Detection of Anomalies in Activity Patterns of Lone Occupants from Electricity Usage Data

Kuanlong Leong*, Cyril Leung†, Chunyan Miao‡ and Yu Christine Chen§
*†§Department of Electrical and Computer Engineering
The University of British Columbia (UBC), Canada
†‡ Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly
Nanyang Technological University (NTU), Singapore
Email: {*kuanlongl, †cleung, §chen}@ece.ubc.ca, ‡ascymiao@ntu.edu.sg

***Abstract*—As the global population ages, assisted living technologies for the elderly are becoming more popular. A person normally performs activities of daily living (ADLs) on a regular basis. A person's ability to perform recurring ADLs indicates the person's wellness. Anomalies in activity patterns of a person might indicate changes in the person's wellness. A method is proposed in this paper for detecting anomalies in activity patterns of a lone occupant using his/her electricity consumption data. The proposed method infers anomalies in activity patterns of an occupant from electricity consumption patterns instead of explicitly monitoring the underlying individual activities. The proposed method provides a score which is a quantitative assessment of anomalies in electricity consumption pattern of an occupant for a given day. A survey was conducted to obtain the hourly activities of three lone occupants for a month. From the survey, the level of suspicion values which are quantitative assessments of anomalies in activity patterns of the occupants were deduced. Using Fuzzy C-Means (FCM) clustering with Euclidean distance measure, the scores and level of suspicion values were clustered respectively. A day was then classified as regular or irregular from an electricity consumption perspective (score) and an activity perspective (level of suspicion value) respectively. Our results show that anomalies in electricity consumption patterns correlate well with anomalies in the underlying activity patterns.**

I. Introduction

The global population share of people aged 60 and over is expected to increase from 11.7% in 2013 to 21.1% in 2050 [1]. The number of elderly people is expected to reach 2 billion in 2050 from 841 million in 2013. The demand for elderly care facilities and services, such as nursing homes, is expected to increase with the ageing population. A U.S. survey found that 87% of people aged 65 and over prefer to age in their own homes [2]. Thus, there is a need to ensure a safe and independent ageing environment for these elderly people, and to help them live in their preferred environments for as long as possible.

Assisted living technologies for detecting medical emergencies and assessing the wellness of the elderly are becoming more popular [3]. While the most direct way of detecting medical emergencies is arguably by monitoring physiological data such as heart rate or blood pressure, estimating the wellness of a person usually involves monitoring the activities of daily living (ADLs). An important component of assisted living technologies is activity recognition and monitoring [3].

A person normally performs ADLs on a regular basis. The ability to perform ADLs regularly implies that the person is at least physically able to maintain a regular lifestyle. It also indicates that the wellness of the person is at a certain level. Large deviations from a regular daily routine may indicate changes in the person's ability to perform ADLs. Such deviations may be used to alert relatives or caregivers to look into the cause of the deviations in a timely manner.

To establish the regular activity patterns of a person, a number of approaches have been proposed to monitor the individual ADLs [3]. Individual activities are commonly monitored using ambient sensors (e.g., motion sensors and force sensors) placed smartly in the monitored environment or by cameras. Individual activities are then recognized as sequences of sensor events or sequences of images. The regular activity patterns of a person can then be established based on the sequences of recognized activities. However, interpreting events/activities captured by sensors could be challenging and time-consuming [3].

ADLs usually involve using electric home appliances. Activities that consume electricity might be inferred from the household electricity consumption patterns. If a person's activity patterns are exactly the same every day, the electricity consumption patterns will also probably be the same every day. The electricity consumption patterns of a monitored environment would reflect the activity patterns of the person. If the activity patterns of a person can be sufficiently represented by the electricity consumption patterns, then these could be used to detect anomalies in the underlying activities and there might be no need to explicitly monitor the underlying individual activities.

In this paper, we propose a method for detecting anomalies in activity patterns of a lone occupant by monitoring the household electricity consumption. This work differs from related works in one fundamental perspective. This work intends to detect anomalies in activity patterns of a person without explicitly monitoring the individual activities or actions of the person. Instead of attempting to use the raw electricity consumption data for detection, we use features extracted from the raw electricity consumption data. The extracted features are designed to effectively reflect the underlying activities.

This work aims to show that anomalies in activity patterns of an occupant can be effectively inferred from electricity consumption measurements of his/her residence. Using the proposed method, we classify a day as regular or irregular from an electricity consumption perspective. As a benchmark, we classify the same day as regular or irregular from an activity perspective. Finally, we compare the two classification results, and find that anomalies in activity patterns can be effectively inferred from electricity consumption data.

II. Related Works

Various assisted living technologies have been proposed to alleviate health problems, detect medical emergencies and improve the wellness of the elderly [4]–[20]. One important assisted living technology is health status monitoring. Physiological parameters are possibly the best indicators of the health status of a person. Practical approaches proposed in [4]–[7] can monitor the physiological parameters such as electrocardiogram signals, blood glucose concentrations, blood pressure, body temperature, heart rate and respiratory rate using ZigBee or Bluetooth for communication.

Another important assisted living technology is fall detection. Fall detection approaches can be divided into two categories: wearable and non-wearable. Approaches using wearable accelerometers were proposed in [8], [9]. How the wearable device is worn and the person's willingness to wear the device are critical to the effectiveness of this approach. Audio-based [10], [11] and visual-based [12], [13] non-wearable approaches have also been proposed. Audio-based approaches can be adversely affected by background noise while visual-based approaches can be harmfully affected by occlusions.

Another important assisted living technology is activity recognition and monitoring. Activity recognition helps determine what daily activities are essential to a person. Long term monitoring helps determine the regular activity patterns of a person. That a person is able to perform recurring ADLs indicates that the person's wellnes is at a certain level. Activity recognition approaches can generally be categorized as video-based [14] or sensor-based [15]–[20]. In [15]–[19], the authors designed a wireless sensor network which can interpret the wellness of a person by monitoring various home accessories such as the bed, microwave oven, toilet, dining chair, etc. The wellness of a person was interpreted according to how well the monitored person performed the essential daily activities in terms of the home appliances' active and inactive durations. The difficulties of this approach are as follows: 1) determining a sufficient number of sensors for monitoring ADLs, 2) storing the sensor data efficiently, 3) annotating the activities deduced from the sensor data, and 4) classifying regular and irregular activities accurately.

In [20], the authors proposed to detect anomalies in activity patterns of a person based on temporal relations between events captured by some sensors such as motion sensors and light sensors. For instance, if event A always occurs before event B but the recurring temporal relation between A and B is violated on a particular day, it would be noted as an anomaly. It would be impractical to investigate all events that occur during a day and hence the authors only focused on the temporal relations between the most frequent events. However, annotating the events captured by the sensors and identifying the temporal relations between events could still be time-consuming.

This work proposes to infer anomalies in activity patterns of a person from the household electricity consumption patterns. Electricity consumption data over a period of time is essentially a time series or a sequence. A straightforward way would be to compare the electricity consumption sequences with one another using some similarity measure. Similar sequences are assigned to the same cluster and there could be multiple clusters. Each cluster can then be assigned a class label (e.g., regular or irregular) if there is sufficient information. Approaches to cluster time series data can be divided into three categories: 1) raw data-based, 2) feature-based, and 3) model-based [21].

However, most similarity measures are too sensitive to slight changes in the raw hourly electricity consumption data. For instance, two 24-hour hourly electricity consumption sequences can be clustered into two different clusters due to trivial differences even though the underlying activity patterns are almost the same. Therefore, applying a clustering technique such as Fuzzy C-Means (FCM) clustering [22] to raw electricity consumption sequences for inferring anomalies in activity patterns might not be effective. We observed that some key features (e.g., maximum, minimum) of electricity consumption patterns are more relevant to the underlying activity patterns. In other words, the key features could be better indicators of the underlying activity patterns than the raw electricity consumption data. Therefore, it might be better to work with features selected or extracted from the raw electricity consumption data for detecting anomalies in activity patterns.

III. Proposed Method for Assessing Anomalies in Electricity Consumption Data Based on FCM Clustering

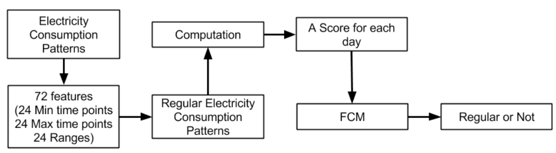


Fig. 1. A flowchart of the proposed method based on FCM clustering

In this section, we propose a method for assessing anomalies in electricity consumption data based on FCM clustering. The proposed method classifies a day as regular or irregular from an electricity consumption perspective for lone occupants. First, 72 selected features are extracted from the raw daily electricity consumption data. Next, regular electricity consumption patterns are deduced from a collection of the extracted features over multiple days (e.g., 30 days). Third, a model is formed and the 72 features extracted from the electricity consumption pattern of a given day are converted to a quantitative score. After computing the scores for multiple days (e.g., 30 days), FCM clustering is applied to the scores for learning the threshold value between the higher scores and the lower scores. A day is deemed regular if its corresponding score is above the threshold. In contrast, a day is deemed irregular if its corresponding score is below the threshold. A flowchart of the proposed method based on clustering is shown in Fig. 1.

A. Datasets

The proposed method is based on hourly electricity consumption data. Data used in this work were collected by the BC Hydro smart meters [23]. There were three participating lone occupants (two elderly people and one young adult) in this work. The hourly electricity consumption data were simply downloaded by each participant from the BC Hydro website. A summary of the three datasets is shown in Table I.

TABLE I
SUMMARY OF THE ELECTRICITY CONSUMPTION DATASETS

Home	# of Occupant	Type of Home	Data Length (days)
A	1	Apartment	30
B	1	Apartment	365
C	1	Detached	365

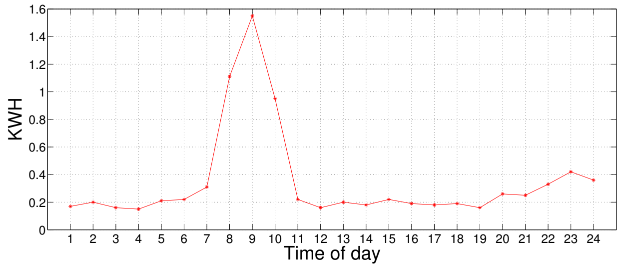
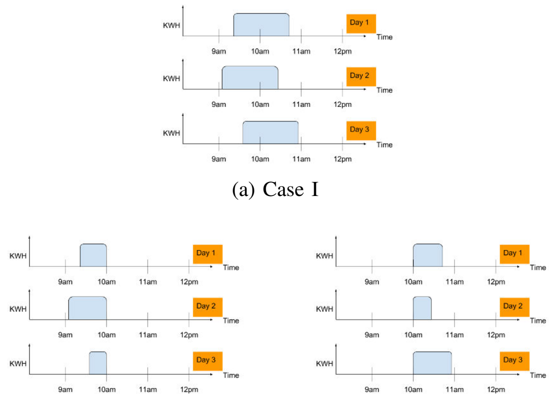


Fig. 2. The typical hourly electricity consumption pattern of Home C during a day



(a) Case I

(b) Case II

(c) Case III

Fig. 3. An illustration of a regular electricity consumption pattern

B. Electricity Features

Features are needed to represent the regular (recurring) electricity consumption patterns of an occupant. The basic assumption made in this work is that people perform ADLs on a regular basis. Many of the ADLs involve using electric home appliances. It is therefore plausible that these activities may be inferred from the electricity consumption data. For instance, the occupant of Home C normally wakes up at 8AM, uses electric cooking appliances right away for two to three hours, is away from home from 11AM to 9PM, and goes to bed at 11PM. The hourly electricity consumption pattern of a typical day at Home C is shown in Fig. 2. As can be seen from Fig. 2, the electricity consumption pattern matches quite well the activity pattern. This provides strong evidence that the activity patterns of the occupant are reflected in the electricity consumption patterns. Instead of attempting to recognize the underlying individual activities from the electricity consumption data, the proposed method uses electricity consumption patterns to represent activity patterns of the occupant without explicitly considering the underlying individual activities.

The datasets in this work provide the hourly electricity consumption data of each home. However, an activity is not limited to begin and end within a single pre-framed hourly interval in the datasets. An activity can begin at any time in one hourly interval and end at any time in another hourly interval. For instance, if there is unrealistically only one given activity that can occur within two particular consecutive hours and it consumes the exact same amount of electricity every day, the amount of electricity consumed during the two hours will always be the same every day as illustrated in Fig. 3a. This is an example of a regular electricity consumption pattern. However, this pattern might not be observable if we only consider the amounts of electricity consumed during individual hours as illustrated in Fig. 3b and Fig. 3c.

It can easily be seen that a similar argument can also be made for any multiple consecutive hours. This suggests that considering the multiple-hour total consumption can help reflect the underlying activity patterns of the occupant. Therefore, at the end of each hour, the proposed method calculates the total electricity consumption for the previous two to 24 hours (moving total electricity consumption). Consequently, at the end of each hour, we not only have the amount of electricity consumed for the previous one hour, but also the amounts of the total electricity consumption for the previous two to 24 hours, as illustrated in Fig. 4. This step of the proposed method does not add any new information to the datasets; it simply computes the 2-hour to 24-hour moving total curves using the hourly consumption data in the datasets.

The raw electricity consumption data might not be the best representations of the underlying activity patterns because unimportant parts of the raw electricity patterns might adversely affect the detection of anomalies in the underlying activity patterns. Although the activity patterns of two given days are the same, their corresponding consumption patterns may not be exactly the same. Detecting anomalies may be

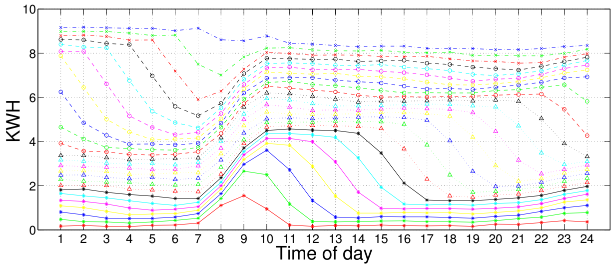


Fig. 4. The 1-hour to 24-hour moving total electricity consumption for one day. Each dot on the bottom curve represents the consumption for the previous one hour, each dot on the second curve from the bottom represents the total consumption for the previous two hours, and so on.

ineffective if we compare one raw consumption pattern to another using some conventional distance measure such as Euclidean distance measure. Rather, we should choose features which are more representative of the underlying activities from the raw electricity consumption patterns.

The local maxima and local minima of electricity consumption patterns are likely to be good indicators of the active and inactive hours of an occupant. Therefore, the times at which the maxima and minima occur may be good indicators of the underlying activity patterns. The difference between the maximum and minimum consumptions might indicate the electricity consumption of the relevant activities. Therefore, the features chosen to represent daily electricity consumption patterns are: 1) the time at which the maximum occurs for each moving total curve, 2) the time at which the minimum occurs for each moving total curve, and 3) the range (difference between the maximum and minimum) of each moving total curve.

In the case that the extremum (maximum or minimum) is not unique, one time point among the multiple time points sharing the same extremum is chosen uniformly at random. The features chosen to represent the electricity consumption pattern of an occupant for a day are the 24 time points of maxima, 24 time points of minima, and the 24 ranges.

One set of the 72 features obtained from one day only represents the electricity consumption pattern on that particular day. To deduce the regular electricity consumption patterns of an occupant, these features need to be collected over multiple days (e.g., 30 days).

C. Regular Electricity Consumption Patterns

Regular electricity consumption patterns of an occupant are deduced from a collection of the electricity features over multiple days. For the proposed method, the regular consumption patterns of an occupant are described by 1) how likely a maximum or minimum of a moving total curve occurs at a particular time, and 2) how likely the range (maximum minus minimum) of a moving total curve happens to have a certain value. The time at which a maximum or minimum consumption normally occurs reflects the active and inactive hours of the occupant. The normal range of a moving total curve indicates the typical total electricity consumption of the relevant activities.

Using the maximum time points and the minimum time points of each day over multiple days (e.g., 30 days), the probability of each time point (hour) being the maximum or minimum of each moving total curve can be approximated. The pseudocode for estimating the occurrence probability of the maximum (minimum) at each hour for the hourly consumption curve is given in Algorithm 1.

Algorithm 1 The pseudocode for estimating the occurrence probability of the maximum (minimum) at each hour for the hourly consumption curve

Step 1. Obtain the hourly consumption data for each hour for one day (i.e., the bottom curve in Fig. 4)
Step 2. Record the time point (hour) at which the maximum (minimum) consumption occurs
Step 3. Repeat **Steps 1** and **2** for each day over multiple days
Step 4. Estimate the probability that the maximum (minimum) would occur at 1AM according to the records
Step 5. Repeat **Step 4** for the remaining 23 time points

To calculate the occurrence probability of the maximum (minimum) at each time point (hour) for the 2-hour moving total curve, Algorithm 1 obtains the *2-hour total electricity consumption* for each hour for one day (i.e., the second to bottom curve in Fig. 4) in **Step 1**. **Step 2** to **Step 5** are then performed. Corresponding modifications to **Step 1** of Algorithm 1 can be made in order to estimate the occurrence probabilities of the maxima (minima) for the 3-hour to 24-hour moving total curves.

TABLE II
PART OF A SAMPLE MAXIMA PROBABILITY MATRIX. EACH ROW CORRESPONDS TO A MOVING TOTAL WHEREAS EACH COLUMN CORRESPONDS TO A TIME POINT (HOUR) OF A DAY

ProbMax	...	6	7	8	9	10	11	12	...
1	...	0	0	0.47	0.47	0.03	0	0	...
2	...	0	0	0.03	0.63	0.3	0	0	...
3	...	0	0	0	0.13	0.77	0.07	0	...
4	...	0	0	0	0.07	0.4	0.5	0	...
5	...	0	0	0	0.03	0.2	0.53	0.2	...
6	...	0	0	0	0.03	0.2	0.37	0.23	...
7	...	0	0	0	0.03	0.07	0.3	0.33	...
8	...	0	0	0	0.03	0.03	0.27	0.27	...
9	...	0	0	0.03	0	0.07	0.13	0.23	...
...	...	...	...	...	...	...	...	...	...

Following the previously described procedure, a total of $24 \times 24 = 576$ occurrence probabilities are obtained for the maxima and an equal number for the minima. These probabilities can be arranged in two 24×24 matrices, one for the maxima and one for the minima. Part of a sample maxima probability matrix is illustrated in Table. II, where each row represents a moving total and each column represents a time point. For instance, it shows that there is a 63% chance that the maximum of the 2-hour moving total curve occurs at 9AM. The minima probability matrix is not shown here, but it can be interpreted in a similar manner. These two matrices help represent the regular consumption patterns of an occupant and they can indicate the normal active and inactive hours of the occupant.

Using the ranges of the moving total curves over multiple days, the mean and standard deviation of the range of each moving total curve can be computed. The pseudocode for estimating the mean and standard deviation of the range of the hourly consumption curve is given in Algorithm 2.

Algorithm 2 The pseudocode for estimating the mean and standard deviation of the range of the hourly consumption curve

Step 1. Obtain the hourly consumption for each hour for one day (i.e., the bottom curve in Fig. 4)
Step 2. Obtain the maximum and minimum
Step 3. Compute the range (maximum minus minimum) and record it
Step 4. Repeat **Steps 1** to **3** for each day over multiple days
Step 5. Compute the mean and standard deviation of the recorded ranges

To compute the mean and standard deviation of the range of the 2-hour moving total curve, we simply use the *2-hour total electricity consumption* for each hour for one day (i.e., the second to bottom curve in Fig. 4) in **Step 1** of Algorithm 2. Corresponding modifications to **Step 1** of Algorithm 2 can be made in order to compute the means and standard deviations of the ranges of the 3-hour to 24-hour moving totals.

Following the previously described procedure, we will then have the means and standard deviations of the ranges of all 24 moving total curves. The means and standard deviations of the ranges help represent the regular consumption patterns of an occupant and they provide an indication of the normal electricity consumptions of the underlying activities.

To sum up, the regular (recurring) electricity consumption patterns of an occupant are numerically represented by the two probability matrices (one for the minima and one for the maxima) and the 24 means and standard deviations.

D. Model for Providing Quantitative Scores

A model can then be formed based on the two probability matrices, the means and the standard deviations. The model is used to detect anomalies in electricity consumption patterns of an occupant and to provide quantitative scores of the detected anomalies. To quantitatively assess the anomalies in the electricity consumption pattern of a given day, the electricity features of the given day need to be provided. As discussed in Section III-B, the electricity features include the 24 maximum time points, 24 minimum time points and 24 ranges.

E. Score for a Day

First, the 24 maximum (minimum) time points of a given day are converted to sub-scores as described in Algorithm 3 (see Table III for the threshold values). Following the procedure described in Algorithm 3, the 24 maximum time points and 24 minimum time points will be converted to 48 sub-scores.

Algorithm 3 The pseudocode for converting the 24 maximum (minimum) time points of a given day into sub-scores

Step 1. Obtain the time of day at which the maximum (minimum) occurs for the hourly consumption curve
Step 2. Retrieve the occurrence probability of that time point being the maximum (minimum) from the maxima (minima) probability matrix of the model
Step 3. Return a *positive score* (e.g., +1) if the probability is greater than or equal to a *pre-defined threshold*; otherwise, return a *negative score* (e.g., -1)
Step 4. Obtain the time of day at which the maximum (minimum) occurs for each of the remaining 23 moving total curves and repeat **Steps 2** and **3**

Next, the 24 ranges of the given day are converted to sub-scores as described in Algorithm 4 (see Table III for the threshold values).

Algorithm 4 The pseudocode for converting the 24 ranges of a given day into sub-scores

Step 1. Obtain the range of the hourly consumption curve of the given day
Step 2. Retrieve the mean and standard deviation of the range previously computed as described in Algorithm 2
Step 3. Calculate the Z-Score (standardized value) of the range of the given day
Step 4. Return a *positive score* (e.g., +1) if the absolute Z-Score is less than or equal to a *pre-defined threshold*; otherwise, return a *negative score* (e.g., -1)
Step 5. Obtain the range of each of the remaining 23 moving total curves and repeat **Steps 2** to **4**

A Z-Score (also know as standard score) measures the distance between an observation and its mean in terms of number of standard deviations [24]. Following the above procedure, the 24 ranges will be converted into 24 sub-scores.

The final score for a given day, which is a quantitative assessment of the anomalies detected in the electricity consumption pattern, is the sum of the 72 sub-scores. A more positive final score for a day indicates that the consumption pattern on that day has less deviation from the regular consumption patterns. Conversely, a more negative final score for a day indicates that there is a large deviation from the regular consumption patterns.

F. Design Variables

To deduce the regular electricity consumption patterns and form a model for assessing anomalies in electricity consumption patterns for an occupant, we need to determine the *number* *of days of consumption data* to be included. To compute the score for a day, we need to determine the *threshold values* and the *magnitudes of the sub-scores* for converting the maximum time points, minimum time points and ranges into 72 sub-scores. The design variable values are set manually for each participating home by observation. The actual configurations for Homes A, B and C in this work are shown in Table III.

TABLE III
CONFIGURATIONS OF THE DESIGN VARIABLES FOR HOMES A, B AND C

	Home A	Home B	Home C
Length of Data (days)	30	60	60
Probability Threshold			
-Maximum time point	0.1	0.1	0.15
-Minimum time point	0.15	0.1	0.15
Score Assignment	+/-1	+/-1	+/-1
Range Z-Score Threshold	1.5	1	1
Score Assignment	+/-1	+/-1	+/-1

G. Clustering of the Scores Based on FCM Clustering

After computing the scores for a period of time (e.g., 30 days) for an occupant, we use FCM clustering for learning the threshold between the lower scores and the higher scores. A low score for a day indicates that there is a large deviation from the regular electricity consumption patterns.

First, FCM clustering with Euclidean distance measure is applied to the scores. The scores are then clustered into two clusters: the *regular* cluster for the higher scores and the *irregular* cluster for the lower scores. Next, the boundary (threshold) between the two clusters is defined as the average of the two centers. The center of a cluster is defined as the average of all scores belonging to the cluster.

H. Classification of the Daily Electricity Consumption Patterns

A *regular* score for a day indicates that the corresponding daily electricity consumption pattern is regular. Whereas, an *irregular* score for a day indicates that the corresponding daily electricity consumption pattern is irregular. A day is then classified as regular or irregular from an electricity consumption perspective.

A day is regular if the score for the day is above the threshold; otherwise, a day is irregular. To validate the classification results from this electricity consumption perspective, we have conducted a survey for classifying a day as regular or irregular from an activity perspective. We aim to show that anomalies detected in electricity consumption patterns by the proposed method can effectively reflect anomalies in activity patterns.

IV. SURVEY FOR VALIDATING THE EFFECTIVENESS OF THE PROPOSED METHOD

To validate the correlation between the household electricity consumption data and activity patterns of an occupant, a survey was conducted to obtain the hourly activities of three lone occupants for one month. Each occupant was asked to record his/her hourly activities on a timesheet. Regular activity patterns were deduced from the activity data for each occupant. A level of suspicion value for a day, which indicates the anomalies in activity patterns, was then computed. FCM clustering was then applied to a collection of the level of suspicion values. A day was then classified as regular or irregular from an activity perspective. A flowchart of the procedure for classifying a day as regular or irregular from an activity perspective is shown in Fig. 5.

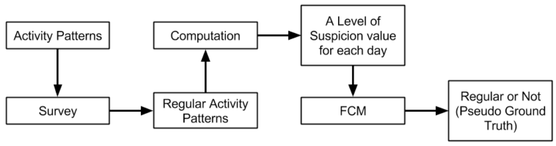


Fig. 5. A flowchart of the procedure for classifying a day as regular or irregular from an activity perspective

A. Timesheet

A part of the survey timesheet is shown in Table. IV. Listed activities on the timesheet includes the essential ADLs such as sleeping and dining. Each occupant was asked to mark the listed activities that took place during any part of each hour on the timesheet every day for one month.

TABLE IV
A PART OF THE SURVEY TIMESHEET

	00:01-01:00	01:01-02:00	02:01-03:00	... ...	23:01-00:00
Not at home					
Sleeping					
Bathing					
Dining at home					
Electric cooking appliances (oven, rice cooker, etc.)					
Dishwasher					
House cleaning (vacuum cleaner, etc.)					
Entertainment (TV, computer, radio, etc.)					
Laundry (washer, dryer, iron, etc.)					
Heater /Air conditioner					
Other energy-consuming activities/appliances*					

B. Regular Activity Patterns

After collecting the hourly activity data, regular (recurring) activity patterns of each occupant were represented by the following six features:

1) Daily highly probable activities (DHPA)
2) Daily less probable activities (DLPA)
3) Hourly highly probable activities (HHPA)

4) Hourly less probable activities (HLPA)
5) Daily less probable durations of activities (DLPD)
6) Less probable daily energy consumption (LPDEC)

Classifying an activity and a duration of a given activity as probable or not probable involves determining some threshold values. The actual configurations of the threshold values for each home are shown in Table V.

TABLE V
THRESHOLD VALUES FOR REGULAR ACTIVITY PATTERNS FOR HOMES A, B AND C

Feature	Home A	Home B	Home C
	Probability		
DHPA	$\geq$0.8	$\geq$0.8	$\geq$0.9
DLPA	$\leq$0.2	$\leq$0.2	$\leq$0.2
HHPA	$\geq$0.9	$\geq$0.8	$\geq$0.8
HLPA	$\leq$0.05	$\leq$0.05	$\leq$0.05
DLPD	$\leq$0.05	$\leq$0.05	$\leq$0.05
	$\lvert Z$-Score$\rvert$		
LPDEC	$\geq$1	$\geq$0.95	$\geq$1.1

TABLE VI
A LIST OF POTENTIAL THRESHOLD VALUES FOR HOME B

Feature	Potential Threshold Values
	Probability
DHPA	$\geq$ 0.8 / 0.9
DLPA	$\leq$ 0.2 / 0.1
HHPA	$\geq$ 0.8 / 0.85 / 0.9 / 0.95
HLPA	$\leq$ 0.15 / 0.1 / 0.05
DLPD	$\leq$ 0.15 / 0.1 / 0.05
	|Z-Score|
LPDEC	$\geq$ 0.9 / 0.95 / 1 / 1.2 / 1.3

C. Level of Suspicion

A level of suspicion value for a day was then calculated. The level of suspicion value is a quantitative assessment of anomalies in a daily activity pattern. The initial level of suspicion value for a day is zero. An absence of a highly probable activity pattern or an occurrence of a less probable activity pattern will increase the level of suspicion value for the given day by a certain number. A high level of suspicion value of a day indicates a large deviation from the regular activity patterns.

The best case scenario is to have only a few days with noticeably higher level of suspicion values and to have most of the other days with lower level of suspicion values. In order to achieve this, combinations of the six threshold values need to be considered. For example, a list of the potential threshold values for Home B are shown in Table VI. The potential threshold values were chosen by observation. The accumulated level of suspicion value for a day is equal to the sum of the level of suspicion values raised by the above six features. If the daily activity pattern of a certain day is significantly different from the regular activity patterns, the accumulated level of suspicion value of that particular day will probably be raised by most of the six features. If the (accumulated) level of suspicion value of a day is relatively low and is only raised by one feature, it probably indicates that the threshold value for that particular feature can be tightened (i.e., select the next threshold value on the list).

To select the appropriate threshold values for the six features, the selection process begins with the loosest values (the leftmost values) on the list. The next threshold value on the list for each feature is chosen if that helps differentiate the days that are currently having noticeably higher level of suspicion values from the other days. The selection process finishes if one of the following events is encountered:

1) The selected set of threshold values is already the tightest possible combination (i.e., the rightmost values on the list).
2) Choosing the next threshold value on the list for any of the six features would significantly harm the differentiation between the highly suspicious (possibly anomalous) days and the possibly normal days.

D. Clustering of the Level of Suspicion Values Based on FCM Clustering

After computing the daily level of suspicion values for the survey period for each occupant, we again used FCM clustering for learning the threshold between the lower values and the higher values. First, FCM clustering with Euclidean distance measure was applied to the level of suspicion values. The values were then clustered into two clusters: the *regular* cluster for the lower values and the *irregular* cluster for the higher values. Next, the boundary (threshold) between the two clusters was defined as the average of the two centers.

E. Pseudo Ground Truth

A *regular* level of suspicion value for a day indicates that the corresponding daily activity pattern is regular. Whereas, a *irregular* level of suspicion value for a day indicates that the corresponding daily activity pattern is irregular. A day was then classified as regular or irregular from an activity perspective. A day is regular if the level of suspicion for the day is below the threshold; otherwise, a day is irregular. The classification results from this activity perspective are used as the pseudo ground truth in this work.

V. EXPERIMENT

This work aims to show that anomalies in activity patterns of an occupant can be inferred from electricity consumption measurements of his/her residence. First, we obtained the activity data of three lone occupant from a survey. We then classified each day of the survey period as regular or irregular from an activity perspective for each occupant. The classification results from this activity perspective were used as the pseudo ground truth. Next, we used our proposed method for classifying each day of the same period as regular or irregular from an electricity consumption perspective. Finally, we compared the classification results provided by the proposed method with the pseudo ground truth.

The proposed method requires extra steps for extracting features from the raw consumption data and computing scores before the FCM clustering. The additional steps would be questionable if the proposed method does not perform better than a raw data-based approach. In [25], it was found that the Euclidean distance measure is the overall best similarity measure for clustering hourly energy consumption data for five university buildings among the other three well-known measures (Mahalanobis distance [26], Dynamic Time Warping distance [27] and Pearson's correlation). Therefore, we chose to compare the proposed method with a raw data-based approach based on FCM clustering and Euclidean distance measure for showing that the extra steps of the proposed method help improve the accuracy performance. The FCM clustering process terminates if one of the following conditions is met:

1) The number of iterations has reached 100.
2) The improvement in the objective function J_m is less than 10^{-5} between two consecutive iterations.

The objective function J_m, which is to be minimized in each iteration, is

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{K} {\mu_{ij}}^m \left|\left|x_i - c_j\right|\right|^2 \qquad (1)$$

where N is the number of data points, K is the number of clusters, μ_{ij} is the degree of membership of data point x_i in cluster c_j, and m is the fuzzy exponent which was chosen to be 2 in this experiment. A flowchart of the experiment is shown in Fig. 6.

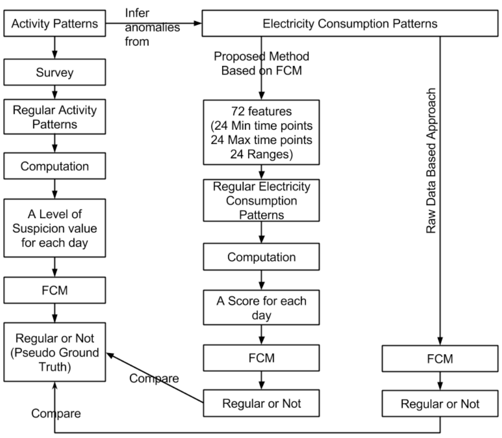


Fig. 6. A flowchart of the experiment

VI. Performance Evaluation

There is a total of 91 classified days based on the pseudo ground truth for the three participating occupants in this work. We used 61 days for training and the remaining 30 days for testing. The training set was chosen by observation so that it contains some days that are likely irregular . The average performances of the proposed method and the raw data-based approach for the three homes are summarized in Table VII. Sensitivity indicates the proportion of irregular days that were correctly labelled by the proposed method or the raw data-based approach. Specificity indicates the proportion of regular days that were correctly labelled by the proposed method or the raw data-based approach.

TABLE VII
THE PERFORMANCES OF THE PROPOSED METHOD AND THE RAW DATA-BASED APPROACH

	Proposed Method	Raw Data-Based Approach
Training Set		
Accuracy	57/61 (93.44%)	54/61 (88.52%)
Sensitivity	13/14 (92.86%)	13/14 (92.86%)
Specificity	44/47 (93.62%)	41/47 (87.23%)
Test Set		
Accuracy	26/30 (86.67%)	22/30 (73.33%)
Sensitivity	4/5 (80%)	2/5 (40%)
Specificity	22/25 (88%)	20/25 (80%)

The results show that anomalies in electricity consumption patterns detected by the proposed method correlate well with anomalies in activity patterns. In other words, anomalies in activity patterns can be effectively inferred from electricity consumption patterns by the proposed method. The results also show that the proposed method outperforms the raw data-based approach. It indicates that the features extracted from the raw consumption data better represent the underlying activities than the raw consumption data. The proposed method yields a better accuracy performance as compared to the raw data-based approach at the expense of the complexity for extracting features from electricity consumption data and computing scores which assess anomalies detected. Our results suggest that the correlation between anomalies in electricity consumption patterns and anomalies in activity patterns are more significant for the two elderly people than the young adult. This is likely because the two elderly people tend to have activity patterns with much less variation from day to day; therefore, what *regular patterns* are is clearer to the proposed method.

As can be seen from Table VII, the sensitivity of the raw-data based approach drops drastically in the testing phase as compared to the training phase. It might indicate that the classifier for the raw data-based approach fits the training data too tightly (i.e., overfitting). Adopting a cross validation technique might help reduce the overfitting effect on the performance evaluation. However, the amount of data available for this experiment is relatively small and the dataset is highly unbalanced (i.e., there are more regular days than irregular days). Adopting a cross validation technique for the performance evaluation would likely result in generating training sets which have no irregular day. Therefore, such validation approach was not adopted in this experiment.

VII. Conclusion

In this paper, we proposed a method based on FCM clustering for inferring anomalies in activity patterns from electricity consumption data without the need for explicitly monitoring individual activities. Our results show that anomalies in activity patterns of an lone occupant can be effectively inferred from the electricity consumption measurements of his/her residence. The proposed method does not identify whether the anomalies detected indicates a positive or negative change in the occupant's wellness. However, a detected anomaly can be used to trigger an alert to caregivers or relatives who can then investigate in a timely manner.

Acknowledgment

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Grant RGPIN 1731-2013, by the UBC Faculty of Applied Science, and by the National Research Foundation Singapore under its Interactive Digital Media (IDM) Strategic Research Programme.

References

[1] United Nations, Department of Economic and Social Affairs, Population Division, "World population ageing 2013," 2013.

[2] R. Harrell, J. Lynott, S. Guzman, and C. Lampkin, "What is livable? community preferences of older adults," *American Association of Retired Persons (AARP) Public Policy Institute*, April 2014.

[3] P. Rashidi and A. Mihailidis, "A survey on ambient-assisted living tools for older adults," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 579–590, 2013.

[4] B. Liu, Y. Zhang, and Z. Liu, "Wearable monitoring system with multiple physiological parameters," in *Medical Devices and Biosensors (MDBS), 5th International Summer School and Symposium on*, 2008, pp. 268–271.

[5] E. Sardini, M. Serpelloni, and M. Ometto, "Multi-parameters wireless shirt for physiological monitoring," in *Medical Measurements and Applications Proceedings (MeMeA), IEEE International Workshop on*, 2011, pp. 316–321.

[6] K. Malhi, S. C. Mukhopadhyay, J. Schnepper, M. Haefke, and H. Ewald, "A zigbee-based wearable physiological parameters monitoring system," *IEEE Sensors Journal*, vol. 12, no. 3, pp. 423–430, 2012.

[7] R. Logier *et al.*, "A multi sensing method for robust measurement of physiological parameters in wearable devices," in *Engineering in Medicine and Biology Society (EMBS), 36th Annual International Conference of the IEEE on*, 2014, pp. 994–997.

[8] J. Chen, K. Kwong, D. Chang, J. Luk, and R. Bajcsy, "Wearable sensors for reliable fall detection," in *Engineering in Medicine and Biology Society (EMBS), 27th Annual International Conference of the IEEE on*, 2006, pp. 3551–3554.

[9] T. Tamrat, M. Griffin, S. Rupcic, S. Kachnowski, T. Taylor, and J. Barfield, "Operationalizing a wireless wearable fall detection sensor for older adults," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 6th International Conference of the IEEE on*, 2012, pp. 297–302.

[10] M. Popescu, Y. Li, M. Skubic, and M. Rantz, "An acoustic fall detector system that uses sound height information to reduce the false alarm rate," in *Engineering in Medicine and Biology Society (EMBS), 30th Annual International Conference of the IEEE on*, 2008, pp. 4628–4631.

[11] X. Zhuang, J. Huang, G. Potamianos, and M. Hasegawa-Johnson, "Acoustic fall detection using gaussian mixture models and GMM supervectors," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP) of the IEEE on*, 2009, pp. 69–72.

[12] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Robust video surveillance for fall detection based on human shape deformation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 21, no. 5, pp. 611–622, 2011.

[13] V. Vaidehi, K. Ganapathy, K. Mohan, A. Aldrin, and K. Nirmal, "Video based automatic fall detection in indoor environment," in *International Conference on Recent Trends in Information Technology (ICRTIT) of the IEEE on*, 2011, pp. 1016–1020.

[14] Z. Zhou, X. Chen, Y.-C. Chung, Z. He, T. X. Han, and J. M. Keller, "Activity analysis, summarization, and visualization for indoor human activity monitoring," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 11, pp. 1489–1498, 2008.

[15] N. Suryadevara, A. Gaddam, S. Mukhopadhyay, and R. Rayudu, "Wellness determination of inhabitant based on daily activity behaviour in real-time monitoring using sensor networks," in *Sensing Technology (ICST), 5th International Conference of the IEEE on*, 2011, pp. 474–481.

[16] N. K. Suryadevara and S. C. Mukhopadhyay, "Wireless sensor network based home monitoring system for wellness determination of elderly," *IEEE Sensors Journal*, vol. 12, no. 6, pp. 1965–1972, 2012.

[17] N. Suryadevara, S. Mukhopadhyay, R. Wang, R. K. Rayudu, and Y. Huang, "Reliable measurement of wireless sensor network data for forecasting wellness of elderly at smart home," in *Instrumentation and Measurement Technology Conference (I2MTC) of the IEEE on*, 2013, pp. 16–21.

[18] N. Suryadevara and S. Mukhopadhyay, "An intelligent system for continuous monitoring of wellness of an inhabitant for sustainable future," in *Humanitarian Technology Conference (R10-HTC) of the IEEE on*, 2014, pp. 70–75.

[19] N. K. Suryadevara and S. C. Mukhopadhyay, "Determining wellness through an ambient assisted living environment," *IEEE Intelligent Systems*, vol. 29, no. 3, pp. 30–37, 2014.

[20] V. Jakkula, D. J. Cook *et al.*, "Anomaly detection using temporal data mining in a smart home environment," *Methods of Information in Medicine*, vol. 47, no. 1, pp. 70–75, 2008.

[21] T. W. Liao, "Clustering of time series data – a survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.

[22] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2, pp. 191–203, 1984.

[23] BC Hydro. Smart metering program. Accessed: 2016-01-21. [Online]. Available: https://www.bchydro.com/index.html

[24] E. Kreyszig, *Advanced Engineering Mathematics Tenth Edition*. Wiley, 2011, p. 1014.

[25] F. Iglesias and W. Kastner, "Analysis of similarity measures in times series clustering for the discovery of building energy patterns," *Energies*, vol. 6, no. 2, pp. 579–597, 2013.

[26] P. C. Mahalanobis, "On the generalized distance in statistics," in *Proceedings of the National Institute of Sciences*, vol. 2, 1936, pp. 49–55.

[27] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Association for the Advancement of Artificial Intelligence (AAAI) Workshop on Knowledge Discovery in Databases on*, 1994, pp. 359–370.