
EMERGING TOPICS
IN COMPUTING

Received 1 September 2016; revised 21 February 2017; accepted 2 April 2017.
Date of publication 7 April 2017; date of current version 4 September 2019.

Digital Object Identifier 10.1109/TETC.2017.2692098

Mining Energy Consumption Behavior Patterns for Households in Smart Grid

SHAILENDRA SINGH AND ABDULSALAM YASSINE, (Member, IEEE)

The authors are with the Distributed and Collaborative Virtual Environments Research Laboratory, School of Electrical Engineering and Computer Science, University of Ottawa, 800 King Edward Street, Ottawa ONK1N 6N5, Canada

CORRESPONDING AUTHOR: S. SINGH (ssing120@uottawa.ca)

ABSTRACT Inarguably, buying-in consumer confidence through respecting their energy consumption behavior and preferences in various energy programs is imperative but also demanding. Household energy consumption patterns, which provide great insight into consumers energy consumption behavioral traits, can be learned by understanding user activities along with appliances used and their time of use. Such information can be retrieved from the context-rich smart meters big data. However, the main challenge is how to extract complex interdependencies among multiple appliances operating concurrently, and identify appliances responsible for major energy consumption. Furthermore, due to the continuous generation of energy consumption data, over a period of time, appliance associations can change. Therefore, they need to be captured regularly and continuously. In this paper, we propose an unsupervised progressive incremental data mining mechanism applied to smart meters energy consumption data through frequent pattern mining to overcome these challenges. This can establish a foundation for efficient energy demand management while ameliorating end-user participation. The details and the results of evaluation of the proposed mechanism using real smart meters dataset are also presented in this paper.

INDEX TERMS Incremental progressive data mining, frequent pattern mining, behavioral analysis, energy consumption patterns

I. INTRODUCTION

With the massive amounts of data continuously being collected through household smart meters, it is becoming of interest to energy producers, utility companies, and end-users to mine energy consumption behavioral patterns from the ever growing data [1]. The use of energy in residential premises is related to the activities that users perform, the time at which appliances are used, and the interdependencies with-in appliances that may be used simultaneously. For example, a consumer may find it convenient to operate the dishwasher while the dryer is on or like to work on the computer or watch TV while cooking or listening to music. These relationships; i.e., appliance-appliance and appliance-time are vital to understanding power consumption in a typical house. smart meters big data as a large time series can contain many frequent patterns of appliance usage [2]. The dynamic discovery of these patterns is important for many decision-making processes, such as consumer energy consumption behavior analysis, demand response optimization as well as improving energy reduction recommendations.

Behavioral analytics of energy consumption for households has several practical applications, particularly with the introduction of Open Automatic Demand Response (OpenADR) protocol [3], [4]. OpenADR protocol was originally proposed for commercial customers, but with the introduction of the smart-grid, the protocol was modified in 2009 to accommodate residential consumers [5]. This is indeed a significant step since households are accountable for 40 percent of total energy consumption according to the US energy information administration report for 2015 and 2016 [6]. However, till date, utilities have not been very successful in getting wide acceptance of demand response mechanisms from residential consumers compared to industrial consumers. The study in [7] shows that one of the major barriers to effective implementation of demand response techniques is a lack of good knowledge of actual energy consumption in households. Unlike industrial consumers who can define, forecast and follow energy consumption patterns, a residential consumer has a direct impact of behavioral attributes over energy consumption. Furthermore, current implementations of demand response mechanisms pay little attention, if at all, to variable usage correlation relationships among appliances and, as a result, do not meet the user's preference requirements. In this regard, it is of paramount importance to understand the fine granularities of consumers' energy usage behavior, which can lead to customer participation and can directly be translated into customer preferences; for example, acknowledgment of benefits versus discomfort from modifying customer's behavior according to suggested demand response plans.

Technically, the aforementioned objective is very challenging because generation of energy consumption data from smart meters is an ongoing process, and over a period of time, appliance associations can change or new ones can be established; this requires a continuous progressive learning approach to comprehensively capture these variations for an updated reflection of consumers' energy consumption behavior. Additionally, due to the heterogeneity of how consumers go about their energy use, determining which appliances should participate in the demand response events can be very costly, since it is not feasible to regularly contact every consumer and obtain their energy consumption characteristics, which might vary according to time. At the same time, besides identifying appliances responsible for peak power load, it is important to identify the appliances (manual and automatic operation) responsible for the substantial contribution towards household energy usage. But, it is challenging to recognize such appliances without analyzing raw energy consumption data.

To tackle the above challenges, this paper proposes an unsupervised progressive incremental data mining mechanism applied to smart meters energy consumption data through frequent pattern mining. This technique addresses the need to comprehensively consider human behavioral variations such as high uncertainty in order of use, varying time of use, and increased or reduced frequency of use of appliances; thus, providing a foundation for a data-driven, well informed and time appropriate decision-making process. The main contributions of this paper are as follows:

- A behavioral energy consumption pattern mining for appliance-appliance and appliance-time associations derived from smart meter data to provide insight into consumers' energy consumption decision patterns. Additionally, the Appliances of Interest (AoI) are identified, which usually have smaller power load footprint but contribute major energy consumption due to extensive use. Deriving these individual temporal and behavioral choices from consumption patterns is vital to achieving efficient energy demand management in order to gain consumers' buy-in.
- Incremental progressive data mining as a means for the smart-grid environment to capture consumers' behavioral variations. Which extends FP-growth and k-means algorithms to incorporate incremental data mining behavior. Our model does the data mining in an online and distributed fashion at the individual household level while supporting decision making towards energy efficiency with improved accuracy. The technique employs k-means clustering through dynamic programming, dynamic determination of k (number of clusters) through use the of *Silhouette coefficient* for k-means clustering, and application of Kulczynski measure (*Kulc*) along with Imbalance Ratio (*IR*), in Frequent Pattern mining, as pattern interestingness measures to eliminate uninteresting patterns and rules.

For the evaluation of the proposed mechanism, the UK Domestic Appliance Level Electricity dataset is used [8]; which is a context-rich dataset and includes time series data of the energy consumption collected from 2012 to 2015 with a time resolution of 6 seconds for five houses having 109 appliances from Southern England. It must be noted that disaggregated data using advance Nonintrusive Load Monitoring (NILM) and energy disaggregation solutions can be easily exchanged between utilities and houses in a smart-grid environment. The dataset is used to conduct an in-depth analysis of the raw energy consumption data to substantiate the results. We also used results from data clustering to support the findings on AoI appliances.

The organization of this paper is as follows: The next Section II discusses the related work. In Section III, the proposed model is presented followed by evaluation results in the Section V. Finally, we conclude the paper and discuss future directions in Section VI.

II. RELATED WORK

Learning consumers' behavioral characteristics towards energy consumption is one key to the success of many or most of energy savings/efficiency programs. A technical report by [9] and a study by [10] provide extensive arguments in support of exploiting behavioral energy consumption information to encourage and obtain greater energy efficiency. Authors in [11] studied the perspective of extracting information about consumers from energy consumption patterns. The impact of behavioral changes towards energy savings was also examined by [12] and [13], and end-user participation towards effective and improved energy savings were emphasized. Mining appliance usage and association in the form of frequent patterns from context-aware smart meter data can reveal surprising underlying information. In this section, a review on existing studies that are directly related to behavior analytics and frequent pattern mining of smart meters big data is presented.

The study by [14] analyzed Hidden Markov Model and cluster analysis to establish a correlation between energy consumption patterns and user behavioral characteristics to support energy program enrollments. In the work [15] and [16], authors utilized historical energy consumption behavior to decide appliance scheduling for efficient energy usage, and emphasized on active participation by end-user to achieve a significant reduction in household energy consumption; but, they fail to consider inter-appliance associations. The study in [17] proposed a methodology to extract user behavioral characteristics in form of activity of daily living (ADL) tasks with context and temporal information to support efficient power management. In [18], personal appliance usage habits were learned from appliance usage patterns, which is dependent on ADLs. Studies [19] and [20] presented a demand side management technique to control peak load hikes by shifting the operation of high power load appliances while minimizing discomfort to consumers, but did not completely consider behavioral variations that may occur with time such as weekdays versus weekends, months and seasons. In [21] authors proposed to analyze the impact of price variations over occupants energy consumption behavior across houses. The work by [22] studied the energy consumption data to determine groups of customers based on energy usage behavior and related variability.

Parallel to the above studies, several research papers discussed mining methods of energy consumption. The approach provided by [23] inspected the rule mining to examine related behavioral characteristics and identify the associations between energy consumption and the time of appliance usage to assist energy conservation, demand response and anomaly detection, but, lacks a formal rule mining mechanism and fails to consider inter-appliance association at a higher degree. The work presented by [2] and [24] used sequential pattern mining to understand the electrical appliance usage patterns with a goal to conserve energy. Similarly, authors in [25] proposed an algorithm to mine probabilistic correlation patterns among appliances, which is extended in [26] through incremental sequential mining to discover correlation patterns among appliances using a new algorithm offering memory reduction with improved performance. Authors in [27] further extended their approach to mine time interval data for probabilistic temporal patterns through discovering appliance usage patterns. The study by [28] proposed a pattern recognition technique to identify ADL from electrical appliance use but uses only one appliance; this work was extended in [29] to identify five concurrent operating electrical appliances using real power consumption of appliances. Context-aware association mining through frequent pattern recognition was studied in [30], where the aim is to discover consumption and operation patterns, and effectively regulate power consumption to save energy.

The study by [31] used frequent pattern mining through Apriori approach to determine current activity, and reveal irregularities in power consumption behavior. Paper [32] utilized sequential association rule mining in conjunction with hierarchical clustering to extract the appliance associations with time and consumers activities to forecast energy consumption. The work in [33] proposed a new algorithm to consider the incremental generation of data and mining appliance associations incrementally. Similarly in [34], appliance association and sequential rule mining were studied to generate and define energy demand patterns. Authors in [35] suggested an auto-regression model to compute energy consumption profile for residential consumers to facilitate energy savings recommendations, but do not consider consumers behavioral attributes. The methodology proposed by [36] uses a two-step clustering approach to analyzing load shapes and propose segmentation schemes with selection strategies for energy programs and related pricing or marketing. Authors in [37] used sequential pattern (activity) mining to predict future activities enabling power consumption forecast, and present short term load forecasting using activity sequence patterns with Support Vector Regression (SVR).

The authors in [38] suggested a clustering approach to identify the distribution of consumers' temporal consumption patterns, however, the study does not consider appliance level usage details, which are a direct reflection of consumers' comfort and does not provide a correlation between generated rules and energy consumption characterization. Study [39] proposed k-means along with SOM to compute baseline load estimate by clustering similar load patterns for demand response purpose. The study in [40] uses clustering as a means to group customers according to load consumption patterns to improvise on load forecasting at the system level. Similarly, the authors in [41] used k-means clustering to discover consumers' segmentation and use socio-economic inputs with Support Vector Machine (SVM) to predict load profiles towards demand side management planning. The work in [32] proposed a methodology to disclose usage pattern using hierarchical and c-means clustering, multidimensional scaling, grade data analysis and sequential association rule mining; while considering appliances' ON and OFF events; but, the study does not consider the duration of appliance usage or the expected variations in the sequence of appliance usage. However, authors in [42] and [43] suggested filtering techniques that can be utilized to address the uncertainty in such models, but needs to be explored.

The above approaches do not take into account human behavioral variations, such as the high uncertainty in the order of use of appliances to complete an activity, or variations, such as increased or reduced frequency of use of appliances due to seasons. These dissimilarities have a direct influence on energy consumption patterns, which results in increased number of patterns to be analyzed along with different interpretation of the same event/activity at different occasions. Also, it is very critical to respect consumers comfort level to buy-in their confidence, which is one main ingredient to the success of any energy-savings or related programs. We address these shortcomings by adapting incremental progressive unsupervised machine learning through frequent pattern mining, translating energy consumption patterns into frequent patterns representing inter-appliance associations, as an illustration of a consumer's behavior and expected comfort, while explaining home energy consumption.

III. PROPOSED MODEL

This section covers proposed a model to extract critical information from data in the form of frequent patterns and association rules, defining inter-appliance association/correlation, and appliance clusters over time for the appliance-time association through incremental and progressive data mining techniques. Figure 1 represents our proposed model with its distinct four phases: data preparation, frequent pattern

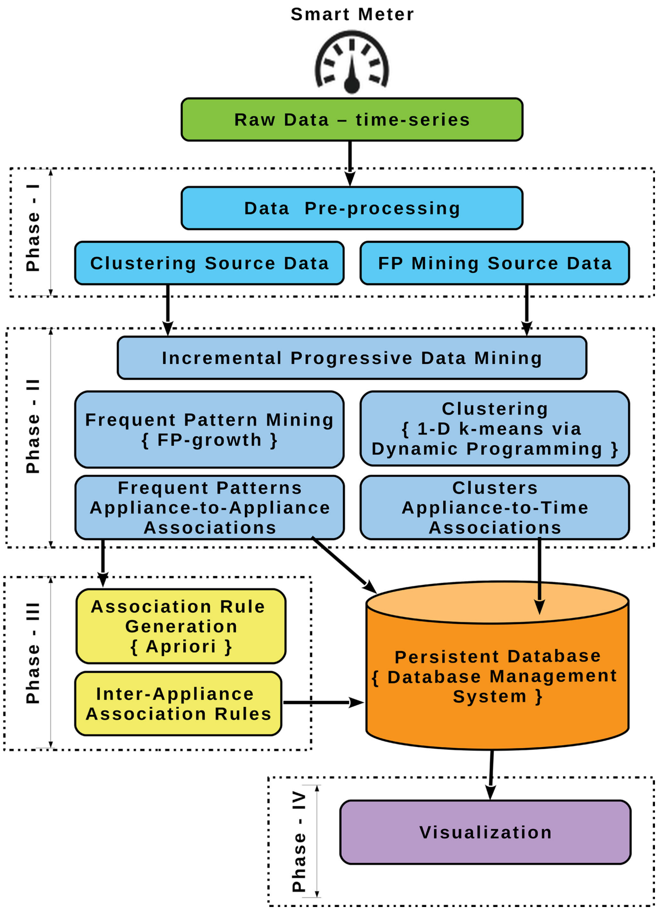

FIGURE 1. Model: Incremental progressive data mining-frequent pattern & association rules, and clustering.

mining and clustering, association rules generation, and visualization. In this section, these phases are discussed and details about the proposed mechanism along with related theoretical background are provided.

A. DATA PREPARATION

Smart meters time-series raw data, which is a high time-resolution data, is transformed into 1 min resolution load data; subsequently translated into a 30 minutes time-resolution source data; i.e., 24 * 2 = 48 readings per day, while recording usage duration, average load, and energy consumption for each active appliance. All the appliances registered active during the 30 minutes time interval are included into the source database for frequent pattern data mining. The dataset [8] has over 400 million raw records of five houses with a time resolution of 6 seconds. Which is reduced to 20 million during preprocessing without loss of accuracy or precision. Table 1, shows an example of the resulting ready to mine source data format comprising four appliances from one house.

B. FREQUENT PATTERN MINING

Frequent patterns are repeated patterns or itemsets, which often appear in a dataset. Considering smart meter data, an itemset could comprise, for example, of laptop and washing machine, that often present themselves together in a frequent pattern. Hence, frequent pattern mining can help discover association and/or correlation among appliances, which defines the relationship among data interpreting consumer energy consumption behavior.

TABLE 1. Frequent Pattern Source Database.

Start Time	End Time	Active Appliances
2013-08-01 07:00	2013-08-01 07:30	'2 3 4 12'
2013-08-01 07:30	2013-08-01 08:00	'3 4 12'
2013-08-01 08:00	2013-08-01 08:30	'2 4 12'
2013-08-01 08:30	2013-08-01 09:00	'4 12'
2013-08-01 09:00	2013-08-01 09:30	'2 3 12'
2013-08-01 09:30	2013-08-01 10:00	'2 3 4'
2013-08-01 10:00	2013-08-01 10:30	'2'
2013-08-01 10:30	2013-08-01 11:00	'12'
2013-08-01 11:00	2013-08-01 11:30	'2 12'
2013-08-01 11:30	2013-08-01 12:00	'3 12'
2013-08-01 12:30	2013-08-01 13:00	'2 4'

2 = Laptop, 3 = Monitor, 4 = Speakers, 12 = Washing Machine

C. FREQUENT ITEMSETS AND ASSOCIATION RULES

In this section, an introduction on the preliminary background on frequent pattern mining based on [44] is presented. Let $\Gamma = \{I_1, I_2, \ldots, I_k\}$ be an itemset containing k items (appliances), which is referred to as k-itemset (l_k). Let *DB*, represent a transaction database with a set of transactions as described in Table 1, where each transaction Υ is an itemset having $\Upsilon \subseteq \Gamma$ and $\Upsilon \neq \emptyset$. The frequency of appearance of an itemset is the number of transactions that contain the itemset, defined as the support count, or the count of the itemset. Let, X and Y be set of items, such that $X \subseteq \Upsilon$ and $Y \subseteq \Upsilon$. Itemsets X and Y are considered frequent itemsets or patterns, if their respective *support* s_X and s_Y, the percentage of transactions the itemset appears in the transaction database *DB*, are greater than or equal to *minsup*; where *minsup* is the pre-defined minimum support threshold. *support* can be viewed as the probability of the itemset in the transaction database *DB*. This is referred to as the relative support, whereas the frequency of occurrence is known as the absolute support. Hence, if the relative support of an itemset $X(s_X)$ [or $Y(s_Y)$] satisfies a pre-defined minimum support threshold *minsup*, then the absolute support of X (or Y) satisfies the corresponding minimum support count threshold.

Association rules are the results of the second iteration of the frequent pattern mining process, where already discovered frequent itemsets/patterns are processed to generate association rules. Rules, of form $\{X \Rightarrow Y\}$, are generated using *support-confidence* framework, where *support* $s_{X\Rightarrow Y}$ [Equation (1)] is the percentage of transactions containing $(X \cup Y)$ in transaction database *DB*, which also can be seen as the probability $P(X \cup Y)$. The *confidence* $c_{X\Rightarrow Y}$ [Equation (3)] is defined as the percentage of transactions in *DB* containing X that also contain Y, which is the conditional probability, $P(Y|X)$ [44]. Equations (1) and (3) capture the above notions respectively

$$support(X \Rightarrow Y) = s_{X\Rightarrow Y} = P(X \cup Y) = support(X \cup Y) \tag{1}$$

$$absolute\_support(X \Rightarrow Y) = support\_count(X \cup Y) \tag{2}$$

$$confidence(X \Rightarrow Y) = P(Y|X) = \frac{support(X \cup Y)}{support(X)} = \frac{support\_count(X \cup Y)}{support\_count(X)}. \tag{3}$$

Hence, an association rule established as $\{X \Rightarrow Y\}$, having $X \subset \Gamma$, $Y \subset \Gamma$, $X \cap Y = \emptyset$, $X \neq \emptyset$, and $Y \neq \emptyset$, with *support* $s_{X\Rightarrow Y} \geq minsup$ and *confidence* $c_{X\Rightarrow Y} \geq minconf$ is classified as strong, where *minconf* is a pre-defined minimum confidence threshold. Additionally, the association rule's support $s_{X\Rightarrow Y}$ will automatically satisfy the minimum support threshold as the rules are essentially generated from frequent patterns X, and Y having respective *support* $s_X, s_Y \geq minsup$. Thus, once the *support* for X, Y, and $(X \cup Y)$ are determined, corresponding association rules $\{X \Rightarrow Y\}$ and $\{Y \Rightarrow X\}$ can be extracted, which satisfies *minsup* and *minconf*; i.e., the association rule generation process can be deduced to a two-step operation; first, frequent pattern mining, and second, generating strong association rules of interest [44].

D. DISCOVERING INTER-APPLIANCE ASSOCIATIONS

In this section, the proposed approach towards incremental progressive frequent pattern mining along with additional interestingness measures for the discovery of correlation among appliances is discussed.

The mining of frequent patterns is generally considered as an off-line and costly process on large databases. In a real world application, transaction data generation is a continuous process, where new transactions are generated and old transactions may become obsolete as the time progresses, thereby, invalidating existing frequent patterns and/or establishing new frequent pattern and associations. Therefore, an incremental and progressive update strategy is imperative, where these variations/updates are taken into account and the discovered frequent patterns are duly maintained. For example, an appliance such as a room heater generally will be used during winter and one can expect reduced usage frequency during other seasons. As an effect, a significant gain during winter but decrease during other seasons will be registered. As a result, room-heater should appear higher on the list of frequent patterns and association rules during winter, but much lower during summer or spring. This objective can be achieved through progressive incremental data mining while eliminating the need to re-mine the entire database at regular intervals. Frequent pattern mining in a large database can be accomplished through pattern growth approach [45], [46]. We extend this pattern growth approach and present an incremental frequent pattern mining strategy of a progressive manner, which is discussed next.

1) FP-GROWTH : A PATTERN-GROWTH APPROACH, WITHOUT CANDIDATE GENERATION FOR MINING FREQUENT ITEMSETS

Apriori [47] algorithm with candidate generation can suffer from the following problems:

- Breadth-first approach; i.e., level-wise search.
- Generate a large number of candidate sets.
- Repeatedly search through the entire database to find support for an itemset.

To overcome these deficiencies, the work in [45] and [46] propose a pattern growth or FP-growth approach, which exploits the depth-first divide-and-conquer technique. To start with, it generates a compact representation of the transactions from the database in the form of a frequent pattern tree or FP-tree. FP-tree preserves the association information, derived from each individual transaction, along with the support count for each constituent item. Next, *conditional databases*(*tree*) for each frequent item is extracted from FP-Tree to mine frequent patterns, which the item under consideration is part of. This way, only the divided portion relevant to the item and its associated growing patterns are inspected, while addressing the shortcomings of the Apriori [47] approach.

2) INCREMENTAL FREQUENT PATTERN EXTRACTION

Our proposed technique exploits the benefits of pattern growth strategy and extends it to achieve incremental progressive mining of frequent patterns by mining in a quantum of 24 hours; i.e., frequent patterns are extracted from data comprising of appliance usage tuples for a 24 hour period, in a progressive manner. With this approach, we mine only a portion of the entire database at each iteration thus reducing the memory overhead for the FP-growth strategy and achieve improved efficiency.

In our proposed approach, available data is recursively mined in quanta of 24 hours, and a frequent patterns' discovered database, represented in Table 3, is maintained across successive mining exercises. In other words, data mining can be viewed as a process conducted at the end of each day in an incremental manner. During each consecutive mining operation, the support count and database size for the existing frequent patterns are incremented and new patterns, with applicable support count and database size, are added to the persistent database. Moreover, we cease the use of the minimum support threshold *minsup* at the mining stage to eliminate any candidate patterns, resulting in the discovery of all the possible frequent patterns. This change in technique is incorporated to avoid missing the candidate patterns, which can become frequent if the time quantum is increased or the complete database is mined in a single operation. At the end of the mining process, *database_size* is updated for all of the frequent patterns in the frequent patterns discovered database Table 3 to ensure the correct computation of *support*.

Frequent patterns discovered database Table 3, can be maintained in-memory using hash table data structure or off the memory in a Database Management System. The latter approach reduces memory requirements at the cost of a

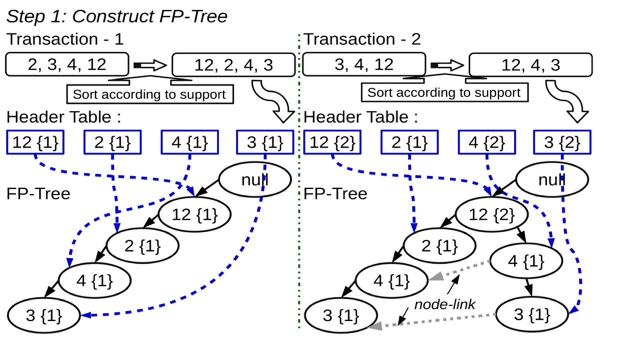

FIGURE 2. **Step 1-FP-Tree construction : Scanning database and adding transactions.**

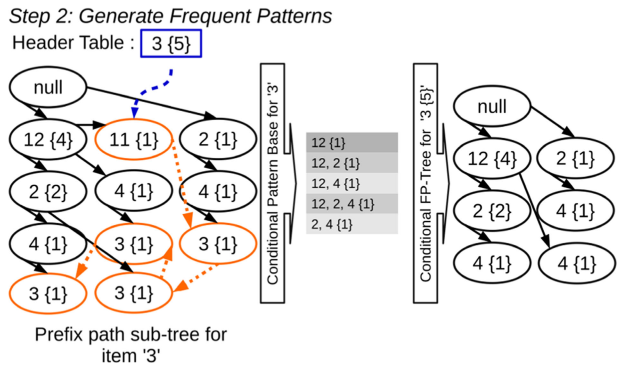

FIGURE 4. **Step 2-Generating frequent pattern : Conditional FP-Tree.**

marginal increase in processing time, whereas the former approach reduces processing time but requires more memory. In the smart meter environment, although quicker processing time is of importance, the persistence of the information discovered through days, months or years is more vital to achieving useful results for the future. Therefore, in this paper, a permanent storage using a Database Management System is used over in-memory volatile storage.

Algorithm 1. Incremental Frequent Pattern Mining

Require: Transaction database (*DB*)
Ensure: Incremental discovery of frequent patterns, stored in frequent patterns discovered database (*FP_DB*)
1: **for all** Transaction data slice db_{24} in quanta of 24 hours in database *DB* **do** {Data is processed in slices of 24 hour period}
2: Determine database size $Database\_Size_{db_{24}}$ for data slice db_{24}
3: Construct Frequent Pattern Tree (FP-Tree) {As described in Step-1 Constructing Frequent Pattern Tree (FP-Tree)}
4: Generate Frequent Patterns, while calling function *save_ update_frequent_pattern* to save frequent patterns to frequent patterns discovered database *FP_DB* {As described in Step-2 Generating Frequent Patterns}
5: **end for**
6: For all Frequent Patterns in Database *FP_DB* increment *Database Size* by $Database\_Size_{db_{24}}$

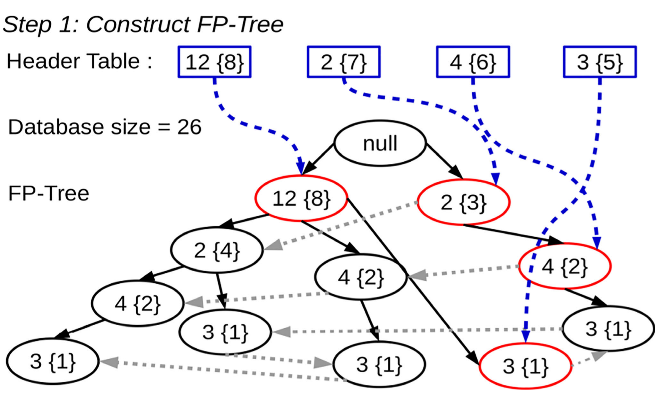

FIGURE 3. **Step 1-FP-Tree construction : Final frequent pattern tree.**

The proposed progressive incremental data mining approach is outlined by the Algorithm 1; which extends the two-step process of frequent pattern mining using FP-growth. Step 1; i.e., FP-Tree construction, is described in Algorithm 3 and Figures 2 and 3. Step 2; i.e., frequent pattern generation, is covered in Algorithm 5 and Figures 4 and 5. Algorithm 2 explains the mechanism to achieve persistent storage of frequent patterns discovered by the mining process into a permanent storage such as a Data Base Management System.

Two step process for frequent pattern mining based on FP-growth:

Step-1 Constructing Frequent Pattern Tree (FP-Tree). It takes two scans of the transaction database (lines 1 and 4 in Algorithm 3; first to create the list of 1-itemset (an itemset comprising only one item) frequent itemsets with support [presented in Table 2, sorted in decreasing order of support], and second to construct the FP-Tree [45], [46]. In our setting, we do not eliminate 1-itemset frequent itemsets based on minimum support threshold (*minsup*) for the reasons discussed earlier, which is a departure from the original algorithm proposed by [45], [46].

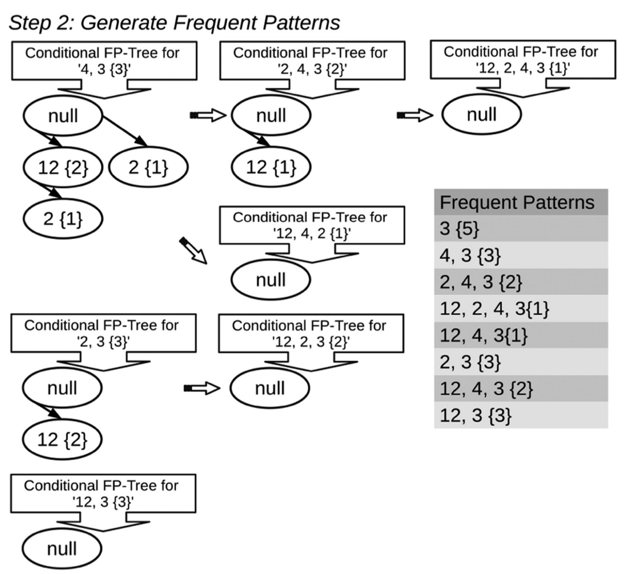

FIGURE 5. **Step 2-Generating frequent pattern : Recursive mining.**

TABLE 2. List: 1-Itemset Frequent Itemsets with Support.

1-itemset	12	2	4	3
support	8	7	6	5

2 = Laptop, 3 = Monitor, 4 = Speakers, 12 = Washing Machine

Algorithm 2. Function *save_update_frequent_pattern*

Require: Frequent Pattern extracted *FP_extracted* (New), support count *absolute_support* for *FP_extracted*, Frequent pattern discovered database *FP_DB*

Ensure: Add or update Frequent Pattern in frequent patterns discovered database

Search a frequent pattern ($FP = FP\_extracted$) in *FP_DB*
if Frequent Pattern found **then**
  Increment support count for *FP* by *absolute_support*.
else
  Add a new Frequent Pattern with support count *absolute_support* and *Database_size* = 0.
end if

Algorithm 3. Step-1 Constructing FP-Tree

Require: Given transaction database *DB*

Ensure: FP-Tree T

1: Scan *DB*, generate list F with all the 1-itemset frequent items, and determine support of each frequent item. {Database Scan 1: create list of 1-itemset frequent itemsets with support}
2: Sort F in descending order of support.
3: Create FP-Tree root $T \leftarrow null$.
4: **for all** Transactions (Tr) $\in DB$ **do** {Database Scan 2: create FP-Tree}
5:   Sort items in Tr according to order of F.
6:   Item list of pattern $[P = p|P']$, where p is first element and P' remaining list
7:   Call Function $insert\_tree([P = p|P'], T)$.
8: **end for**

FP-Tree is constructed by reading one transaction at a time (lines 4 to 6 in Algorithm 3) from the Frequent Pattern Source Database [Table 1], sorting the items according to the list of 1-itemsets [Table 2] in decreasing order of global support, and mapping it to a path in the FP-Tree. The fixed order of the items ensures an overlap of the path for the transactions having an identical prefix; i.e., sharing items. Further, on the addition of a new transaction, the support count is incremented for each item node in the prefix path shared among transactions as explained in Algorithm 4. A header table comprising of 1-itemsets, sorted in decreasing order of global support, is maintained throughout the complete process. This table stores pointers to the last items added to the tree of particular 1-itemset. Additionally, a *node-link* is added from the recently added item node to the preceding item node of the same 1-itemset type. A *node-link* facilitates traversal (trace) of all the 1-itemsets in FP-Tree for a specific item. Therefore, a FP-Tree is a compact representation of the database which preserves the critical information of *absolute support* for each item/itemset and transaction patterns. This process is depicted in Figure 4 by the addition of two transactions to the FP-Tree. Figure 5 presents the final FP-Tree constructed from Frequent Pattern Source Database [Table 1].

Algorithm 4. Function $insert\_tree([p|P'], T)$

Require: Frequent item list of pattern $[P = p|P']$ where p = first element and P' = remaining list, and FP-Tree T.

Ensure: Add items from an item list to FP-Tree T.

Initialize current root node $R_C \leftarrow root\_node$ (FP-Tree).
for all Item/element $item_i \in P$ **do**
  Search for a node N in T, having $N = item_i$.
  if Node found **then**
    Increment support count for N by 1 (one).
    $R_C \leftarrow N$ {Capture as current root}
  else
    Create new node N, with support count 1 (one); where parent-link linked to R_C, and node-link linked to nodes with same item.
    $R_C \leftarrow N$ {Capture as current root}
  end if
end for

Step-2 Generating Frequent Patterns. Once the FP-Tree is created, a bottom-up recursive elimination approach making use of the divide and conquer scheme is employed to generate a complete set of frequent patterns from the FP-Tree, this is explained in Algorithm 5. FP-tree mining is accomplished by starting from each frequent length-1 pattern (as an initial suffix pattern) and constructing its' *conditional-pattern-base* from prefix paths extracted from the FP-tree, where suffix co-occur with a prefix. A *conditional-pattern-base* can be considered as a collection of transactions containing a particular itemset, but removing it from transactions. The header table, as shown in Figure 3 acts as the source for 1-itemsets; which are ordered in the increasing order of global support, so the process starts from the leaf (bottom) nodes of the tree and traverses towards the root (up). Later, *conditional-FP-tree*'s are created from the *conditional-pattern-base* and mined recursively to extract frequent patterns until the resulting tree is empty or comprises a single path. Lastly, frequent patterns from single paths are derived by producing all of the combinations of the sub-paths. The pattern growth is accomplished through the concatenation of the suffix pattern with the frequent patterns generated from a conditional FP-tree. The *node-link* enables extraction of *conditional-pattern-base* and creation of *conditional-FP-tree* by aiding the trace of nodes for a given 1-itemset suffix pattern, which is frequent. Figures 4 and 5 illustrate the procedure for the extraction of frequent patterns.

For the purpose of making the frequent patterns available for future manipulation and utilization, we store all the frequent patterns extracted into a Database Management System as explained in Algorithm 2.

Algorithm 5. Step-2: FP-Growth: Generating Frequent Patterns

Require: FP-Tree T, Current itemset suffix S.
Ensure: Frequent Patterns
if T is a single path **then** {Mine single path FP-Tree for frequent patterns}
$FP\_single\_path \leftarrow null$
  for all Combination C of nodes in T **do** {support(C) = minimum support of nodes C}
    Generate frequent patterns $FP_S = C \cup S$ {Algorithm 3}
    Update Frequent Pattern Discovered Database FP_DB {Algorithm 2}
    $FP\_single\_path = FP\_single\_path \cup FP_S$
  end for
  FP_single_path represents Frequent patterns generated
else
  for all $item_i$ of nodes in T **do** {T is multipath tree Mine multipath FP-Tree for frequent patterns}
    Generate frequent pattern $FP_M = item_i \cup S$ {Algorithm (refalgo:FP-Tree)}
    Update Frequent Pattern Discovered Database FP_DB {Algorithm 2}
    $FP\_multipath = FP\_multipath \cup FP_M$
    Determine itemset suffix $S_i = item_i \cup S$
    Extract conditional prefix path or conditional pattern-base for $item_i$ by using node-link and parent-link.
    Generate conditional FP-Tree T_i from conditional prefix path or conditional pattern-base.
    if FP-Tree $T_i \neq \emptyset$ **then**
      Call FP-growth(FP-Tree T_i, Current itemset suffix S_i)
    end if
  end for
  $FP\_multipath$ represents Frequent patterns generated
end if
Final set of frequent patterns generated = $FP\_single\_path \cup FP\_multipath$

The results of frequent pattern mining are represented in the Table 3. Thus, a frequent pattern of this form, for example, "Laptop, Monitor, Speakers", can be interpreted to represent association among appliances or inter-appliance associations. The respective probabilistic evidence, supporting the occurrence of these association, can be computed as shown in Equations (1) and (2).

E. ASSOCIATION RULES GENERATION USING CORRELATION ANALYSIS

In a large database, the number of association rules generated can be very large. Although, it is entirely dependent on the application of results, but the reduction in numbers of association rules can help narrow down the search space for the useful and the strong association rules. This can be achieved through the application of statistical interestingness measures. In general, FP-growth [45], [46] and Apriori [47] algorithms use *support-confidence* framework to generate frequent patterns and extract association rules, while eliminating uninteresting rules by comparing *support* and *confidence* with *minsup* and

TABLE 3. Frequent Patterns: Frequent Patterns Discovered Database.

Frequent Pattern	Absolute Support	Database Size
'2 3'	3,939	7,899
'2 4'	2,840	7,899
'2 3 4'	2,649	7,899
'3 4'	2,649	7,899
'2 3 4 12'	1,299	7,899

2 = Laptop, 3 = Monitor, 4 = Speakers, 12 = Washing Machine

minconf respectively. However, *support* and *confidence* do not evaluate correlation of the rule's *antecedent* and *consequent*. This turns out to be less effective in eliminating uninteresting association rules. Therefore, it is important to learn correlation relationship among the rules constituents to determine the positive or negative impact of one's presence over other and remove the rules which are not of interest. A measure of correlation such as *Lift*, *Kulc* and/or *IR* can help and supplement the *support-confidence* framework and provide more insight into the association relationship [44]. Consequently, the correlation rule can be expressed as

$$X \Rightarrow Y[support, confidence, correlation]. \tag{4}$$

Lift: measures dependency and correlation of rule's *antecedent* and *consequent*; *lift* is define as

$$lift(X,Y) = \frac{P(X \cup Y)}{P(X).P(Y)} = \frac{P(Y|X)}{P(Y)} = \frac{confidence(X \Rightarrow Y)}{support(Y)}. \tag{5}$$

Where, if $Lift < 1.0$, then X is negatively correlated to Y; i.e., the occurrence of X indicates the absence Y or vice-versa. If $Lift > 1.0$, then X and Y are positively correlated; i.e., occurrence of X indicates presence of Y or vice-versa; whereas if $Lift = 1.0$, then X and Y are independent with no correlation among them.

Ironically, this commonly used correlation measure *Lift* is affected by null-transactions. Null-transaction, are the transactions where itemsets under consideration are not part of it ($X \in$ null-transaction *or* $Y \in$ null-transaction), and in a large database, null-transactions can outbalance the support count for itemsets. Hence, this approach fails when contemplating a low minimum support threshold or searching for extended patterns as explained by the study in [44]. This study suggests using null-invariant interestingness measures of the Kulczynski measure (*Kulc*) along with the Imbalance Ratio to supplement *support-confidence*/*lift* framework to extract more interesting rules.

Kulczynski Measure (Kulc) [44]: *Kulc* of X and Y, is an average of the *confidence* measures for X and Y, which, by definition of the *confidence* can be translated into the average of their conditional probabilities. *Kulc* measure is null-invariant and is defined as

$$Kulc(X,Y) = \frac{1}{2}(P(X|Y) + P(Y|X)). \tag{6}$$

Where, if $Kulc = 0.0$, then X is negatively correlated to Y; i.e., the occurrence of X indicates the absence of Y or vice-versa. If $Kulc = 1.0$, then X and Y are positively correlated; i.e., occurrence of X indicates presence Y or vice-versa, whereas $Kulc = 0.50$ indicates X and Y are independent having no correlation.

Imbalance Ratio [44]: *IR* measures the imbalance of *antecedent* and *consequent* for the rule. It is defined as

$$IR(X,Y) = \frac{|s_X - s_Y|}{s_X + s_Y - s_{(X \cup Y)}}. \tag{7}$$

Where $IR = 0.0$ and $IR = 1.0$ represent perfectly balanced and very skewed scenario respectively. Imbalance ratio is null-invariant and is not influenced by the database size.

Association Rule Generation: It is an effortless process to extract the association rules from frequent itemsets discovered from the transactions in a database *DB*. Association rules can be derived, as explained in Algorithm 6, where we introduce the use of the correlation measures of *Kulc* to extend the Apriori [47] approach in order to filter out uninteresting association rules along with the measure of imbalance ratio *IR* to explain it.

Algorithm 6. Apriori Association Rule Generation

Require: Frequent patterns discovered database *FP_DB*, minimum support *minsup*, minimum confidence *minconf*, minimum Kulczynski Measure *minkulc*
Ensure: Association Rules

```
for all Frequent itemset FP_i in FP_DB do
  Generate all subsets subset_{FP_i} of FP_i, having subset_{FP_i} ≠ ∅
  for all Subset subset_{FP_ij} in subset_{FP_i} do
    X ← subset_{FP_ij}
    Y ← (FP_i − subset_{FP_ij})
    support(s_{X⇒Y}) ← support(X ∪ Y)
    confidence(c_{X⇒Y}) ← support(X∪Y) / support(X)
    Kulc ← 1/2 {confidence(X ⇒ Y) + confidence(Y ⇒ X)}
    if support(s_{X⇒Y}) ≥ minsup AND confidence(c_{X⇒Y})
       ≥ minconf AND Kulc ≥ minkulc then
      Output rule: X ⇒ Y As subset_{FP_ij} ⇒ (FP_i − subset_{FP_ij})
    end if
  end for
end for
```

F. APPLIANCE-TIME ASSOCIATION DISCOVERY THROUGH CLUSTERING ANALYSIS: INCREMENTAL K-MEANS

Apart from learning the inter-appliance association, it is of critical interest to understand the appliance usage time with respect to hour of day (00:00-23:59), time of day (Morning, Afternoon, Evening, Night), weekday, week, month and/or season (winter, summer/spring, fall). This underlying information in the smart meter time series data facilitates discovery of the appliance-time associations; which is critical to analyze consumers energy consumption behavior. These appliance-time associations can identify peak load/energy consumption hours and/or explain the behavioral characteristics of the consumers. Appliance-time associations can be considered as the grouping of sufficiently close time-stamps, when the relevant appliance has been recorded as active or operational, to form a class or cluster for a given appliance. The clusters or classes constructed will describe appliance-to-time associations, while the respective size of clusters, defined as the count of members in the cluster, will establish relative strength for clusters. Therefore, the discovery of appliance-time associations can be translated into clustering of time-stamps for appliances into brackets of time-spans/slices; where each cluster belongs an appliance with respective times-stamps (data points) as members of the cluster.

Cluster or clustering analysis is the process of creating classes (unsupervised classification) or groups/segments (automatic segmentation) or partitions or subsets among a set of data called clusters. The members of a cluster must possess similarity with one another, but should be dissimilar from the members of the other clusters. The distinct advantage of the clustering analysis is the non-supervised nature of the process [48]. The proposed approach exploits this advantage to discover appliance-time associations. We choose a 30-minute time-span/slice, for cluster segmentation, which will sufficiently capture the associations while minimizing the number of segments created; i.e., creating maximum 48 clusters for a day, whereas other clustering bases such as time-of-day, weekday, week, month and seasons have natural segmentation. Algorithm 7 extends the k-means cluster analysis by the dynamic programming algorithm, proposed in [49], to achieve incremental data mining for the discovery of the appliance-time associations. And, make use of the *silhouette score* calculated based on the *eluclidean* distance to determine the optimal number of the clusters; i.e., k. The proposed algorithm in [49] ensures optimality and quick runtime. Table 4 shows the sample results of cluster analysis.

TABLE 4. Cluster Analysis: Clusters Discovered Database.

Appliance	Cluster ID	Size	Centroid	SSE	Distance From Centroid
2	1	113	630	0	0
2	2	118	660	0	0
:	:	:	:	:	:
2	14	154	1,020	0	0
2	15	151	1,050	0	0
:	:	:	:	:	:
2	40	10	1,380	0	0
2	41	8	1,410	0	0
:	:	:	:	:	:
2	48	51	150	0	0

2 = Laptop

TABLE 5. Raw Data: Sample.

Timestamp	Appliance power consumption
1404249794	88
1404249856	99
1404249917	99
1404249978	98

Recorded at appliance level using plug-in individual appliance monitors (IAMs).

Algorithm 7. Incremental Clustering: k-Means

Require: Transaction database *DB*, permissible time distance between clusters *time-span*/*slice*
Ensure: Incremental clustering, clusters and related configuration stored in discovered clusters database *CL_DB*

for all Transaction data slice db_{24} in quanta of 24 hours in database *DB* **do** {Data is processed in slices of 24 hour period}
  Determine optimal k for data slice db_{24} by computing *silhouette coefficient*(*width*) for clustering
  Discover k clusters CL_{24} in db_{24} using one-dimension k-means clustering via dynamic programming, while capturing SSE, Silhouette coefficient (width), and data points with distance from centroid
  for all Cluster c in CL_{24} **do**
    Search for closest cluster c_{ts} in *CL_DB*, with cluster center within a distance of *time-span*/*slice*
    if Cluster found **then**
      Merge clusters c_{ts} and c while evaluating quality of cluster by computing *silhouette coefficient*(*width*) and save to database
    else
      Add new cluster c with respective parameters and information
    end if
  end for
end for

IV. DATASET AND SYSTEM SETUP

We make use of the context-rich UK Domestic Appliance Level Electricity dataset [8], which includes time series data of power consumption collected between 2012 and 2015. The dataset contains time series data for five houses with a total of 109 appliances, having a time resolution of 6 seconds, from Southern England published by UK Energy Research Centre Energy Data Centre (UKERC-EDC). Energy consumption was conducted at appliance level using plug-in individual appliance monitors (IAMs) [8]. A sample of raw data is presented in Table 5. Additionally, we generated a synthetic dataset, having over 1.2 million raw energy consumption data points with a time resolution of 1 minute from the smart meter of one house with 21 appliances to conduct initial experiments.

The underlying system for the proposed model is developed in Python, and the data is stored in MySQL and MongoDB databases on Ubuntu 14.04 LTS 64-bit system.

TABLE 6. House 2 : Appliances Associations.

Appliances ↓	Support (%)			
Time Period	→ 7 days	30 days	25 %	Full
Laptop, Monitor, Speakers	66.38	51.22	29.33	33.54
Laptop, Monitor	82.53	70.93	38.83	49.87
Monitor, Speakers	66.38	51.22	29.33	33.54
Laptop, Speakers	66.38	51.98	30.28	35.95
Laptop, Monitor, Washing Machine				22.09
Monitor, Washing Machine				22.48
Laptop, Washing Machine				24.19
Speakers, Washing Machine				26.04

Note: Incremental progressive inter-appliance association discovery with minsup ≥ 0.2.

V. EVALUATION RESULTS

We carried out exhaustive incremental frequent pattern mining on the energy consumption data from five houses of the dataset UK-Dale [8] along with the synthetic dataset to examine intermediate and final results. The outcome of the evaluation comprehensively supports our hypothesis of the undeviating influence of human behavior over energy consumption patterns at a household, which can be learned from inter-appliance associations. Moreover, we conducted a comprehensive analysis of power load and energy consumption patterns to verify and explain our results. Due to space restrictions, we present results from three houses while augmenting our argument that consumer behavior affects the household energy usage.

Table 6, exhibits the incremental appliance-appliance association discovery and formation. We noted that the association relationships change over a period of time, which are directly affected by consumers' behavior while revealing personal preferences depicting expected comfort. Also, time: season, month, week, weekday, time of day and hour of day has strong influence over the consumers' behavior with an immediate impact on energy usage. Therefore, it is critical to learn these variations at regular intervals, most suitably near-real time, to take into account for designing energy programs. Tables 7 and 8 represent extracted appliances associations and appliances usage priority respectively, which constitute the behavioral appliance usage patterns establishing household energy consumption patterns. Appliance usage priorities are derived from the frequency of usage of the respective appliance. Figure 6 presents the visualization of the appliance associations for House 2; a tree structure exhibiting the associative relationship among the appliances with their respective *support* in the database.

With reference to House 2, we observe that the four appliances, Washing Machine, Laptop, Monitor, and Speakers, exhibit strong association rules. Further, energy consumption curve compliments the results of frequent pattern and association rules discovery. It can be noted that the energy consumption patterns of different appliances (refer the marked portion in energy consumption curves in Figure 13), exhibit close similarities, which is expected. Thus, we can infer that

TABLE 7. Appliance Association Rules.

	Sr.	Association Rule	Support	Confidence	Kulc	IR
House 1	1	Tv ⇒ Amp_Livingroom	0.15	0.96	0.85	0.23
	2	Amp_Livingroom ⇒ Tv	0.15	0.73	0.85	0.23
	3	Lcd_Office ⇒ Laptop	0.14	0.93	0.75	0.38
	4	Tv, Amp_Livingroom ⇒ Subwoofer_Livingroom	0.13	0.87	0.78	0.19
	5	Kitchen_Lights, Subwoofer_Livingroom ⇒ Amp_Livingroom	0.08	0.85	0.62	0.51
	6	Amp_Livingroom, Livingroom_Lamp_Tv ⇒ Subwoofer_Livingroom	0.08	0.98	0.69	0.6
	7	Subwoofer_Livingroom, Livingroom_Lamp_Tv ⇒ Amp_Livingroom	0.08	0.97	0.67	0.62
House 2	1	Monitor ⇒ Laptop	0.50	0.99	0.96	0.06
	2	Laptop ⇒ Monitor	0.50	0.93	0.96	0.06
	3	Speakers ⇒ Laptop	0.36	0.74	0.70	0.08
	4	Monitor, Speakers ⇒ Laptop	0.34	1.00	0.81	0.38
	5	Laptop, Speakers ⇒ Monitor	0.34	0.93	0.80	0.27
	6	Monitor, Washing Machine ⇒ Laptop	0.22	0.98	0.70	0.58
	7	Laptop, Washing Machine ⇒ Monitor	0.22	0.91	0.68	0.50
	8	Laptop, Washing Machine ⇒ Speakers	0.18	0.74	0.56	0.45
	9	Monitor, Speakers, Washing Machine ⇒ Laptop	0.16	1.00	0.65	0.69
	10	Laptop, Speakers, Washing Machine ⇒ Monitor	0.16	0.91	0.62	0.62
	11	Monitor, Washing Machine ⇒ Laptop, Speakers	0.16	0.73	0.59	0.32
House 5	1	Washer_Dryer ⇒ Microwave	0.82	0.92	0.9	0.04
	2	Microwave ⇒ Washer_Dryer	0.82	0.88	0.9	0.04
	3	I7_Desktop ⇒ Washer_Dryer	0.70	0.91	0.85	0.13
	4	Washer_Dryer ⇒ I7_Desktop	0.70	0.79	0.85	0.13
	5	I7_Desktop ⇒ Microwave	0.70	0.91	0.83	0.16
	6	Microwave ⇒ I7_Desktop	0.70	0.75	0.83	0.16
	7	I7_Desktop Microwave ⇒ Washer_Dryer	0.63	0.91	0.81	0.2
	8	I7_Desktop Washer_Dryer ⇒ Microwave	0.63	0.90	0.79	0.23
	9	I7_Desktop ⇒ Microwave Washer_Dryer	0.63	0.83	0.8	0.05
	10	Microwave Washer_Dryer ⇒ I7_Desktop	0.63	0.77	0.8	0.05
	11	Washer_Dryer ⇒ I7_Desktop Microwave	0.63	0.71	0.81	0.2

House 1 : $Kulc \geq 0.60$, $minsup \geq 0.08$, $minconf \geq 0.70$, House 2 : $Kulc \geq 0.55$, $minsup \geq 0.10$, $minconf \geq 0.70$
House 5 : $Kulc \geq 0.75$, $minsup \geq 0.60$, $minconf \geq 0.70$

TABLE 8. Appliance Usage Priority.

	Sr.	Appliance	Relative Support (%)
House 1	1	Kitchen_Lights	42.82
	2	Kitchen_Lamp2	31.10
	3	Laptop	25.56
	4	Amp_Livingroom	21.01
	5	Subwoofer_Livingroom	19.46
	6	Livingroom_S_Lamp	18.08
	7	Kitchen_Phone&Stereo	17.86
	8	LCD_Office	15.44
	9	Livingroom_Lamp_TV	11.08
	10	Office_Lamp2	9.28
	11	Washing_Machine	8.63
	12	Livingroom_S_Lamp2	8.16
	13	Office_Lamp3	7.72
	14	Kettle	5.77
House 2	1	Washing Machine	62.74
	2	Laptop	53.84
	3	Monitor	50.34
	4	Speakers	48.83
	5	Laptop 2	15.43
	6	Running Machine	09.61
	7	Kettle	06.13
House 5	1	Microwave	92.86
	2	Washer_Dryer	88.96
	3	I7_Desktop	76.89
	4	Network_Attached_Storage	55.33
	5	Sky_Hd_Box	55.33
	6	Home_Theatre_Amp	53.96

House 1 : $minsup \geq 0.05$, House 2 : $minsup \geq 0.05$
House 5 : $minsup \geq 0.50$

these appliances were in fact used simultaneously as shown by our data mining outcomes. Similar observations are recorded for other houses as well. Additionally, from the frequent patterns and association rules, we can notice the occupants' behavioral traits. For example, Laptop is used along with Washing Machine; i.e., the occupant likes to work on the computer while washing clothes and listening to music. Also, from appliance usage priority table, Washing Machine was found to be the highest used appliance for the home. Analogous results are obtained for Houses 1 and 5. For House 1 Kitchen_Lights, Subwoofer, Amp, and TV show strong associations and hint that the occupants like to watch TV or listen to music while cooking. The Kitchen_Lights is the most used appliance in the house. Whereas, for House 5 Microwave, Washer_Dryer, and I7_Desktop display strong associations while Microwave is the most frequently used appliance in the house. It can, therefore, be inferred that the occupants at House 5 enjoy working on the computer while cooking and/or washing/drying clothes.

Moreover, appliance usage priorities, seen in the Table 8, can determine the Appliance of Interest: the major energy consuming appliances. AoIs are identified as the appliances most frequently used with higher usage period; therefore, they act as major contributors towards household energy consumption. These AoI appliances might have a small power rating, but contribute larger a portion of energy consumption compared to appliances having a higher power rating.

IEEE TRANSACTIONS ON
EMERGING TOPICS
IN COMPUTING

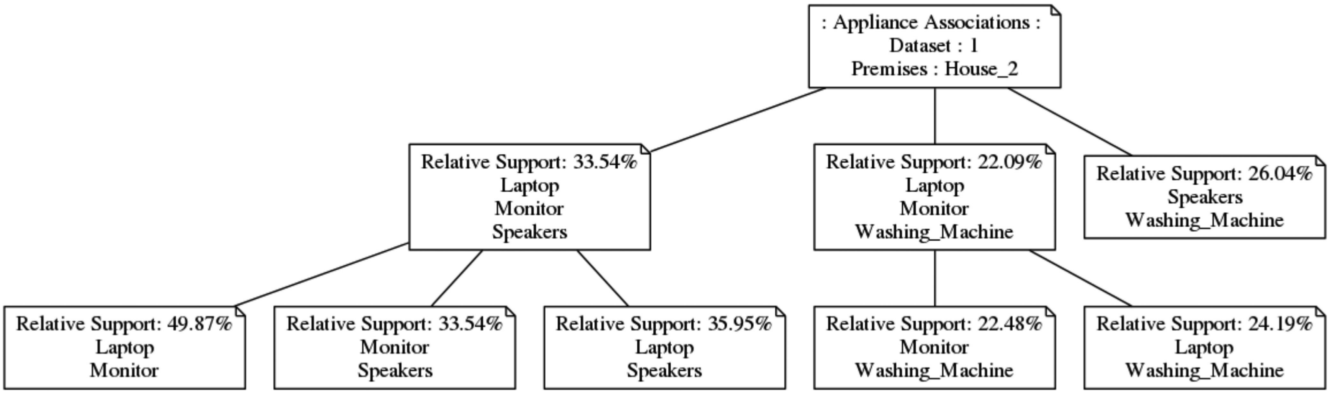

FIGURE 6. **House 2: Appliance associations in full database ($minsup \geq 0.2$).**

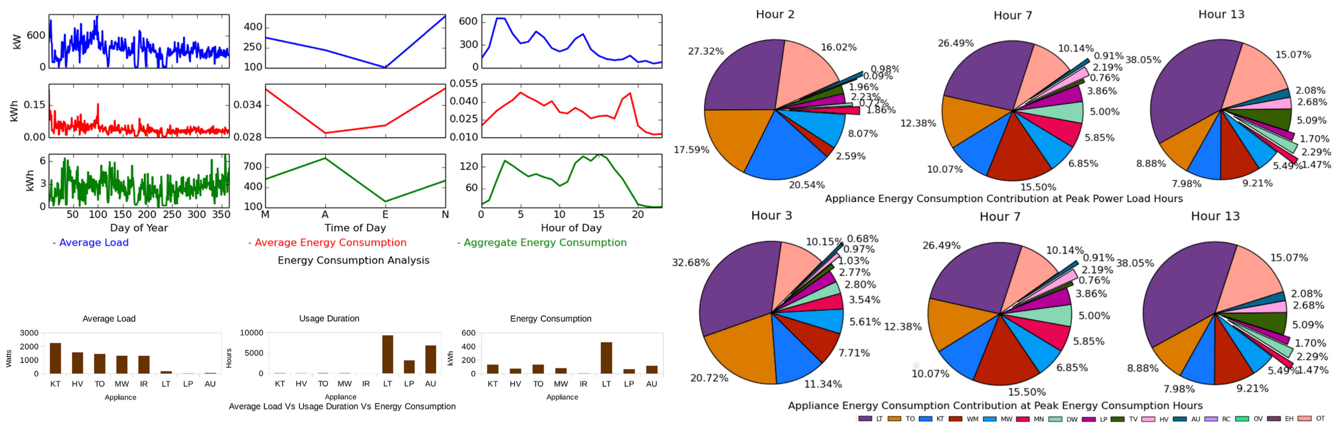

FIGURE 7. **House 1 : Energy consumption analysis [KT-Kettle, HV-Hoover, TO-Toaster, MW-Microwave, IR-Iron, LT-Lights, LP-Laptop or Desktop, AU-Audio/music equipment : Home theater or Amp or Speakers, MN-Monitor, VC-Vaccum Cleaner, EH-Electric Hob, HD-Hairdryer, WM-Machine Machine or Dryer, OV-Oven].**

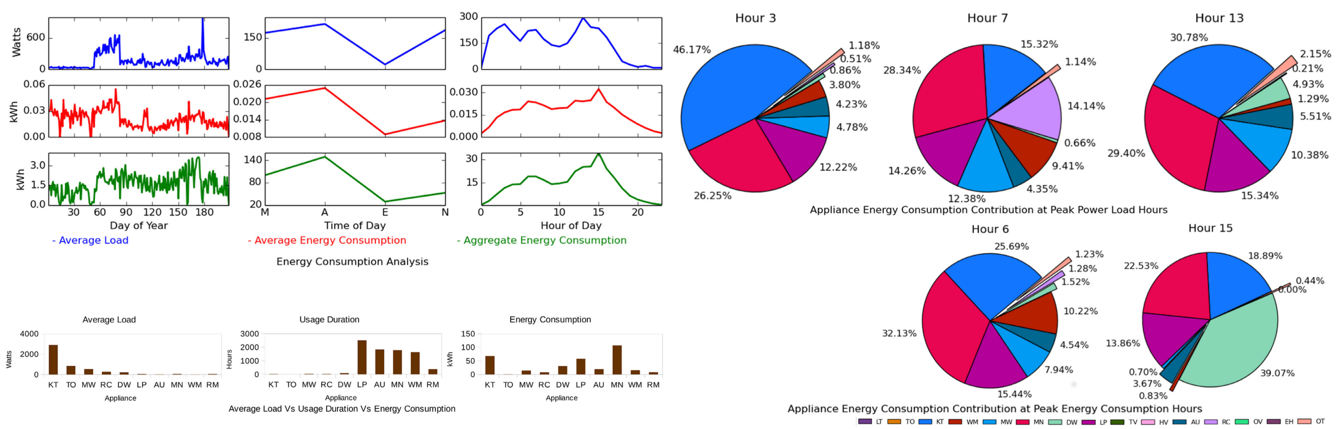

FIGURE 8. **House 2 : Energy consumption analysis [KT-Kettle, HV-Hoover, TO-Toaster, MW-Microwave, IR-Iron, LT-Lights, LP-Laptop or Desktop, AU-Audio/music equipment : Home theater or Amp or Speakers, MN-Monitor, VC-Vaccum Cleaner, EH-Electric Hob, HD-Hairdryer, WM-Machine Machine or Dryer, OV-Oven].**

Additionally, appliances with high power rating might contribute towards peak power load (kW) momentarily, but overall energy (kWh) consumption is noted to be higher for appliances identified as AoI appliances that have short or very short operating time but long usage duration for these group of appliances.

We conducted extensive energy consumption, peak load, peak energy usage, and appliance usage duration analysis to verify our results and found homogeneous trends. Figures 7, 8, and 9, demonstrate the energy consumption analysis conducted for the three houses with trend plots for average load (Watts), average energy consumption (kWh) and aggregate energy

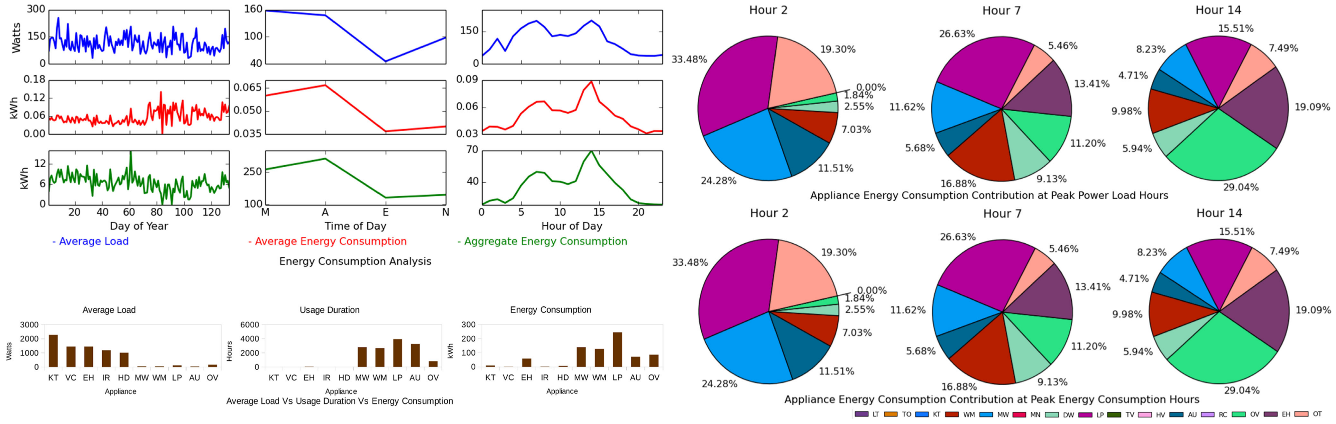

FIGURE 9. House 5 : Energy consumption analysis [KT-Kettle, HV-Hoover, TO-Toaster, MW-Microwave, IR-Iron, LT-Lights, LP-Laptop or Desktop, AU-Audio/music equipment : Home theater or Amp or Speakers, MN-Monitor, VC-Vaccum Cleaner, EH-Electric Hob, HD-Hairdryer, WM-Machine Machine or Dryer, OV-Oven].

consumption (kWh) on a daily, time of day and hour of day scales. From the trend plots, we can identify the peak load (Watts) and peak energy consumption (kWh) hours. The bar charts analyzed average load (Watts) against usage duration (Hours) and energy consumption (kWh) for two groups of appliances; i.e., appliances contributing to peak load (kW) and appliances with highest usage duration for the house. Additionally, the two sets of pie charts analyze the appliances energy consumption contribution during peak hours, that is peak load (kW) and peak energy consumption (kWh). From the energy consumption analyses, it can be noted that appliances contributing to peak power loads such as Kettle, Toaster, Microwave, and Rice_Cooker, due to higher power ratings, are not the only appliances contributing towards household energy consumption. On the contrary, the AoI appliances such as Monitor, Laptop, and Speakers, having small power ratings, contribute a larger portion of energy usage. This situation is explained by the respective usage duration for these groups of appliances. The low power appliances are used for a longer duration, whereas high power appliances usage duration are small or very-very small. For example, in House 2, appliances Kettle, Toaster, and Microwave had high power rating but the major portion of energy usage was contributed by Monitor and Laptop, which were relatively small power footprint appliances. Similar observations were noted for House 1, where appliances such as Hoover, Toaster, and Microwave had high power ratings but consume a relatively small amount of energy when compared to Lights and Laptop, which were low power appliances. Additionally, for House 5, appliances such as Laptop, Audio/music equipment, and Oven made-up the group of AoI appliances against high power rating appliances such as Kettle, Vacuum Cleaner, Electric Hob, and Hairdryer.

Additionally, Figures 10, 11, and 12 are results of the cluster analysis representing appliance-time associations. The marked

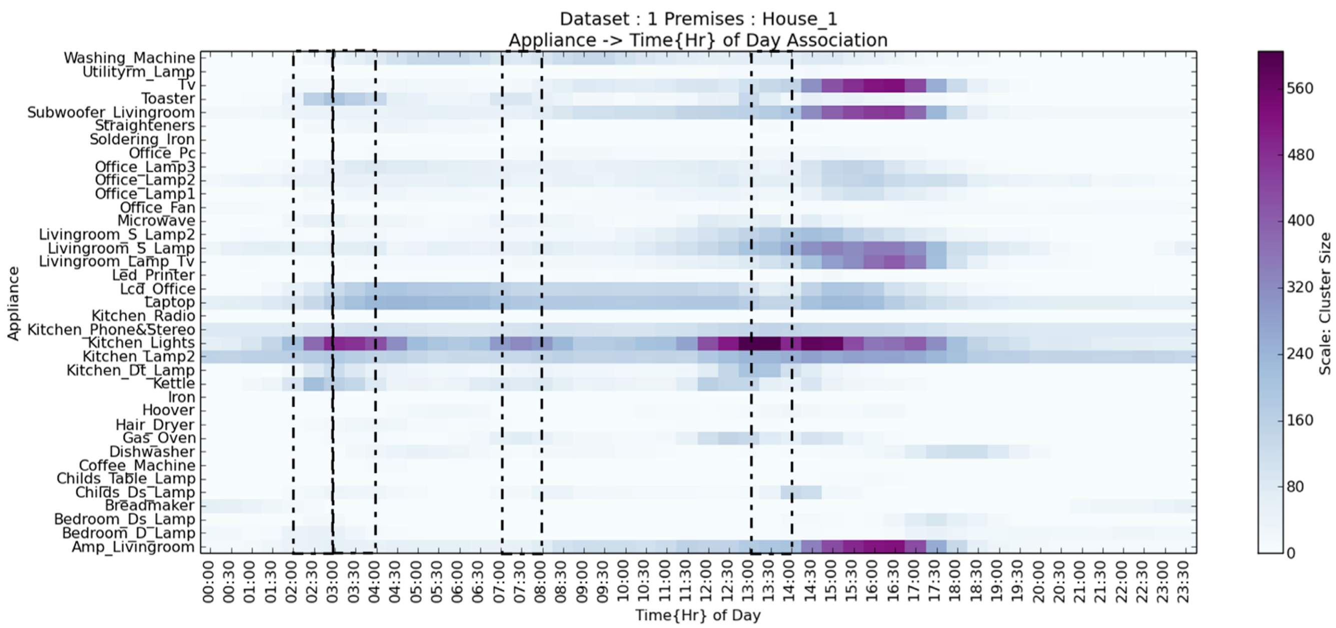

FIGURE 10. House 1: Appliance to time association.

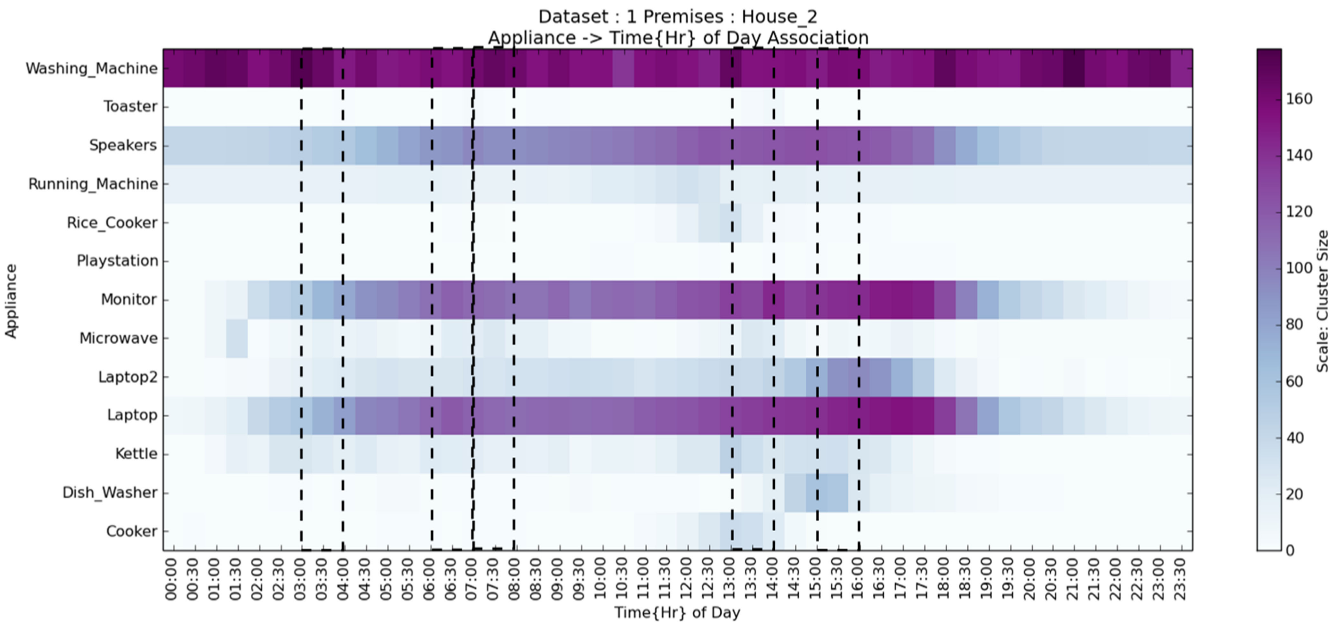

FIGURE 11. House 2: Appliance to time association.

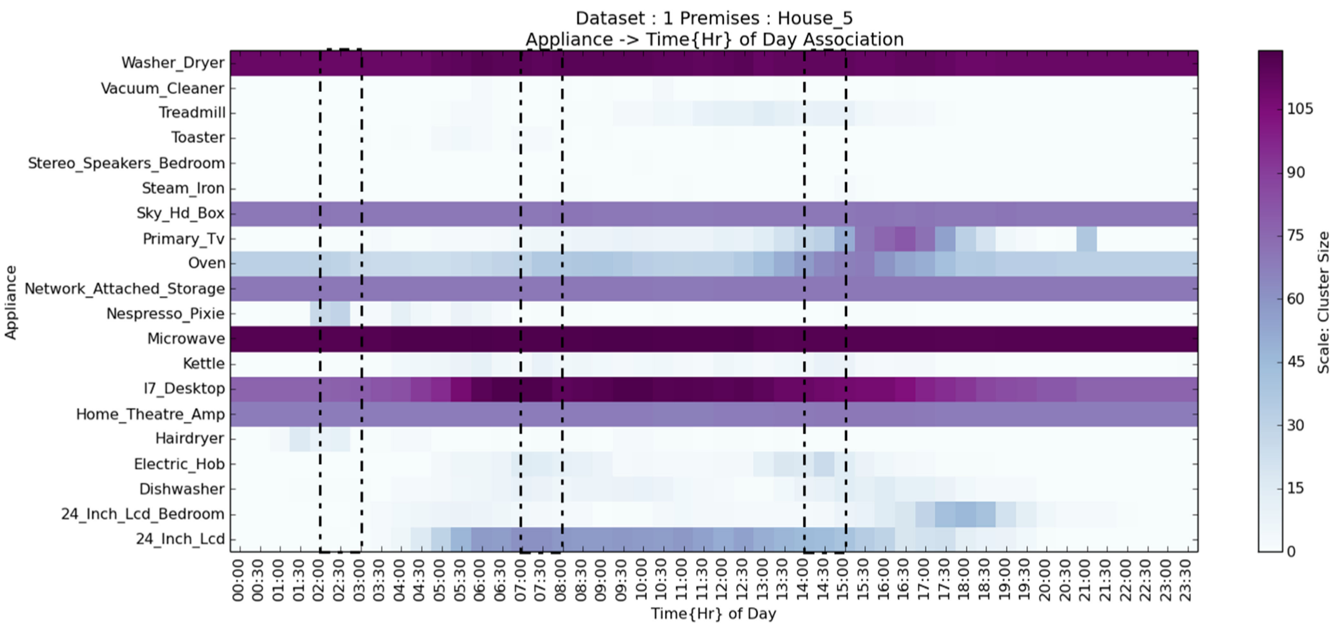

FIGURE 12. House 5: Appliance to time association.

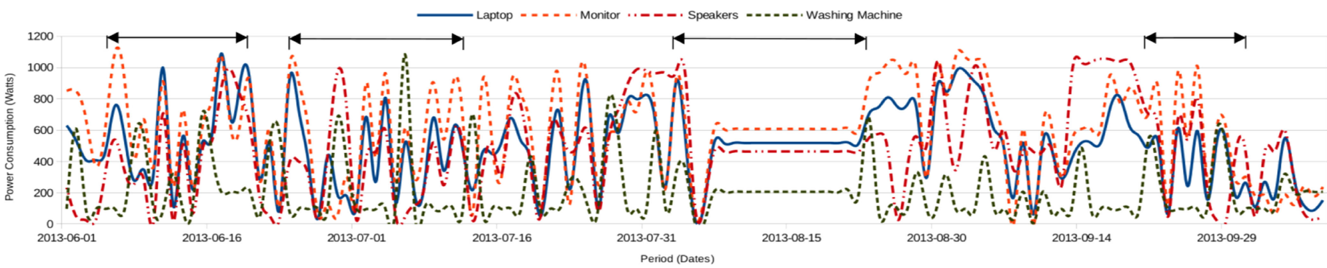

FIGURE 13. House 2: Appliances (laptop, monitor, speakers, washing machine) association via energy curves.

segments represent the peak load and energy consumption hours. We include only the appliance-hour of day association in this paper, but similar results are obtained for other identified units of time such as seasons, months, weeks, weekdays, and time of day. These appliance-time association results support our above findings while representing the concentration of usage of the appliance during specific hours; i.e., during peak load and peak energy consumption hours.

Furthermore, we observed appliances such as the Washer and Dryer, which were considered as first candidates for various demand side management programs, were not the major contributors to household energy consumption; rather, appliances such as Kettle, Lights, and Laptop were the largest contributors towards energy usage, which can assist establishing individual households behavioral attributes or household energy profile more effectively and accurately. Considering this kind of knowledge and incorporating them in various energy programs can increase active participation of consumers and attract their total involvement for a lasting effect.

From a demand side management scheduling perspective, the appliances contributing to peak power load should be considered first in any energy saving plan, but the user preferences should be accounted to minimize any inconveniences to the user's day-to-day lifestyle. Also, due consideration is required towards the manually operated appliances, which characterize consumers energy consumption behavior and devise energy savings plan in maximum possible agreement with the consumers to extend maximum benefits to them and simultaneously gain from the increased participation.

VI. CONCLUSION AND FUTURE WORK

In this paper, we demonstrated that the appliance associations are a direct reflection of the consumer energy usage behavior while revealing personal preferences depicting expected comfort. These must be indispensable input parameters to the energy saving programs and related decision-making processes while buying in the much-required consumer confidence to achieve successful persistent results. We identified Appliances of Interest the major energy consuming appliances for the house, which require equal attention from both consumers and utilities to attain energy efficiency. In our future work, we will extend our analysis of energy consumption behavior to include the prediction of multiple appliances usage on a short-term and long-term basis along with energy consumption forecasts. Also, we will explore the discovery of frequent energy consuming appliance discovery to refine our prediction results and accuracy by considering actual energy consumption and usage duration rather just on the frequency of usage of appliances.

REFERENCES

[1] T. Yu, N. Chawla, and S. Simoff, "Data analysis challenges in the future energy domain," in *Computational Intelligent Data Analysis for Sustainable Development*. London, U.K.: Chapman and Hall/CRC, 2013, pp. 181–242. [Online]. Available: https://books.google.ca/books?id=iKx4qyZPvCcC

[2] D. Schweizer, M. Zehnder, H. Wache, H. F. Witschel, D. Zanatta, and M. Rodriguez, "Using consumer behavior data to reduce energy consumption in smart homes: Applying machine learning to save energy without lowering comfort of inhabitants," in *Proc. IEEE 14th Int. Conf. Mach. Learning Appl.*, 2015, pp. 1123–1129.

[3] openADR, 2017. [Online]. Available: http://www.openadr.org/

[4] A. Piette, *et al.*, "Open automated demand response communications specification (version 1.0)," 2009. [Online]. Available: https://drrc.lbl.gov/sites/all/files/cec-500–2009-063.pdf

[5] K. Herter, J. Rasin, and T. Perry, "Development and demonstration of the open automated demand response standard for the residential sector," Nov. 2009, https://drrc.lbl.gov/sites/all/files/lbnl-6531e.pdf

[6] How much energy is consumed in residential and commercial buildings in the united states? 2016. [Online]. Available: http://www.eia.gov/tools/faqs/faq.cfm?id=86&t=1

[7] D. Torstensson and F. Wallin, "Potential and barriers of demand response at households customers," in *Proc. 7th Int. Conf. Appl. Energy Energy Procedia Clean Efficient Affordable Energy Sustainable Future*, 2015, pp. 1189–1196.

[8] K. Jack and K. William, "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes," *Sci. Data*, vol. 2, 2015, Art. no. 150007.

[9] B. Anca-Diana, G. Nigel, and M. Gareth, "Achieving energy efficiency through behaviour change: What does it take?" 2013. [Online]. Available: http://www.eea.europa.eu/publications/achieving-energy-efficiency-through-behaviour

[10] A. Zipperer, *et al.*, "Electric energy management in the smart home: Perspectives on enabling technologies and consumer behavior," *Proc. IEEE*, vol. 101, no. 11, pp. 2397–2408, Nov. 2013.

[11] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proc. 2nd ACM Workshop Embedded Sens. Syst. Energy-Efficiency Building*, 2010, pp. 61–66. [Online]. Available: http://doi.acm.org/10.1145/1878431.1878446

[12] G. Wood and M. Newborough, "Dynamic energy-consumption indicators for domestic appliances: Environment, behaviour and design," *Energy Buildings*, vol. 35, no. 8, pp. 821–841, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378778802002414

[13] G. Wood and M. Newborough, "Influencing user behaviour with energy information display systems for intelligent homes," *Int. J. Energy Res.*, vol. 31, no. 1, pp. 56–78, 2007. [Online]. Available: http://dx.DOI.org/10.1002/er.1228

[14] A. Albert and R. Rajagopal, "Smart meter driven segmentation: What your consumption says about you," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4019–4030, Nov. 2013.

[15] C. O. Adika and L. Wang, "Autonomous appliance scheduling for household energy management," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 673–682, Mar. 2014.

[16] C. O. Adika and L. Wang, "Autonomous appliance scheduling based on time of use probabilities and load clustering," in *Proc. 10th Int. Power Energy Conf.*, 2012, pp. 42–47.

[17] H. S. Cho, T. Yamazaki, and M. Hahn, "AERO: Extraction of user's activities from electric power consumption data," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 2011–2018, Aug. 2010.

[18] X. Zhang, T. Kato, and T. Matsuyama, "Learning a context-aware personal model of appliance usage patterns in smart home," in *Proc. IEEE Innovative Smart Grid Technol.-Asia*, 2014, pp. 73–78.

[19] N. Zhang, L. F. Ochoa, and D. S. Kirschen, "Investigating the impact of demand side management on residential customers," in *Proc. 2nd IEEE PES Int. Conf. Exhib. Innovative Smart Grid Technol.*, 2011, pp. 1–6.

[20] F. D. Silva and O. Mohammed, "Demand side load control with smart meters," in *Proc. IEEE Power Energy Soc. General Meeting*, 2013, pp. 1–5.

[21] S. Waczowicz, *et al.*, "Demand response clustering-how do dynamic prices affect household electricity consumption?" in *Proc. IEEE Eindhoven PowerTech*, 2015, pp. 1–6.

[22] S. Haben, C. Singleton, and P. Grindrod, "Analysis and clustering of residential customers energy behavioral demand using smart meter data," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 136–144, Jan. 2016.

[23] S. Rollins and N. Banerjee, "Using rule mining to understand appliance energy consumption patterns," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2014, pp. 29–37.

[24] M. Hassani, C. Beecks, D. Tows, and T. Seidl, "Mining sequential patterns of event streams in a smart home application," in *Proc. LWA Workshops KDML FGWM IR FGDB*, Oct. 2015. [Online]. Available: http://ceur-ws.org

[25] Y.-C. Chen, C.-C. Chen, W.-C. Peng, and W.-C. Lee, *Mining Correlation Patterns Among Appliances Smart Home Environment*. Cham, Switzerland: Springer, 2014, pp. 222–233. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-06605-9_19

[26] Y. C. Chen, H. C. Hung, B. Y. Chiang, S. Y. Peng, and P. J. Chen, "Incrementally mining usage correlations among appliances in smart homes," in *Proc. 18th Int. Conf. Netw.-Based Inf. Syst.*, 2015, pp. 273–279.

[27] Y. C. Chen, W. C. Peng, and S. Y. Lee, "Mining temporal patterns in time interval-based data," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 12, pp. 3318–3331, Dec. 2015.

[28] S. Rahimi, A. D. C. Chan, and R. A. Goubran, "Usage monitoring of electrical devices in a smart home," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 5307–5310.

[29] S. Rahimi, A. D. C. Chan, and R. A. Goubran, "Nonintrusive load monitoring of electrical devices in health smart homes," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, 2012, pp. 2313–2316.

[30] Y. S. Liao, H. Y. Liao, D. R. Liu, W. T. Fan, and H. Omar, "Intelligent power resource allocation by context-based usage mining," in *Proc. IIAI 4th Int. Congr. Adv. Appl. Informat.*, 2015, pp. 546–550.

[31] A. Alhamoud, *et al.*, *Extracting Human Behavior Patterns from Appliance-Level Power Consumption Data*. Cham, Switzerland: Springer, 2015, pp. 52–67. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-15582-1_4

[32] K. Gajowniczek and T. Zabkowski, "Data mining techniques for detecting household characteristics based on smart meter data," *Energies*, vol. 8, no. 7, 2015, Art. no. 7407. [Online]. Available: http://www.mdpi.com/1996-1073/8/7/7407

[33] Y. C. Chen, H. C. Hung, B. Y. Chiang, S. Y. Peng, and P. J. Chen, "Incrementally mining usage correlations among appliances in smart homes," in *Proc. 18th Int. Conf. Netw.-Based Inf. Syst.*, 2015, pp. 273–279.

[34] L. Ong, M. Bergés, and H. Y. Noh, "Exploring sequential and association rule mining for pattern-based energy demand characterization," in *Proc. 5th ACM Workshop Embedded Syst. Energy-Efficient Buildings*, 2013, pp. 25:1–25:2. [Online]. Available: http://DOI.acm.org/10.1145/2528282.2528308

[35] O. Ardakanian, N. Koochakzadeh, R. P. Singh, L. Golab, and S. Keshav, "Computing electricity consumption profiles from household smart meter data," in *Proc. Workshops EDBT/ICDT Joint Conf.*, 2014, pp. 140–147. [Online]. Available: http://ceur-ws.org/Vol-1133#paper-22

[36] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 420–430, Jan. 2014.

[37] Y. Ding, J. Borges, M. A. Neumann, and M. Beigl, "Using sequence mining to understand daily activity patterns for load forecasting enhancement," 2015. [Online]. Available: https://www.teco.edu/michael/publication/2015_ISC2_144.pdf

[38] C. Chelmis, J. Kolte, and V. K. Prasanna, "Big data analytics for demand response: Clustering over space and time," in *Proc. IEEE Int. Conf. Big Data*, 2015, pp. 2223–2232.

[39] S. Park, S. Ryu, Y. Choi, and H. Kim, "A framework for baseline load estimation in demand response: Data mining approach," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2014, pp. 638–643.

[40] F. L. Quilumba, W. J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 911–918, Mar. 2015.

[41] J. L. Viegas, S. M. Vieira, J. M. C. Sousa, R. Melicio, and V. M. F. Mendes, "Electricity demand profile prediction based on household characteristics," in *Proc. 12th Int. Conf. Eur. Energy Market*, 2015, pp. 1–5.

[42] M. Imani and U. Braga-Neto, "Optimal state estimation for boolean dynamical systems using a boolean kalman smoother," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2015, pp. 972–976.

[43] M. Imani and U. M. Braga-Neto, "Maximum-likelihood adaptive filter for partially observed boolean dynamical systems," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 359–371, Jan. 2017.

[44] J. Han, J. Pei, and M. Kamber, "Mining frequent patterns, associations, and correlations: Basic concepts and methods," in *Data Mining: Concepts and Techniques*, 3rd ed. San Mateo, CA, USA: Morgan Kaufmann, 2011, pp. 243–278. [Online]. Available: http://www.sciencedirect.com/science/book/9780123814791

[45] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2000, pp. 1–12. [Online]. Available: http://DOI.acm.org/10.1145/342009.335372

[46] J. Han, J. Pei, Y. Yin, and R. Mao, "Mining frequent patterns without candidate generation: A frequent-pattern tree approach," *Data Mining Knowl. Discovery*, vol. 8, no. 1, pp. 53–87, 2004. [Online]. Available: http://dx.DOI.org/10.1023/B:DAMI.0000005258.31418.83

[47] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *Proc. 20th Int. Conf. Very Large Data Bases*, 1994, pp. 487–499. [Online]. Available: http://dl.acm.org/citation.cfm?id=645920.672836

[48] J. Han, J. Pei, and M. Kamber, "Cluster analysis: Basic concepts and methods," in *Data Mining: Concepts and Techniques*, 3rd ed. San Mateo, CA, USA: Morgan Kaufmann, 2011, pp. 443–494. [Online]. Available: http://www.sciencedirect.com/science/book/9780123814791

[49] W. Haizhou and S. Mingzhou, "Ckmeans.1d.dp: Optimal k-means clustering in one dimension by dynamic programming," *R J.*, vol. 3, no. 2, pp. 29–33, Dec. 2011. [Online]. Available: http://journal.r-project.org/archive/2011-2/RJournal_2011-2_Wang+Song.pdf

SHAILENDRA SINGH received the BEng degree in electronics engineering from Pt. Ravishankar Shukla University, Raipur, India. He is currently working toward the MASc degree in electrical and computer engineering from the University of Ottawa, Canada. He has more than 18 years of industry experience, and currently a research assistant in the Distributed and Collaborative Virtual Environment Research Laboratory, School of Electrical Engineering and Computer Science, University of Ottawa. His research interests include software defined networks, artificial intelligence (AI), smart environments, and cyber-physical systems.

ABDULSALAM YASSINE received the BSc degree in electrical engineering from Beirut Arab University, Lebanon, in 1993, and the MSc and PhD degrees in electrical and computer engineering from University of Ottawa, Canada, in 2004 and 2010, respectively. Between 2001 and 2013, he was member of the technical staff in the Wireless Communication Division at Nortel Networks and later at Alcatel-Lucent, Ottawa, Canada. From 2013 until 2016, he was a post-doctoral fellow at DISCOVER Laboratory, School of Electrical Engineering and Computer Science, University of Ottawa. He is currently an assistant professor in the Department of Software Engineering, Lakehead University, Canada. His current research interests include mostly focused on systems and networks, multimedia, artificial intelligence (AI), smart environments, and smart grids. He is a member of the IEEE.