IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 27, 2017, accepted June 13, 2017, date of publication June 27, 2017, date of current version July 31, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2719921

Mining Human Activity Patterns From Smart Home Big Data for Health Care Applications

ABDULSALAM YASSINE[1], (Member, IEEE), SHAILENDRA SINGH[2], AND ATIF ALAMRI[3,4], (Member, IEEE)
[1]Department of Software Engineering, Lakehead University, Thunder Bay, ON P7B 5E1, Canada
[2]DISCOVER Laboratory, Department of Electrical Engineering and Computer Science, Ottawa, Ontario, K1N 6N5, Canada
[3]Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[4]Research Chair of Pervasive and Mobile Computing, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding author: Abdulsalam Yassine (ayassine@lakeheadu.ca)

ABSTRACT Nowadays, there is an ever-increasing migration of people to urban areas. Health care service is one of the most challenging aspects that is greatly affected by the vast influx of people to city centers. Consequently, cities around the world are investing heavily in digital transformation in an effort to provide healthier ecosystems for people. In such a transformation, millions of homes are being equipped with smart devices (e.g., smart meters, sensors, and so on), which generate massive volumes of fine-grained and indexical data that can be analyzed to support smart city services. In this paper, we propose a model that utilizes smart home big data as a means of learning and discovering human activity patterns for health care applications. We propose the use of frequent pattern mining, cluster analysis, and prediction to measure and analyze energy usage changes sparked by occupants' behavior. Since people's habits are mostly identified by everyday routines, discovering these routines allows us to recognize anomalous activities that may indicate people's difficulties in taking care for themselves, such as not preparing food or not using a shower/bath. This paper addresses the need to analyze temporal energy consumption patterns at the appliance level, which is directly related to human activities. For the evaluation of the proposed mechanism, this paper uses the U.K. Domestic Appliance Level Electricity data set—time series data of power consumption collected from 2012 to 2015 with the time resolution of 6 s for five houses with 109 appliances from Southern England. The data from smart meters are recursively mined in the quantum/data slice of 24 h, and the results are maintained across successive mining exercises. The results of identifying human activity patterns from appliance usage are presented in detail in this paper along with the accuracy of short- and long-term predictions.

INDEX TERMS Big data, smart cities, smart homes, health care applications, behavioral analytics, frequent pattern, cluster analysis, incremental data-mining, association rules, prediction.

I. INTRODUCTION

Studies show that by year 2050, 66% of the world population will be living in urban areas [1]. The demand for health care resources will be greatly affected by this vast influx of people to city centers. This unprecedented demographic change places enormous burden on cities to rethink the traditional approaches of providing health services to residents. In responding to the new needs and challenges, cities are currently embracing massive digital transformation in an effort to support sustainable urban communities, and provide healthier environment [2], [3]. In such transformation, millions of homes are being equipped with smart devices (e.g. smart meters, sensors etc.) which generate massive volumes of fine-grained and indexical data that can be analyzed to support health care services. Advancement of big data mining technologies, which provide means of processing huge amount of data for actionable insights, can aid us in understanding how people go about their life. For example, monitoring the changes of appliance usage inside a smart home can be used to indirectly determine the person's wellbeing based on historical data. Since people's habits are mostly identified by everyday routines, discovering these routines allows us to recognize anomalous activities that may indicate people's difficulties in taking care for themselves, such as not preparing food or not using shower/bath [4], [5]. The underlying correlation between appliance usage inside the smart home and routine activities can be used by health care applications to detect potential health problems. This is not only going to alleviate the burden on health care systems, but also providing 24 hour monitoring service that automatically identify normal and abnormal behaviors for independently living patients or those with self-limiting conditions (e.g. elderly and patients with cognitive impairments).

This paper proposes the use of energy data from smart meters installed at homes to unveil important activities of inhabitants. Our study assumes that there are mechanisms in place to protect people's privacy from being shared or measured for unlawful uses as discussed in [6] and [7]. The proposed model observes and analyzes readings from smart meters to recognize activities and changes in behavior. Disaggregated power consumption readings are directly related to the activities performed at home. For instance, if the ''Oven'' is ON, the operation of this appliance is most likely associated with activity ''Preparing Food''. The time (e.g. morning or evening) of this operation may also indicate the type of the meal such as breakfast or dinner. Furthermore, people often perform more than one activity at the same time such as ''Preparing Food'' and ''Listening to Music'' or ''Watching TV'', which means multiple appliances are operated together. In this context, we analyze consumers' temporal energy consumption patterns at the appliance level to detect multiple appliance usages and predict their operations over short and long term time-frames. This is particularly possible without additional hardware since the smart meter data have time-series notion typically consisting of usage and consumption measurements patterns of component appliances over a time interval [8]. Such endeavor, however, is very challenging since it is not easy to detect usage dependencies among various appliances when their operation overlap or occur at the same time. Furthermore, deriving accurate prediction of human activity patterns is influenced by the probabilistic relationships of appliance usage events that have dynamic time intervals.

To tackle the aforementioned issues, this paper proposes frequent mining and prediction model to measure and analyze energy usage changes sparked by occupants' behavior. The data from smart meters are recursively mined in the quantum/data slice of 24 hours, and the results are maintained across successive mining exercises. We also utilize the Bayesian network, a probabilistic graphical model, to predict the use of multiple appliances and household energy consumption. The proposed model is capable of short-term predictions ranging from next hour up to 24 hours and long-term prediction for days, weeks, months, or seasons. For the evaluation of the proposed mechanism, this research uses the UK Domestic Appliance Level Electricity dataset (UK-Dale) [9] - time series data of power consumption collected from 2012 to 2015 with time resolution of six seconds for five houses with 109 appliances from Southern England. It must be noted that in practice load disaggregation is carried by Non-Intrusive Appliance Load Monitoring (NALM) technique. NALM is a technique used to disaggregate a home's power usage into individual appliances and label them for further mining and analysis. The main contributions of this paper are as follows:

- We propose a human activity pattern mining model based on appliance usage variations in smart homes. The model which utilizes FP-growth for pattern recognition and k-means clustering algorithms is capable of identifying appliance-to-appliance and appliance-to-time associations through incremental mining of energy consumption data. This is not only important to determine activity routines, but also, when utilized by health care application, is capable of detecting sudden changes of human activities that require attention by a health provider.
- We apply a Bayesian network for activity prediction based on individual and multiple appliance usage. This is significant for health applications that incorporate reminders for patients to perform certain activities based on historical data. For added accuracy of the system, the prediction model integrates probabilities of appliance-to-appliance and appliance-to-time associations, thus recognizing activities that occur in certain patterns more accurately.

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 and results analysis in section (IV). Finally, we conclude the paper and discuss future directions in section (V).

II. RELATED WORK

Lately, there has been a growing interest in using smart home technologies for detecting human activity patterns for health monitoring applications. The main goal is to learn occupants' behavioral characteristics as an approach to understand and predict their activities that could indicate health issues. In this section, we review existing work in the literature, which employ smart homes data to analyze users' behavior.

Detecting human activities in smart homes by means of analyzing smart meters data is studied in [10]. The paper proposes two approaches to analyze and detect user's routines. One approach uses Semi-Markov-Model (SMM) for data training and detecting individual habits and the other approach introduces impulse based method to detect Activity in Daily Living (ADL) which focuses on temporal analysis of activities that happen simultaneously. Similarly, the work in [11] proposes human activity detection for wellness monitoring of elderly people using classification of sensors related to the main activities in the smart home. Smart meters data are also used in [4] for activity recognition using Non-intrusive Appliance Load Monitoring(NALM) and Dempster-Shafer (D-S) theory of evidence. The study collects pre-processed data from homes to determine the electrical appliance usage patterns and then employs machine learning-based algorithm to isolate the major activities inside the home. The issue is that the study has to perform two steps on the data to completely isolate the main activities. Exploiting appliance usage patterns and identify them for sudden behavioral change is presented in [12]. The aim of the study is to provide around the clock monitoring system to support people's suffering from Alzheimer or Parkinson disease at minimum intrusion level. The study uses classification techniques to detect abnormal behavior of personal energy usage patterns in the home. Other studies such as [13]–[15] and [16] although do not utilize smart meters data, they use Internet of Things (IoT) infrastructures in smart cities for developing applications that monitor and provide health services for patients.

Using data analytics for smart meters to detect and predict behavioral abnormality for remote health monitoring is discussed in [17] and [18]. Alam *et al.* [17] use everyday appliances usage from smart meter and smart plug data to trace regular activities and learn unique time segment groups of appliance's energy consumption. The study employs hierarchical probabilistic model-based detection to infer about discovered anomalous behavior. This in tern can be used to understand the criticality of some abnormal behaviors for sustaining better health care. In [18] an experimental demonstration for observing and measuring energy consumption of appliances is presented. The study aims to provide a portrait profile of activities of daily living for elderly patients independently living at home. The data is also used to mine important patterns of changes for short-term and long-term anomaly detection of urgent health conditions. The work in [19], uses Bayesian networks to predict occupant behavior from collected smart meters data. The study proposes behavior as a service based on a single appliance, but does not provide a model to be applied for real-world scenarios. Authors in [20] and [21], used time-series multi-label classifier to forecast appliance usage based on decision tree correlations, however, the study takes only the last 24-hour window along with appliance sequential relationships. Chelmis *et al.* [22] suggest a clustering approach to identify the distribution of consumers' temporal consumption patterns, however, the study does not consider appliance level usage details. This might not be applicable for human activity recognition since specific activities require individual and multiple appliance to appliance and time associations. The work in [23] considers the appliances' ON and OFF status to detect usage pattern using hierarchical and c-means clustering. However, the study does not consider the duration of appliance usage or the expected variations in the sequence of appliance usage. The work in [24] proposes graphical model based algorithm to predict human behavior and appliance interdependency patterns and use it to predict multiple appliance usages using a Bayesian model.

The above-discussed approaches do not consider appliance level usage patterns, which is critical in determining human activity variations. Furthermore, our experiments are conducted using a much larger dataset than existing studies although there are similarities in data analytics techniques between the proposed study and existing work.

III. PROPOSED MODEL

Figure 1 represents the proposed model. It starts by cleaning and preparing the data and then applying frequent pattern mining for discovering appliance-to-appliance associations, i.e., determining which appliances are operating together. Then, it uses cluster analysis to determine appliance-to-time associations. With these two processes, the system is able to extract the pattern of appliance usage which is then used as input to the Bayesian network for short-term and long-term activities prediction. The output of the system is utilized by specific health care applications depending on the intended use. For example, a health care provider might only interested in knowing activities related to cognitive impairment where tracking the sequence of daily activities is crucial for reminding the patient when abnormal behavior is detected. Next subsection explain such processes and briefly outlines the theoretical background.

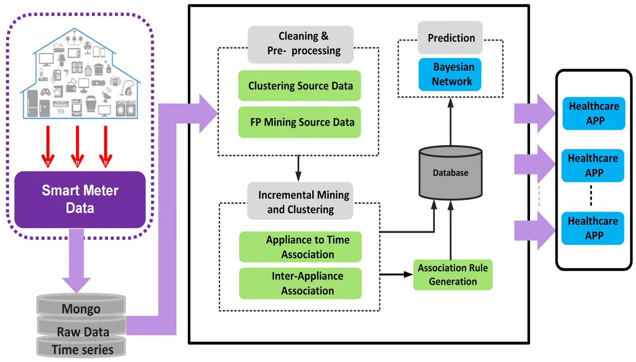

FIGURE 1. Model: Mining frequent patterns and activity predictions for health care applications in smart homes.

TABLE 1. Ready-to-mine source data.

Date	ST	ET	Active Appliances
2013-08-01	07:00	07:30	'2 3 4 12'
2013-08-01	07:30	08:00	'3 4 12'
2013-08-01	08:00	08:30	'2 4 12'
2013-08-01	08:30	09:00	'4 12'

ST= Start Time, ET=End Time, 2 = Laptop
3 = Monitor, 4 = Speakers, 12 = Washing Machine

TABLE 2. Clustering source database.

Appliance	Hour of Day	Time of Day
2	07:30 08:00	M E
3	16:00 16:30	A E
12	13:30 14:00 14:30	A

2 = Laptop, 3 = Monitor, 12 = Washing Machine
M = Morning, A = Afternoon, E = Evening

A. DATA PREPARATION

The dataset used in this study is a collection of smart meters data from five houses in the United Kingdom (UK) [9]. This dataset includes 400 million raw records at time resolution of 6 seconds. In the first stage of the cleaning process we developed customized procedures to remove noises from the data and prepare it for mining. After cleaning and preparation, the dataset is reduced to 20 million. Additionally, we developed a synthetic dataset for preliminary evaluation of the model, having over 1.2 million records. In tables 1 and 2 we show an

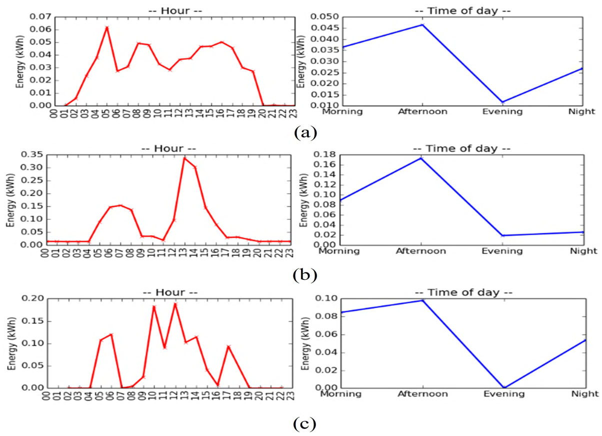

FIGURE 2. Hour and time of the day for active appliances: (a) TV, (b) Oven, (c) Treadmill.

example of the resulting ready to mine source data format comprising four appliances from one house. Smart meters time-series raw data, which is a high time-resolution data, is transformed into a 1-minute resolution load data; subsequently translated into a 30 minutes time-resolution source data, i.e. $24 * 2 = 48$ readings per day per appliance, while recording start time and end time for each active appliance.

B. EXTRACTING FREQUENT PATTERNS OF HUMAN ACTIVITIES

As mentioned earlier, the aim is to discover human activity patterns from smart meters data. For example, activities such as "Watching TV, Cooking, Using Computer, Preparing Food and Cleaning Dishes or Clothes" are usually regular routines. Our aim is to detect the patterns of these activities so that a health care application, that monitors sudden changes in patient's behavior (e.g. patients with cognitive impairment), can send timely alert to health care providers. In pursuing such process, all appliances that are registered active during the 30-minute time interval are included into the source database for frequent pattern data mining. Figure 2 is an example of active appliances that indicate three different activities at home. The energy trace of appliances (TV, Oven and Treadmill) is related to human activities such as leisure/relaxation time, food preparation, and exercising. A simplified example which describes possible relationships between appliance usage and activities is shown in figure 3. Extracting human activity patterns is not only discovering the individual appliance operation, but also the appliance-to-appliance associations; i.e., the patterns of activities that are combined together such as washing clothes while exercising or watching TV. The underlying concept of the model is based on [25] and [26], which propose pattern growth or FP-growth approach using depth-first divide-and-conquer technique. However, this operation usually best performed off line, which might not be applicable for health applications that require prompt reaction for decision making. Therefore, we propose a new technique that 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. Extensive details about the proposed incremental frequent pattern mining can be found in our previous work [27] and [28]; for the sake of completeness in this paper, we briefly describe the preliminaries and provide the algorithm that describes the incremental mining process.

TABLE 3. Frequent patterns: frequent patterns discovered database.

Frequent Pattern	Absolute Support (S)	Database Size (D)	Support or Probability $P = S/D$
'2 3'	3939	7899	0.4986
'2 4'	2840	7899	0.3595
'2 3 4'	2649	7899	0.3353
2 = Laptop, 3 = Monitor, 4 = Speakers			

Let $\Gamma = \{I_1, I_2, \ldots, I_k\}$ be an itemset containing k items which is referred to as k-itemset (l_k). In our case the items are the set of appliances in the database described in table 1, where each transaction Υ is an itemset having $\Upsilon \subseteq \Gamma$ and $\Upsilon \neq \emptyset$. The support of an itemset is the subset of transactions in the database containing all the itemsets. 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, provided respective *support* s_X and s_Y, if the percentage of transactions of the itemset appears in the transaction database *DB* are greater than or equal to *minsup*; where *minsup* is the pre-defined minimum support 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* $s_{X \Rightarrow Y}$ 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* $c_{X \Rightarrow Y}$ is defined as the percentage of transactions in *DB* containing X that also contain Y, which is the conditional probability, $P(Y|X)$ [29]. In the proposed setup, we made sure that all frequent patterns are preserved in the database. In this regard, newly discovered frequent patterns are compared with existing ones in the database and the *support* is updated or else the new pattern is added in the database. Once the mining process is completed, the size of the database *database_size* is updated for all the frequent patterns so that the computation of the *support* is correctly updated. Algorithm (1) outlines the incremental frequent pattern mining process and the results are presented in table 3. Further, from the definition of *support* which is the probability of itemset in the transaction database, the marginal distribution for appliance-appliance association can be computed at a global level as shown in table 3. The calculated marginal distribution determines the probability of appliances being concurrently active.

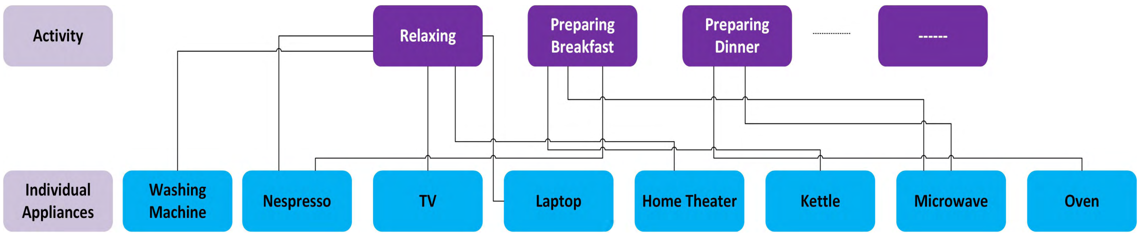

FIGURE 3. **Example of possible relationships between appliance usage and daily activities inside typical smart home.**

Algorithm 1 Incremental Frequent Pattern Mining

Require: Transaction database (*DB*), Frequent pattern discovered database (*FP_DB*)

Ensure: Incremental discovery of frequent patterns, stored in frequent patterns discovered database (*FP_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 database size Database_Size_db24 for data slice/quantum db_24
    Mine Frequent patterns in FP_DB_db24 using extended FP-growth approach
    for all Frequent Pattern FP in FP_DB_db24 do
        Search a frequent pattern FP in FP_DB
        if Frequent Pattern found then
            Update frequent pattern in FP_DB
        else
            Add a new Frequent Pattern to FP_DB
        end if
    end for
    For all Frequent Patterns in Database FP_DB increment Database Size by Database_Size_db24
end for
```

C. CLUSTERING ANALYSIS: INCREMENTAL k-MEANS

Discovering appliance-to-time associations is vital to health applications that monitor patients' activity patterns on a daily basis. In this section, a clustering analysis mechanism is used to discover appliance usage time with respect to hour of day (00:00 - 23:59), time of day (Morning, Afternoon, Evening, Night), weekday, week and/or month of the year.

Appliance-to-time associations are underlying information in the smart meter time series data which include sufficiently close time-stamps, when relevant appliance has been recorded as active or operational. Using this data we can group a class or cluster of appliances that are in operation simultaneously or overlapping. The size of the cluster that describes such associations is defined as the count of members in the cluster as well as its relative strength. Clustering analysis is the process of creating classes (unsupervised classification) or groups/segments (automatic segmentation) or partitions where members 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 [30]. We selected 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 and months have natural segmentation. For a dataset, *DB*, having n data points in Euclidean space. Partitional clustering distributes the data points from *DB* into k clusters, $C_1, C_2, \ldots, C_k$, having centroids $c_1, c_2, \ldots, c_k$ such that $C_i \subset DB$, $C_i \cap C_j = \emptyset$ and $c_i \neq c_j$ for$(1 \leq i, j \leq k)$. An objective function based on Euclidean distance, $distance(x, y) = \sqrt{\sum_j (x_j - y_j)^2}$, is used to measure the cohesion among data points, which reflects the quality of the cluster. This objective function is the sum of the squared error (SSE), $SSE = \sum_{i=1}^{k} \sum_{d \in C_i} distance(d, c_i)^2$, and k-means algorithm seeks to minimize the SSE. And, make use of the *silhouette score* [31] which is calculated based on the *eluclidean* distance to determine the optimal number of the clusters; i.e., k. In our model, incremental progressive clustering is obtained by consolidating existing and newly discovered clusters of each successive mining operation into the database. This incremental process is achieved by making sure all relevant cluster parameters such as centroid, SSE, Silhouette coefficient (width), data points and distance from the centroid are recorded in the database. The algorithm that ensures efficiency and speed of such operation can be found in [32]. Figures 4 and 5 show the clustering analysis for time of the day and day of the week for house 5 in the dataset. Both figures show the relationships between the appliance operation and the time of use. This information along with inter-appliance associations are used for predicting the activities inside houses as described next.

D. BAYESIAN NETWORKS FOR ACTIVITY PREDICTION

In this section, we integrate the frequent patterns and appliance-to-time associations to learn about the use of multiple appliances and build the activity prediction model. The mechanism utilizes Bayesian network which is a directed acyclic graph, where nodes represent random variables and edges indicate probabilistic dependencies. An example of Bayesian network, representing 6 random variables, is shown in figure (6). One of the main features of a Bayesian network is that it includes the concept of causality. For example, the

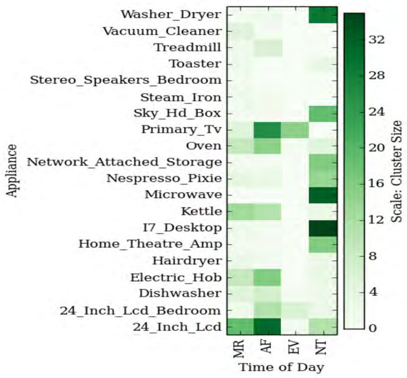

FIGURE 4. Incremental clustering of appliances-to-time of day associations inside one of the five houses using 25% of the Data.

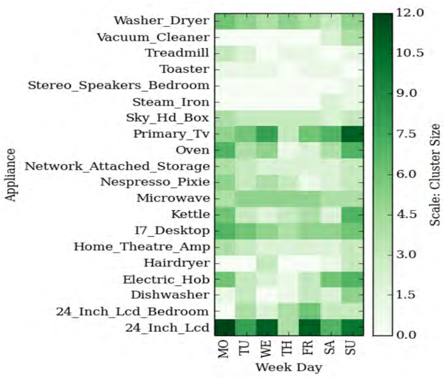

FIGURE 5. Incremental clustering of appliances-to-time of week associations inside one of the five houses using 25% of the Data.

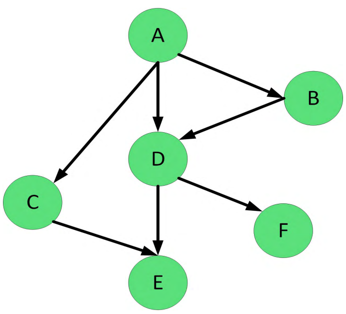

FIGURE 6. Example of Bayesian Network.

link/arc between A to C in figure 6 indicates that node A causes node C, which means that the directed graph in a Bayesian network is acyclic. In addition to the structure, a Bayesian network model provides a compact way of representing the joint probability distribution. In other words, each node or variable is independent of its nondescendants and accompanied by its local conditional probability distributions in the form of a node probability table, which facilitates the computation of the join conditional probability distribution for the model [33], [34]. An important benefit of the Bayesian network is the capability of mitigating missing data, learn relationships, and make use of historical facts and observations while avoiding overfitting of data [35]. A Bayesian network is defined by the probabilistic distribution presented in equation (1) [36], [37].

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i|parents(x_i)) \tag{1}$$

As mentioned above, our probabilistic prediction model is constructed based on integrating probabilities for appliances-to-time associations in terms of hour of day (00:00 - 23:59), time of day (Morning, Afternoon, Evening, Night), weekday, week, month, season, and appliance-to-appliance associations. The topology of the resulting Bayesian network has only one level of input evidence nodes, accompanied by respective unconditional probabilities, converging to one output node. Equation (2) presents the posterior probability or marginal distribution for the proposed prediction model.

$$\begin{aligned} p(.) = p(Hour) \times p(Time\ of\ day) \times p(Weekday) \\ \times p(Week) \times p(Month) \times p(Season) \end{aligned} \tag{2}$$

Table 4 shows sample of the training data. This data is derived from clustering and frequent patterns analysis where the probability of each appliance represents its operation during the specified period. This information is utilized by the Bayesian mechanism to determine and predict active appliances, operating simultaneously, using historical evidence from the cluster analysis (appliance-time association) and frequent pattern mining (appliance-appliance association). Furthermore, appliance prediction results establish the foundation for human activity prediction from the next hour up-to 24 hours (short-term) and days, weeks, or months (long-term). Next, we evaluate our model and provide result analysis.

TABLE 4. Node probability table - marginal distribution for appliances.

	Cluster Center/Centroid				
Appliance	00:00	00:30	01:00	01:30	02:00
2	0.0608	0.0602	0.1257	0.1414	0.1404
3	0.0338	0.0361	0.1078	0.1212	0.1149
4	0.1351	0.1205	0.1198	0.1010	0.0936
8	0.0270	0.0422	0.0719	0.1010	0.0979
10	0.0878	0.0783	0.0778	0.0657	0.0553
11	0.0068	0.0181	0.0180	0.0404	0.0511
12	0.6014	0.5060	0.4731	0.3939	0.4043
13	0.0000	0.0000	0.0000	0.0051	0.0043
15	0.0473	0.1386	0.0060	0.0303	0.0383

2 = Laptop, 3 = Monitor, 4 = Speakers
8 = Kettle, 10 = Running Machine, 11= Laptop2
12= Washing Machine, 13 = Dishwasher, 15 = Microwave

IV. EVALUATION AND RESULTS

For the evaluation of the proposed model, we performed our experiments using the dataset UK-Dale [9] along with the synthetic dataset to inspect intermediate and final results. The (UK-Dale) dataset 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

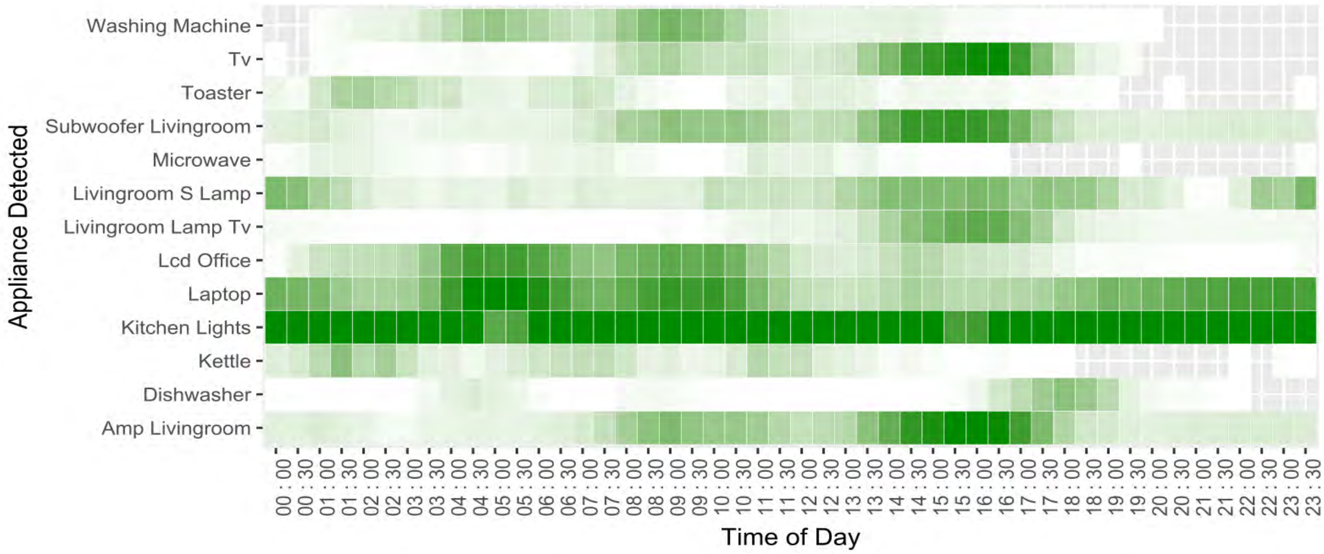

FIGURE 7. House 3: Appliance-Time Associations hour of the day.

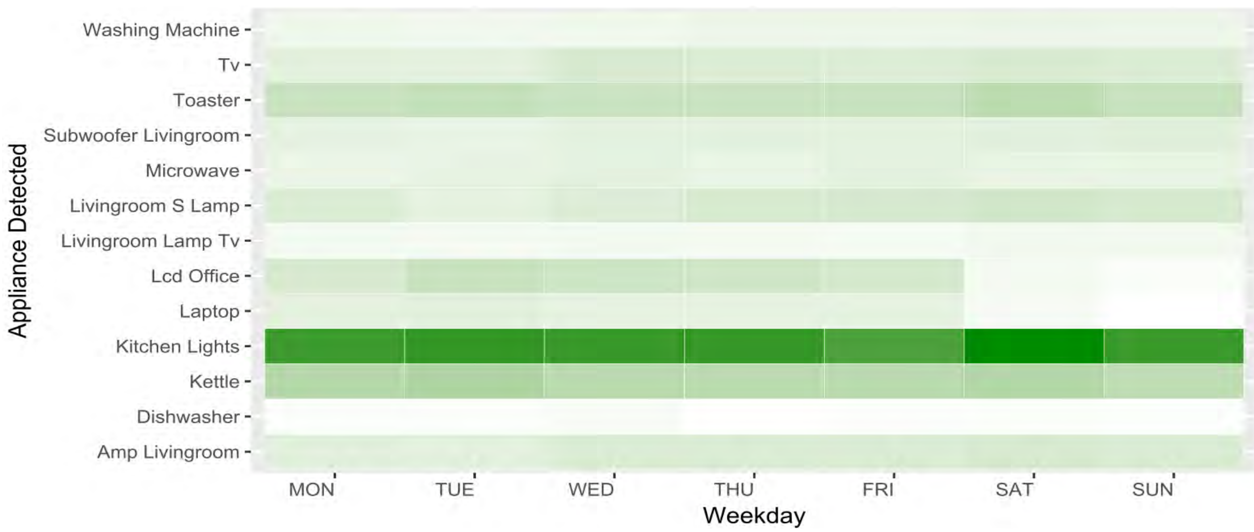

FIGURE 8. House 3: Appliance-Time Associations day of the week.

Southern England published by UK Energy Research Centre Energy Data Centre (UKERC-EDC). This dataset is one of the largest datasets having approximate half a billion records. Energy consumption measurement was conducted at appliance level using plug-in individual appliance monitors (IAMs) [9]. The underlying system for the proposed model is developed in Python, and the data is stored in MySQL and MongoDB databases on a ubuntu 14.04 LTS 64-bit system. The main objective of the experiments is to detect the appliance usage as an indication of human activity patterns and use the prediction model to forecast the short and long term activities inside the house. For a health care application, this means that our model can be used to feed mechanisms such as active monitoring, alert generation, health profiling etc.

A. RESULTS ANALYSIS AND DISCUSSION

The first step in understanding human activities is by extracting associations of appliance usage. Figure 7 and 8 show the appliance-to-time associations discovered for time of the day and weekday respectively for house 2. We can see that between 2:30 and 5:00 PM TV, Toaster, Livingroom Lights are used together in this house with highest concentration during the weekend. Also, the washing machine and Laptop are simultaneously used between 8:30 and 10 am. The washing machine is used almost all weekdays, where the Laptop is not used on the weekends. Considering these facts we can notice the varying effect of time and days on the use of appliances. In table 5, we show appliance-to-appliance association. The result is for 3 houses and it is based on processing 25% of the dataset. One can easily observe from appliance associations

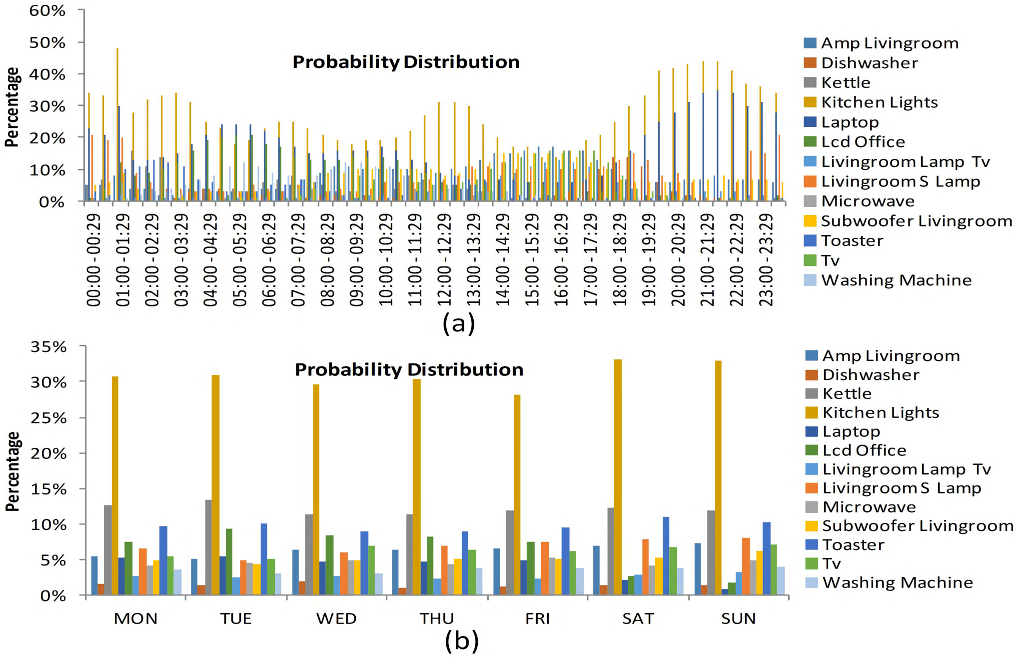

FIGURE 9. Detected appliances with probability distribution over 24-hour period. (a) Probability distribution of appliances for time of the day. (b) Probability distribution of appliances for day of the week.

TABLE 5. Appliance association rules.

	Sr.	Association Rule	Support	Confidence
House 1	1	Tv ⇒ Amp_Livingroom	0.15	0.96
	2	Amp_Livingroom ⇒ Tv	0.15	0.73
	3	Lcd_Office ⇒ Laptop	0.14	0.93
	4	Tv, Amp_Livingroom ⇒ Subwoofer_Livingroom	0.13	0.87
	5	Kitchen_Lights, Subwoofer_Livingroom ⇒ Amp_Livingroom	0.08	0.85
	6	Amp_Livingroom, Livingroom_Lamp_Tv ⇒ Subwoofer_Livingroom	0.08	0.98
	7	Subwoofer_Livingroom, Livingroom_Lamp_Tv ⇒ Amp_Livingroom	0.08	0.97
House 2	1	TV ⇒ Laptop	0.50	0.99
	2	Laptop ⇒ TV	0.50	0.93
	3	Speakers ⇒ Laptop	0.36	0.74
	4	TV, Speakers ⇒ Laptop	0.34	1.00
	5	Laptop, Speakers ⇒ TV	0.34	0.93
	6	TV, Washing Machine ⇒ Laptop	0.22	0.98
	7	Laptop, Washing Machine ⇒ TV	0.22	0.91
	8	Laptop, Washing Machine ⇒ Speakers	0.18	0.74
	9	TV, Speakers, Washing Machine ⇒ Laptop	0.16	1.00
	10	Laptop, Speakers, Washing Machine ⇒ TV	0.16	0.91
	11	TV, Washing Machine ⇒ Laptop, Speakers	0.16	0.73
House 5	1	Washer_Dryer ⇒ Microwave	0.82	0.92
	2	Microwave ⇒ Washer_Dryer	0.82	0.88
	3	Desktop ⇒ Washer_Dryer	0.70	0.91
	4	Washer_Dryer ⇒ Desktop	0.70	0.79
	5	Desktop ⇒ Microwave	0.70	0.91
	6	Microwave ⇒ Desktop	0.70	0.75
	7	Desktop Microwave ⇒ Washer_Dryer	0.63	0.91
	8	Desktop Washer_Dryer ⇒ Microwave	0.63	0.90
	9	Desktop ⇒ Microwave Washer_Dryer	0.63	0.83
	10	Microwave Washer_Dryer ⇒ Desktop	0.63	0.77
	11	Washer_Dryer ⇒ Desktop Microwave	0.63	0.71

House 1 : $minsup \geq$0.08, $minconf \geq$ 0.70
House 2 : $minsup \geq$0.29, $minconf \geq$ 0.65
House 5 : $minsup \geq$0.60, $minconf \geq$ 0.70

that occupants of house 1 like to relax while preparing food. This is evident from strong associations of appliances Kitchen Lights, Subwoofer, Amp, and TV. For house 2, occupants like to use the computer or listen to music while washing clothes. Similarly for house number 5, occupants use the computer while cooking or doing the laundry. Such associations are

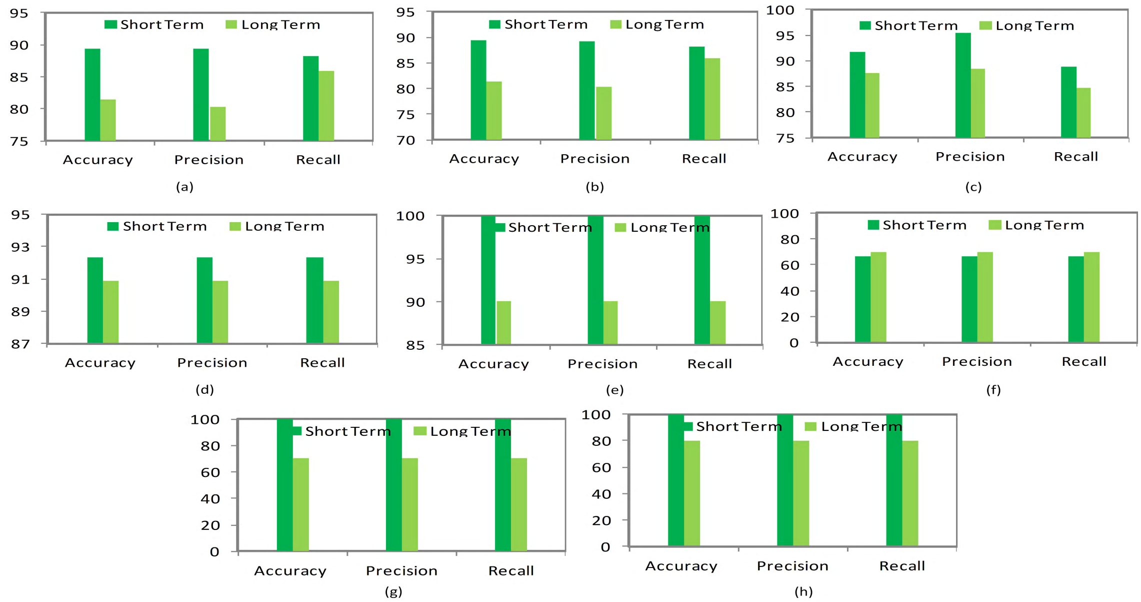

FIGURE 10. Prediction Model Accuracy, Precision , Recall: (a), (b), (c) are short and long-term predictions @ 25 %, 50% and 75% of training data respectively. (d), (e), (f), (g), (h) House Level Accuracy, Precision, Recall for short and long-term prediction @ 75% of the training data.

TABLE 6. House 2: examples of activity recognition results.

No.	Time	Appliances detected	Probability	Activity
1	5:00-7:00	Kitchen lights, Toaster, Kettle	20.794	Preparing Breakfast
2	9:00-11:00	Laptop, Lcd Office, Washing machine	60.45	Using Computer while doing laundry
3	13:00-14:00	Microwave, Laptop, TV	24.47	Relaxing
4	17:00-18:00	TV, Washing machine, Livingroom lights	38.43	Watching TV while doing laundry

characteristics of human behavioral traits that are performed routinely. Table shows few examples of possible activities inside the house based on the probability of appliance associations. These are just samples of human activities that can be discovered by our system and be used to detect anomalies that deviate from normal patterns.

As explained in subsection III-D, the prediction model utilizes appliance-to-appliance and appliance-time associations to predict multiple concurrent operating appliances. Figure 10 shows the probability distribution of appliances for house 2 over time of the day and weekday. Figures 10(d)-(h) shows the model accuracy for short-term, long-term and overall predictions at three stages of incremental data mining process; i.e., 25%, 50% and 75% of the dataset used as training data. Figures 10(d)–(e) show short and long term predictions for the five houses for 75% incremental data mining. The proposed model attains combined accuracy of 81.82(25%), 85.90(50%), 89.58(75%) at each stage, respectively. The obtained short term accuracy for houses 1, 2, 3, 4, and 5 is 92.31%, 100.00%, 66.67%, 100.00%, and 100.00% respectively. The obtained long term accuracy for houses 1, 2, 3, 4, and 5 is 90.91%, 90.00%, 70.00%, 70.00%, and 80.00% respectively.

Based on the above results, we can easily see the strong relationship between appliance usage inside the smart houses and human activity recognition. Learning the appliance-to-appliance and appliance-to-time associations extracted from the frequent pattern mining and cluster analysis are key processes to track patients/people's routines and possibly provide them with health services when needed.

V. CONCLUSION AND FUTURE WORK

In this paper, we presented a model for recognizing human activities patterns from low resolution smart meters data. Occupants' habits and behavior follow a pattern that could be used in health applications to track the wellbeing of individuals living alone or those with self-limiting conditions. Most of these activities can be learned from appliance-to-appliance and appliance-to-time associations. We presented incremental frequent mining and prediction model based on Bayesian network. In our current work, through experiments, we found that 24-hour period was optimal for data mining, but we built the model to operate on any quantum of time. From the experiment results we have demonstrated the applicability of the proposed model to correctly detect multiple appliance usage and make short and long term prediction at high accuracy.

For future work, we are planning to refine the model and introduce distributed learning of big data mining from multiple houses in a near real-time manner. This will help health applications to promptly take actions such as sending alert to patients or care providers. Furthermore, we are planning to build a health ontology model to automatically map discovered appliances to potential activities. This means we can efficiently train the system and increase the accuracy of detecting human activities.

ACKNOWLEDGMENTS

The authors are grateful to the Deanship of Scientific Research, King Saud University, Saudi Arabia, for providing support through the Vice Deanship of Scientific Research Chairs.

REFERENCES

[1] United Nation. (2014). *World Urbanization Prospect*. [Online]. Available: http://dl.acm.org/citation.cfm?id=308574.308676
[2] M. S. Hossain, "Cloud-supported cyber-physical localization framework for patients monitoring," *IEEE Syst. J.*, vol. 11, no. 1, pp. 118–127, Mar. 2017.
[3] M. S. Hossain, G. Muhammad, W. Abdul, B. Song, and B. B. Gupta, "Cloud-assisted secure video transmission and sharing framework for smart cities," *Future Generat. Comput. Syst. J.*, Elsevier 2017, to be published. [Online]. Available: https://doi.org/10.1016/j.future.2017.03.029
[4] J. Liao, L. Stankovic, and V. Stankovic, "Detecting household activity patterns from smart meter data," in *Proc. Int. Conf. Intell. Environ. (IE)*, vol. 6. Jul. 2014, pp. 71–78.
[5] A. Yassine, A. A. N. Shirehjini, and S. Shirmohammadi, "Smart meters big data: Game theoretic model for fair data sharing in deregulated smart grids," *IEEE Access*, vol. 3, pp. 2743–2754, 2015.
[6] A. Yassine and S. Shirmohammadi, "Measuring users' privacy payoff using intelligent agents," in *Proc. IEEE Int. Conf. Comput. Intell. Meas. Syst. Appl.*, May 2009, pp. 169–174.
[7] A. Yassine and S. Shirmohammadi, "A business privacy model for virtual communities," *Int. J. Web Based Commun.*, vol. 5, no. 2, Mar. 2009.
[8] 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. (NBiS)*, 2015, pp. 273–279.
[9] 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, p. 150007, Sep. 2015.
[10] J. Clement, J. Ploennigs, and K. Kabitzsch, "Detecting activities of daily living with smart meters," in *Advance Technology and Societal Change*. Heidelberg, Germany: Springer, 2014, pp. 143–160. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-37988-8_10
[11] Q. Ni, A. B. G. Hernando, and I. P. de la Cruz, "The Elderly's independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development," *Sensors*, vol. 15, no. 5, pp. 11312–11362, 2015. [Online]. Available: http://www.mdpi.com/1424-8220/15/5/11312
[12] C. Chalmers, W. Hurst, M. Mackay, and P. Fergus, "Smart meter profiling for health applications," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2015, pp. 1–7.
[13] M. S. Hossain, "A patient's state recognition system for health care using speech and facial expression," *J. Med. Syst.*, vol. 40, no. 12, pp. 272:1–272:8, Dec. 2016.
[14] M. S. Hossain and G. Muhammad, "Cloud-assisted industrial Internet of Things (IIoT)—Enabled framework for health monitoring," *Comput. Netw.*, vol. 101, pp. 192–202, Jun. 2016.
[15] P. Pouladzadeh, P. Kuhad, S. V. B. Peddi, A. Yassine, and S. Shirmohammadi, "Mobile cloud based food calorie measurement," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2014, pp. 1–6.
[16] S. V. B. Peddi, P. Kuhad, A. Yassine, P. Pouladzadeh, S. Shirmohammadi, and A. A. N. Shirehjini, "An intelligent cloud-based data processing broker for mobile e-health multimedia applications," *Future Generat. Comput. Syst. J.*, vol. 66, pp. 71–86, Jan. 2017.
[17] M. Ul Alam, N. Roy, M. Petruska, and A. Zemp, "Smart-energy group anomaly based behavioral abnormality detection," in *Proc. IEEE Wireless Health (WH)*, Oct. 2016, pp. 1–8.
[18] T. Thomas, C. Cashen, and S. Russ, "Leveraging smart grid technology for home health care," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2013, pp. 274–275.
[19] L. Hawarah, S. E. Ploix, and M. H. Jacomino, in *Proc. Int. Conf. Artif. Intell. Soft Comput. (ICAISC)* Zakopane, Poland, 2010, pp. 372–379. [Online]. Available: http://dx.DOI.org/10.1007/978-3-642-13208-7_47
[20] K. Basu, V. Debusschere, and S. Bacha, "Appliance usage prediction using a time series based classification approach," in *Proc. IEEE 38th Annu. Conf. Ind. Electron. Soc. (IECON)*, May 2012, pp. 1217–1222.
[21] K. Basu, L. Hawarah, N. Arghira, H. Joumaa, and S. Ploix, "A prediction system for home appliance usage," *Energy Buildings*, vol. 67, pp. 668–679, Sep. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378778813000789
[22] 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 (Big Data)*, Apr. 2015, pp. 2223–2232.
[23] K. Gajowniczek and T. Ząbkowski, "Data mining techniques for detecting household characteristics based on smart meter data," *Energies*, vol. 8, no. 7, p. 7407, 2015. [Online]. Available: http://www.mdpi.com/1996-1073/8/7/7407
[24] N. C. Truong, J. McInerney, L. Tran-Thanh, E. Costanza, and S. D. Ramchurn, "Forecasting multi-appliance usage for smart home energy management," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, 2013, pp. 2908–2914. [Online]. Available: http://dl.acm.org/citation.cfm?id=2540128.2540547
[25] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, Dallas, TX, USA, 2000, pp. 1–12. [Online]. Available: http://DOI.acm.org/10.1145/342009.335372
[26] 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
[27] S. Singh and A. Yassine, "Mining energy consumption behavior patterns for households in smart grid," *IEEE Trans. Emerg. Topics Comput.*, to be published.
[28] S. Singh, A. Yassine, and S. Shirmohammadi, "Incremental mining of frequent power consumption patterns from smart meters big data," in *Proc. IEEE Electr. Power Energy Conf. (EPEC)*, Oct. 2016, pp. 1–6.
[29] J. Han, J. Pei, and M. Kamber, "Data mining: Concepts and techniques," in *Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods*, 3rd ed. San Mateo, CA, USA: Morgan Kaufmann, 2011, ch. 6, pp. 243–278. [Online]. Available: http://www.sciencedirect.com/science/book/9780123814791
[30] J. Han, J. Pei, and M. Kamber, "Data mining: Concepts and techniques," in *Cluster Analysis: Basic Concepts and Methods*, 3rd ed. San Mateo, CA, USA: Morgan Kaufmann, 2011, ch. 10, pp. 443–494. [Online]. Available: http://www.sciencedirect.com/science/book/9780123814791
[31] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. 1, pp. 53–65, 1987. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0377042787901257
[32] 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, 2011. [Online]. Available: http://journal.r-project.org/archive/2011-2/RJournal_2011-2_Wang+Song.pdf
[33] I. Ben-Gal, *Bayesian Networks*. Hoboken, NJ, USA: Wiley, 2008. [Online]. Available: http://dx.DOI.org/10.1002/9780470061572.eqr089
[34] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann, 1988.
[35] D. Heckerman, "Bayesian networks for data mining," *Data Mining Knowl. Discovery*, vol. 1, no. 1, pp. 79–119, 1997. [Online]. Available: http://dx.DOI.org/10.1023/A:1009730122752
[36] D. Barber, *Bayesian Reasoning and Machine Learning*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[37] J. Han, J. Pei, and M. Kamber, "Data mining: Concepts and techniques," in *Classification: Advanced Methods*, San Francisco, CA, USA: Morgan Kaufmann, 2011, ch. 9, pp. 243–278. [Online]. Available: http://www.sciencedirect.com/science/book/9780123814791

ABDULSALAM YASSINE received the B.Sc. degree in electrical engineering from Beirut Arab University, Lebanon, in 1993, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Ottawa, Canada, in 2004 and 2010, respectively. Between 2001 and 2013, he was a member of the Technical Staff in the Wireless Communication Division, Nortel Networks, and later at Alcatel-Lucent, Ottawa, Canada. From 2013 to 2016, he was a Postdoctoral Fellow at the DISCOVER Laboratory, School of Electrical Engineering and Computer Science, University of Ottawa. Currently, he is an Assistant Professor with the Department of Software Engineering, Lakehead University, Canada. His research interests are mostly focused on systems and networks, multimedia, artificial intelligence, smart environments, and smart grids.

SHAILENDRA SINGH received the B.Eng. degree in electronics engineering from Pt. Ravishankar Shukla University, Raipur, India. He is currently pursuing the M.A.Sc. degree in electrical and computer engineering from the University of Ottawa, Canada. He has more than 18 years of industry experience and is currently a Research Assistant with 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, smart environments, and cyber-physical systems.

ATIF ALAMRI is currently an Associate Professor with the Information Systems Department, College of Computer and Information Sciences, King Saud University, Saudi Arabia. He is also the Research Chair of Pervasive and Mobile Computing. His research interests include multimedia-assisted health systems, ambient intelligence, and service-oriented architecture. He was a Guest Associate Editor of the IEEE Transactions on Instrumentation and Measurement, a Co-Chair of the first IEEE International Workshop on Multimedia Services and Technologies for E-health, a Technical Program Co-Chair of the 10th IEEE International Symposium on Haptic Audio Visual Environments and Games, and has been or is a Program Committee Member of many conferences in multimedia, virtual environments, and medical applications.