International Journal of the Physical Sciences Vol. 6(31), pp. 7233 - 7241, 30 November, 2011
Available online at http://www.academicjournals.org/IJPS
DOI: 10.5897/IJPS11.1410
ISSN 1992 - 1950 ©2011 Academic Journals

Full Length Research Paper

Temporal modeling and its application for anomaly detection in smart homes

M. R. Alam[1], M. B. I. Reaz[2]* and H. Husain[2]

[1]Institute of Microengineering and Nanoelectronics, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia.
[2]Department of Electrical, Electronic and Systems Engineering, Universiti Kebangsaan Malaysia
43600 Bangi, Selangor, Malaysia.

Accepted 30 September, 2011

Classification and modeling of activity duration provide significant characteristics to estimate the psychological behaviour of a smart home resident. This article validates the fact, which was only an assumption previously, that smart home event duration can be modeled in Gaussian distribution. It proposes a temporal prediction algorithm based on Gaussian distribution to predict the duration of an event interval, which approximates the ending time of the smart home user's activities. It incrementally estimates the ending time of an event that follows the central limit theory of statistical probability. The results and analysis imply that temporal duration follows Gaussian distribution, which expresses almost the same property of Gaussian equation. The algorithm is verified with significant amount of MavHome and MIT PlaceLab smart home sensory data, which exhibit 88.3 and 90.3% prediction accuracies respectively. Finally, the proposed temporal algorithm is utilized for temporal anomaly detection, which has detected 54 and 46 abnormal behaviour when tested with MavLab and MIT PlaceLab data respectively.

Key words: Temporal duration, smart homes, Gaussian distribution, prediction algorithm, temporal prediction, anomaly detection.

INTRODUCTION

Smart home projects have been conducted for the last several decades and they convey different ideas, functions and utilities. Smart home is an application of ubiquitous computing where the home environment is monitored by ambient intelligence to provide the user with context-aware services and facilitate remote home control. It is expanding into different branches of specialization that focus on the interest of the researchers, and user requirements and expectations.

Smart home research involves the understanding of human psychology to predict inhabitant behaviour. The success of smart home research mostly depends on how efficiently human behaviour can be represented into existing computing elements. User activity is a collection of smaller tasks that occur repeatedly following specific temporal pattern. Predicting the time duration of events is an important parameter to identify temporal characteristics of human activity. Unfortunately, the event durations are not persistent. Duration varies according to user habit, willingness, and may even change between weekdays and weekends. The task duration is also influenced by environmental parameters like temperature, humidity, rain, snowfall and so on. The activity duration may also differ during daytime and at night. There should be some method to identify this time-varying pattern and utilize it for temporal prediction.

Temporal duration is a stochastic random variable that eventually follows a statistical distribution. This study investigates the potential of constructing a temporal

*Corresponding author. E-mail: mamun.reaz@gmail.com. Tel: +603-89216311. Fax: +603-89216146.

Table 1. Allen's thirteen temporal relations.

Relation	Pictorial example	Inverse relation	Pictorial example
X before Y	XXX YYY	X after Y	YYY XXX
X meets Y	XXXYYY	Y meets X	YYYXXX

X overlaps Y	XXX
	  YYY
		Y overlaps X	YYY
			  XXX

X during Y	XXX YYYYYYY	Y during X	YYY XXXXXX

X starts Y	XXX
	  YYYYY
		Y starts X	YYY
			  XXXXX

X finishes Y	XXX YYYYYY	Y finishes X	YYY XXXXX
X equals Y	XXX YYY		

prediction algorithm based on the central limit theory. It formulates the temporal characteristics of a smart home inhabitant based on hypothesis test and validation. Initially, we assumed the hypothesis that temporal duration follows Gaussian distribution (Allen, 1983; Jakkula and Cook, 2007; Ozgun and Orhan, 2011; Rashidi and Cook, 2009). Based on this hypothesis, an incremental algorithm was proposed, which recursively constructed a temporal database with statistical mean and standard deviation. We have conducted experiment with practical dataset to validate the model. Experimental results confirm the validation of the hypothesis, that is, smart home temporal duration follows Gaussian distribution.

RELATED WORKS

Although temporal prediction is a potential problem for smart home implementation, there are only a few research outcomes for this problem. A very early study was done by Allen in 1983, which mainly discussed temporal logic of event intervals (Allen, 1983). In his study, Allen argued that time interval is more informative than the point of time. The temporal relations between two events can be classified into thirteen distinct conditions. If *X* and *Y* are two events, then temporal relations can be classified into thirteen different ways considering the inverse of those relations as shown in Table 1.

Based on this logic, Allen presented a constraint propagation algorithm, which incrementally updates its temporal network using predicative logic. In the study, Allen proposed that the temporal logic can be utilized for duration reasoning. However, the method provides only logical duration relationship, which is not a numerical value, that is, if *X* and *Y* are two events, it can only estimate whether the duration of *X* is larger, smaller or equal to *Y*. It does not provide the numerical time duration of *X* or *Y* (in day, hour, minute or second).

Gopalratnam and Cook (2007) assume that the time interval between smart home events approximates a Gaussian distribution. Their Active LeZi algorithm incrementally builds a Gaussian that represents the observed Gaussian distribution of the relative time of smart home events. The mean and standard deviation of the Gaussian is constructed incrementally by recursively defining the values. The resulting algorithm exhibits 70% probability to get the next event within the mean ± standard deviation of the predicted time. The algorithm is based on the hypothesis that the intervals follow Gaussian distribution. But they did not provide any statistical evidence of the assumption. Moreover, it is tested on synthetic data which does not reflect real life scenarios.

Jakkula and Cook (2007) tried to combine the above two algorithms for temporal prediction. They simplified Allen's temporal logic, which only determines the most probable states of thirteen temporal relations. For this purpose, an algorithm is proposed to determine the most frequent relationship between the events. For interval reasoning, the researchers modified Active LeZi to predict between $\mu \pm 2\sigma$ ranges (μ and σ represent mean and standard deviation respectively). Its functionality is similar to Allen's temporal logic, which only estimates the relation between the events. It fails to provide methodologies to predict the task duration for a smart home event.

Mori et al. (2008) used Gaussian mixture model (GMM) to detect the behavioural anomaly of smart home residents. Like other previous researchers, the authors used the concept of Gaussian distribution without validating this hypothesis.

Most of the previous algorithms related to resolving the temporal relationship, which is not the main concern of our problem (Allen, 1983; Gopalratnam and Cook, 2007).

However, these methods provide several guidelines

```
Initialize temporal_database: = null
Initialize task_id: = 1

Loop
  Wait for the sensor data
  If data found
    Grab the sensor_id and status
    Check the sensor_id in the temporal_database
                If sensor_id does not exists
      Insert task_id, sensor_id, current_time as status_time to the
      Temporal_database
      Set task_id: = task_id+1
                Else
      If status = ON
        (Update corresponding event in temporal_database and
        predict the ending time of the event)
        Set status: = ON
        Set status_time: = current_ time
        Predict the ending time: = mean ± 2 * standard deviation
      Else
        (Update corresponding event in temporal_database)
        Set time duration: = current time - status time
        Set standard deviation: = | mean – time duration |
                Set mean: = (mean+ time duration) / 2
Forever
```

Figure 1. Pseudocode of the proposed prediction algorithm.

to develop an algorithm for temporal duration. Several algorithms are based on the hypothesis that temporal interval follows Gaussian distribution, but they do not validate their assumption (Gopalratnam and Cook, 2007; Jakkula and Cook, 2007; Youngblood and Cook, 2007). The pro-posed algorithm validates the hypothesis and presents an efficient technique by intensive analysis of practical smart home data.

THE TEMPORAL PREDICTION ALGORITHM

Human activity can be modeled via utilizing the information generated by sensors attached to home appliances. The difference between starting time and ending time of an electrical appliance indicates the temporal duration of device usage. Contact switches connected to furniture doors are used to measure the duration of open and close status. A user can be tracked with pressure sensors under the floor, which indicate the duration of presence at that location. Therefore, most of the smart home event durations can be estimated from the starting and finishing points. This study proposes a method to find out a relationship between the time durations of smart home user activities.

Suppose, for any appliance, t_s indicates the starting point and t_e indicates the ending point of device usage. Let t_{s_1} be the 1st starting point of the appliance, which ends at t_{e_1}; t_{s_2} be the 2nd staring point of that device, which ends at t_{e_2} and so on. And, t_{s_n} is the *n* th starting point of the appliance, which ends at t_{e_n}. There are *m* appliances that can be represented by $x_1, x_2, x_3, \ldots, x_m$. We have to predict the ending time, t_e of any appliance usage given that the starting time is t_s. It also implies that the system is aware of all previous starting times $t_{s_1}, t_{s_2}, \ldots, t_{s_n}$ and ending times $t_{e_1}, t_{e_2}, \ldots, t_{e_n}$.

Data collection

For our work, we used practical smart home data from MavLab (webpage-ailab, 2011) and PlaceLab (webpage-placelab, 2011). MavLab is the testbed of MavHome at University of Texas in Arlington. The data sample consists of the activities of six inhabitants at MavHome in April 2003. MavHome dataset has 51 different appliances with time and status information. There are a total of 689 sequential sensor events with 326 temporal durations.

We used wire switch data from PlaceLab Intensive Activity 1 (PLIA1) dataset (webpage-placelab, 2011). PLIA1 is a dataset from MIT PlaceLab, which was initiated by MIT House_n research group (Intille et al., 2006). The wire switches detect on/off and open/close events such as doors being opened/closed and knobs being turned using switches built into the infrastructure. There are total 953 sequential events from 30 wire switches, which create total 469 durations of the events.

METHODOLOGY

The problem is to predict the ending time of an event, given the starting time. The ending time is directly related to the duration of the task. Suppose,

$t_{x_{1_1}}, t_{x_{1_2}}, t_{x_{1_3}}, \ldots\ldots, t_{x_{1_n}}$ are the durations of x_1 event.

$t_{x_{2_1}}, t_{x_{2_2}}, t_{x_{2_3}}, \ldots\ldots\, t_{x_{2_n}}$ are the durations of x_2 event.

……………………………

$t_{x_{m_1}}, t_{x_{m_2}}, t_{x_{m_3}} \ldots\ldots\, t_{x_{m_n}}$ are the durations of x_m event.

We have to develop a model considering all these temporal durations according to the corresponding events to predict the finishing time of a smart home event.

An incremental learning algorithm is proposed to represent the dataset in the temporal_database, where temporal_database is the name of the database location. Instead of storing every value of time duration, it processes the mean and standard deviation using only the

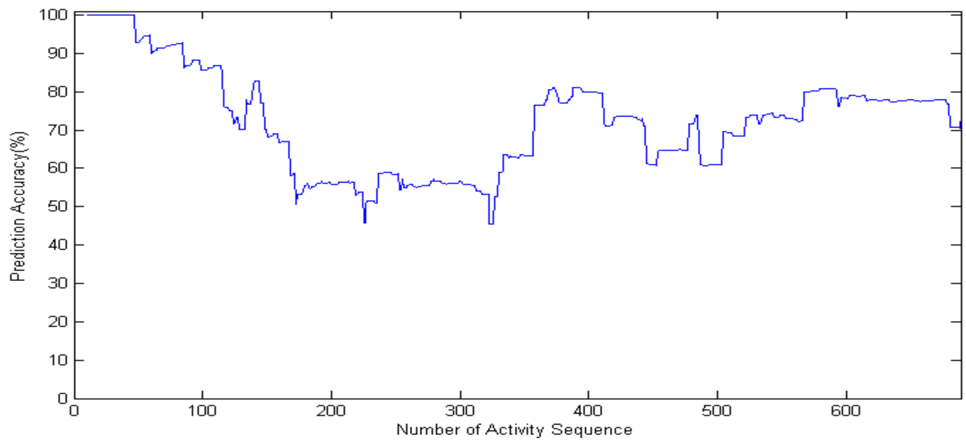

Figure 2. Prediction accuracy when $\mu \pm \sigma$ is used to predict the ending time of MavLab events.

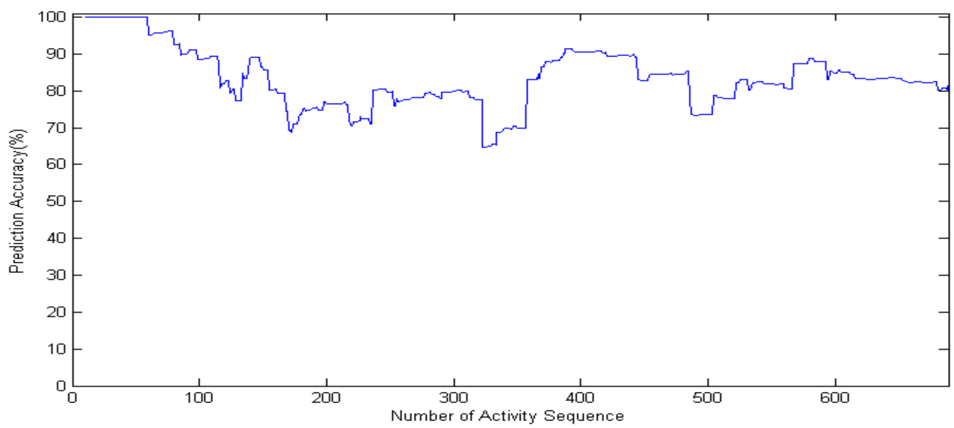

Figure 3. Prediction accuracy when $\mu \pm 2\sigma$ s used to predict the ending time of MavLab events.

previous value, which is the incremental output of all the previous history.

Figure 1 is the pseudocode of the proposed prediction algorithm. Initially, the temporal_database is empty. The program waits for the sensor data. If any sensory information arrives, it checks the sensor_id in the temporal_database, where sensor_id is the identification of sensory information. If sensor_id is not found, it assigns a unique task_id for the sensor, stores the sensor_id, status and current_time as status_time, where task_id is the identification of the sensor, status is the present status of the sensor, current_time is the time the sensor information arrive and status_time is the combination of both status and current_time. At this point, mean and standard deviation is empty. If sensor_id exists, it checks whether the arriving sensor status is ON or OFF. If arriving status is ON and the corresponding temporal_database event status is OFF, it just updates the corresponding database event status to ON and predicts the ending time by calculating $\mu \pm 2\sigma$. If arriving status of the sensor is OFF and the corresponding temporal_database event status is ON, it updates the status to OFF; calculates the event time_duration from last status_ time and current_time and sets status_time to current_time. Then, it updates the deviation from last mean and calculated time_duration, where time_duration is the duration of task execution. Finally, it computes the mean from last mean and calculated time_duration.

PERFORMANCE ANALYSIS AND RESULTS

The algorithm is tested incrementally to evaluate its performance. For this purpose, initially it was trained with the first event, and tested using only that event to estimate the prediction accuracy. Then, it was trained with the first two sequential events and tested using those two events for prediction accuracy calculation. Similarly, the temporal_database was trained with all the events and tested with those events to check whether it can predict the time durations accurately.

Figures 2 to 4 shows the accuracy curves using MavLab data. Figure 2 illustrates the prediction accuracy when the ending times of the events were predicted between $\mu \pm \sigma$. Initially, when the training history is small,

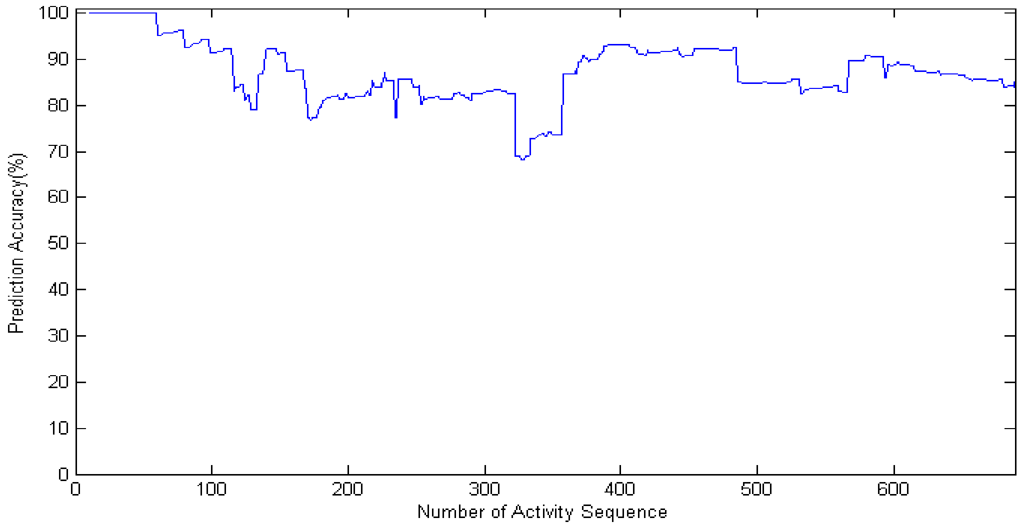

Figure 4. Prediction accuracy when $\mu \pm 3\sigma$ is used to predict the ending time of MavLab events.

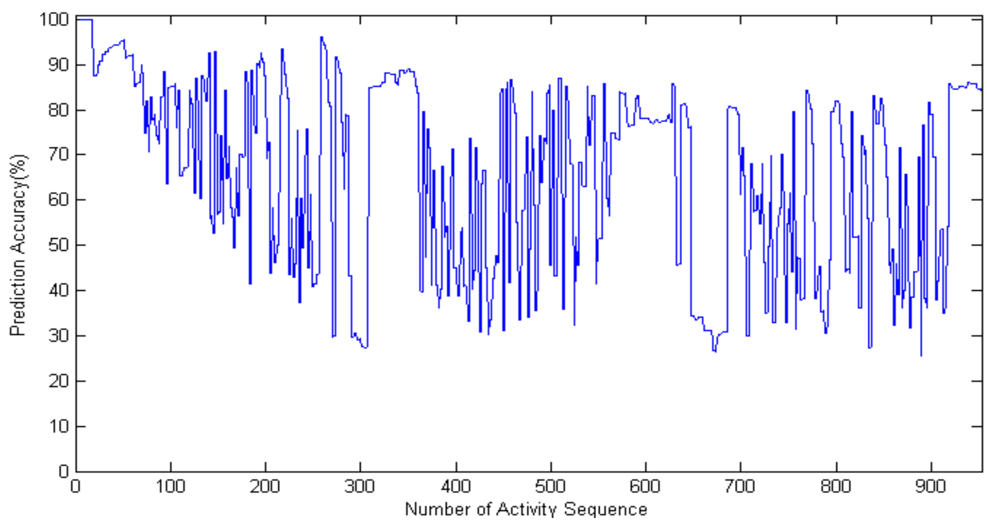

Figure 5. Prediction accuracy when $\mu \pm \sigma$ is used to predict the ending time of PlaceLab events.

it shows immature prediction accuracies. The increment of training sequence tries to converge the temporal database. The prediction accuracy becomes almost stable in the range of 60 to 80%.

The prediction accuracy increases when the predictor estimates the ending point between mean $\mu \pm 2\sigma$. In this case, the algorithm converges after about 350 sequences and shows persistent prediction accuracy between 80-90% as shown in Figure 3. The performance improves if $\mu \pm 3\sigma$ is utilized to verify the algorithm. In this case, the prediction accuracy lies between 90 to 97% as shown in Figure 4. Figures 5 to 7 exhibits prediction accuracies when MIT PlaceLab data is used to verify the algorithm. When the algorithm is tested between $\mu \pm \sigma$ range, the prediction accuracies lie between 50-80% (Figure 5). Figure 6 shows that prediction accuracies increase when $\mu \pm 2\sigma$ is utilized to calculate the prediction accuracy. The average prediction accuracy is 79.4% and most of the time, the curve shows prediction accuracies between 60 to 90%.

The prediction accuracy becomes more stable (Figure 7) when $\mu \pm 3\sigma$ is used as the time duration range. In this case, the curve shows stable prediction accuracies between 80 to 90% (except a few exceptional cases).

The figures (Figures 2 to 7) illustrate some significant properties of the algorithm. Initially, the training sequence

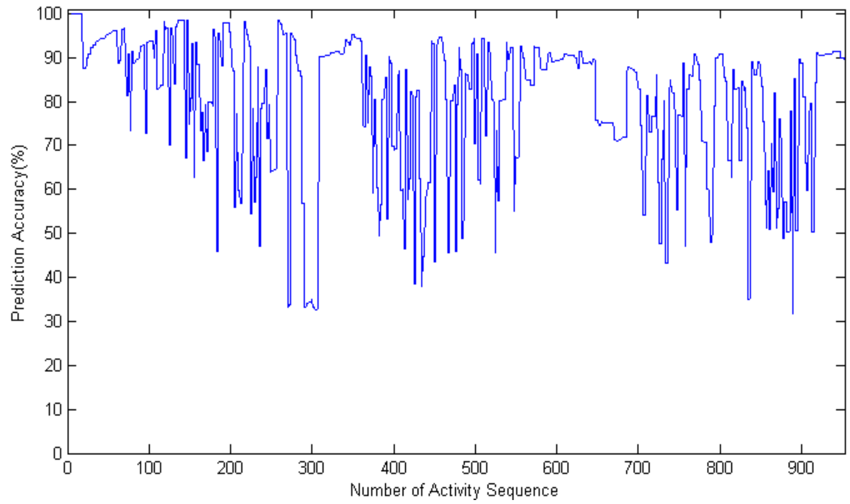

Figure 6. Prediction accuracy when $\mu \pm 2\sigma$ is used to predict the ending time of PlaceLab events.

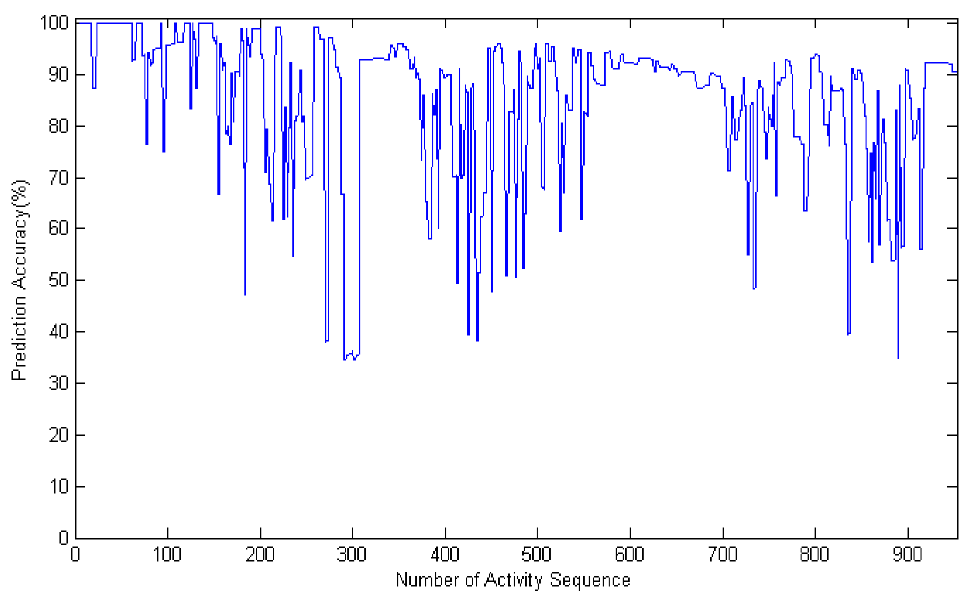

Figure 7. Prediction accuracy when $\mu \pm 3\sigma$ is used to predict the ending time of PlaceLab event.

sequence is not adequate for prediction. At those stages, the algorithm shows immature prediction accuracies. With the passage of time, more events arrive and the algorithm starts to converge. After a certain amount of sensory information, it shows stable prediction accuracies.

The curves of Figures 2 to 7 shows that the prediction accuracy is reasonable compared with previous researches where the predictor used $\mu \pm 2\sigma$ range to determine the ending time. Therefore, this property is utilized in the proposed algorithm to predict the durations.

THE TEMPORAL MODEL

The algorithm is tested from different angles with respect to the standard deviation multiplier to properly identify the temporal pattern of smart home event durations. For this purpose, the temporal database is first trained with all smart home sensor data sequences. Then it is tested for the incremental multipliers of standard deviation which are added to the mean. Figure 8 shows how prediction accuracy increases according to the increment of the multiplier.

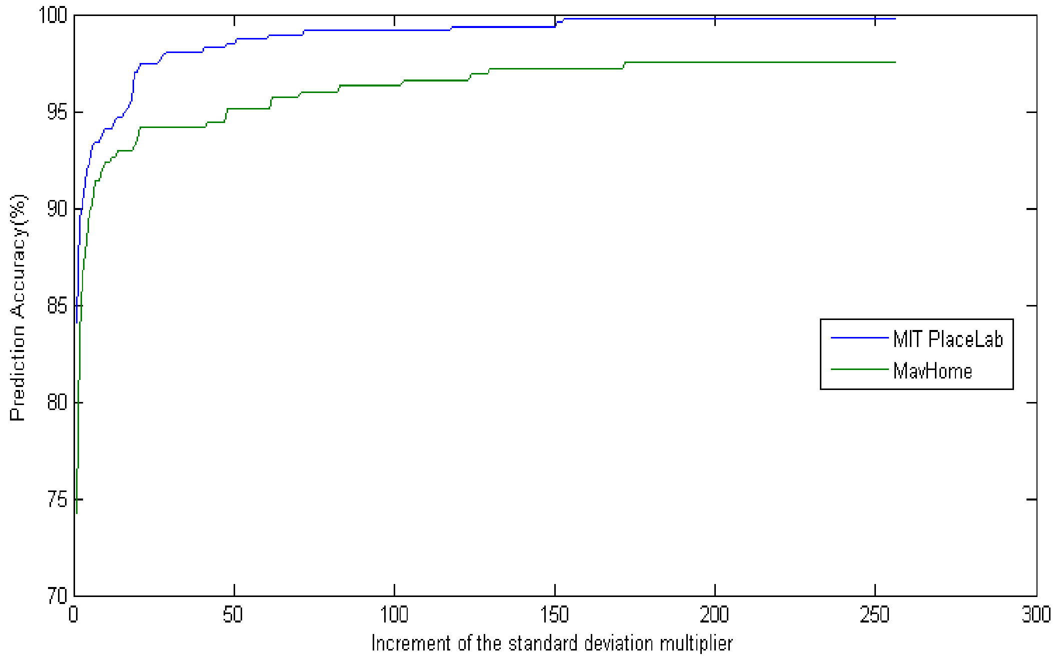

Figure 8. Prediction accuracy increases according to increment of standard deviation multiplier.

The lower accuracy curve shows the prediction accuracy using MavHome data. The curve shows that when $\mu \pm \sigma$ is used for end time prediction; it exhibits 74.2% prediction accuracy. If $\mu \pm 2\sigma$ is applied, the accuracy increases to 83.4%, which is about 9.2% better than the previous result. It shows about 3.1% improvement when $\mu \pm 3\sigma$ is utilized for prediction. If $\mu \pm 4\sigma$ is applied, the accuracy becomes 88.3%, which is 1.8% higher than the previous one. The accuracy curve becomes almost parallel to the *x*-axis with further increments of the multiplier.

The higher accuracy curve shows that for MIT PlaceLab data, the prediction accuracies are all time higher than MavHome data (Figure 8). It shows 90.3% prediction accuracy for $\mu \pm \sigma$, $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$ duration ranges respectively. The curves of Figure 8 resemble the pattern of Gaussian distribution. A Gaussian distribution is expressed by the following probability density function (Montgomery et al., 2004),

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{for, } -\infty < x < \infty \qquad (1)$$

From (1), we can compute that for any Gaussian random variable,

$P(\mu - \sigma < X < \mu + \sigma) = 0.683$
$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.955$
$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.997$

Figure 9 illustrates the temporal model of smart home event duration. Smart home event temporal duration follows a Gaussian distribution that can be represented by a bell-shaped curve and shows a 74.2% probability of predicting the ending time between $\mu - \sigma$ and $\mu + \sigma$ (for MavHome data). 83.4 and 86.5% accuracies are achievable if the durations are [$\mu - 2\sigma$, $\mu + 2\sigma$] and [$\mu - 3\sigma$, $\mu + 3\sigma$] respectively (for MavHome data). Figure 10 shows a similar bell-shaped curve when the algorithm is tested using MIT PlaceLab data. Both the curves prove that smart home event duration follows Gaussian distribution.

TEMPORAL ANOMALY DETECTION

Inhabitant activity prediction algorithms make significant contribution to anomaly detection (Youngblood and Cook, 2007; Rashidi and Cook, 2009), activity identification (Chen et al., 2009; Barnes et al., 1998), assistive services (Barger et al., 2005; Brdiczka et al., 2009; Assim et al., 2006; Adlam et al., 2004) etc. The proposed temporal model and temporal prediction algorithm can effectively detect and identify abnormal behaviour of the residents. The model can easily be implemented into low computing power hardware because of simplicity and effectiveness.

Suppose the inhabitant takes average 30 min for taking bath. Sometimes he may take more than 30 min, but never exceeds 35 min. If the duration exceeds 35 min, it means that there is a possibility of anomaly. This algorithm

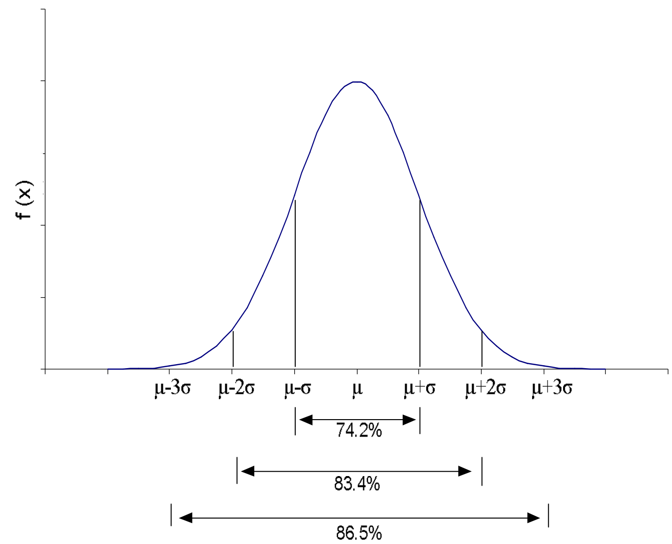

Figure 9. The bell curve for MavHome events' temporal duration which shows that it follows a Gaussian distribution.

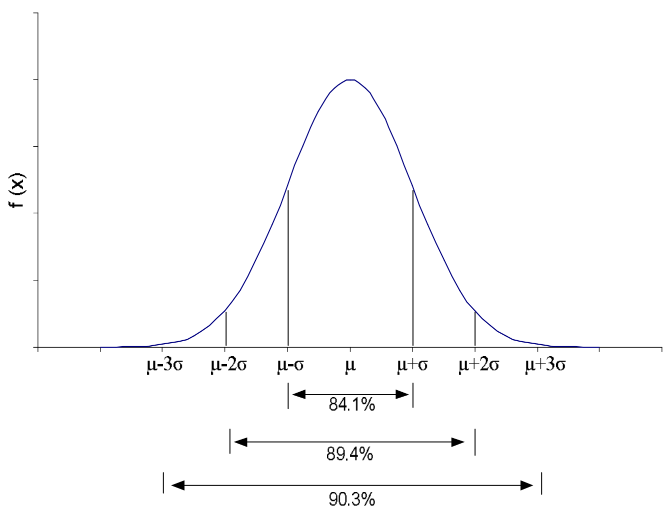

Figure 10. The bell curve for MIT PlaceLab events' temporal duration which shows that it follows a Gaussian distribution.

can be utilized to identify these types of abnormal behaviours. Consider another scenario where the inhabitant forgets to close the main door. This inactivity can be easily detected by average door open time duration and its deviation.

The proposed algorithm can detect the temporal anomaly of a smart home resident. It generates a temporal database of smart home event durations, which consist of average event durations and their deviations. For MIT PlaceLab data, the algorithm generates 30 average durations and 30 standard deviations for every event. When, the algorithm was tested using all the previously seen temporal durations, it detected total 46 temporal anomalies.

When the temporal database was trained with MavHome data, it generated 51 average durations and 51 standard deviations for all the event durations. The algorithms got total 326 durations between the events and it detected 54 abnormal behaviours that exceeded normal time durations.

In case of any abnormal activities, the smart home can generate an alarm or contact the remote health care center for immediate support.

CONCLUSION

This article presents an effective algorithm to predict temporal durations and ending times of smart home events. Although it is a potential problem for smart home event prediction, such an algorithm had not been formulated previously. The paper proposed an algorithm that shows 83.4% prediction accuracy when tested with MavHome smart home data. For MIT PlaceLab data, it exhibits 89.4% prediction accuracy. Several important properties related to the central tendency of the dataset are evaluated to illustrate the actual pattern of temporal durations. It validates the fact that smart home event duration can be modeled in Gaussian distribution, which was only an assumption previously. The proposed algorithm has a major application for temporal anomaly detection. It has detected 54 and 46 abnormal activities of the residents when tested with MavLab and MIT PlaceLab dataset respectively. The proposed temporal duration prediction algorithm and the temporal model present an effective way to represent temporal characteristics of the inhabitants.

ACKNOWLEGDEMENTS

This work was supported by the research grant Geran Galakan Penyelidik Muda (UKM-GGPM-TK-046-2010) and Geran University Penyelidikan (UKM-GUP-2011-043) from University Kebangsaan Malaysia and Long Term Research Grant Scheme (LRGS/TD/2011/UKM/ICT/03) from Ministry of Higher Education (MOHE), Malaysia.

REFERENCES

Adlam T, Faulker R, Orpwood R, Jones K, Macijauskiene J, Budraitiene A (2004). The installation and support of internationally distributed equipment for people with dementia. IEEE Trans. Inform. Technol. Biomed., 8(3): 253–257.

Allen JF (1983). Maintaining Knowledge about Temporal Intervals. Comm. of the ACM. 26: 832-843.

Assim A, Reaz MBI, Ibrahimy MI, Ismail AF, Choong F, Mohd-Yasin F (2006). An AI Based Self-Moderated Smart-Home. Informacije MIDEM – J. Microelectronics, Electronic Components Mate., 36(2): 91-94.

Barnes NM, Edwards NH, Rose DAD, Garner P (1998). Lifestyle monitoring technology for supported independence. Comput. Contr. Eng. J., 9(4): 169–174.

Barger TS, Brown DE, Alwan M (2005). Health-status monitoring through analysis of behavioral patterns. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 35(1): 22-27.

Brdiczka O, Crowley JL, Reignier P (2009). Learning Situation Models in a Smart Home. IEEE Trans. Syst. Man Cybern. B. Cybern., 39(1): 56-63.

Chen CY, Tsoul YP, Liao SC, Lin CT (2009). Implementing the design of smart home and achieving energy conservation. Proceedings of the 7th IEEE Int. Conf. on Industrial Informatics (INDIN) held at Cardiff, Wales. pp. 273-276.

Gopalratnam K, Cook DJ (2007). Online Sequential Prediction via Incremental Parsing: The Active LeZi Algorithm. IEEE Intell. Syst., 22(1): 52-58.

Intille SS, Larson K, Tapia EM, Beaudin J, Kaushik P, Nawyn J, Rockinson R (2006). Using a live-in laboratory for ubiquitous computing research. Proceedings of PERVASIVE. pp. 349-365.

Jakkula VR, Cook DJ (2007). Using Temporal Relations in Smart Environment Data for Activity Prediction. Proc. of the 24th Int. Conf. on Machine Learning held at Corvallis, USA. p. 4.

Montgomery DC, Runger GC, Hubele NF (2004). Engineering Statistics, John Wiley & Sons, Inc.

Mori T, Urushibata R, Shimosaka M, Noguchi H, Sato T (2008). Anomaly Detection Algorithm Based on Life Pattern Extraction from Accumulated Pyroelectric Sensor Data. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems held at Nice, France. Pp. 2545-2552.

Ozgun A, Orhan A (2011). An Ontology-based Visualization for Mobile Geoinformation Services. Int. J. Phys. Sci., 6(4): 993-1000.

Rashidi P, Cook DJ (2009). Keeping the Resident in the Loop: Adapting the Smart Home to the User. IEEE Trans. Syst. Man Cybern., Part A: Systems and Humans. 39(5): 949-959.

Webpage-ailab, http://ailab.eecs.wsu.edu/casas/datasets.html.

Webpage-Placelab, http://architecture.mit.edu/house_n/data/PlaceLab/PLIA1.htm

Youngblood GM, Cook DJ (2007). Data Mining for Hierarchical Model Creation. IEEE Trans. Syst. Man Cybern. C Appl. Rev., 37(4): 561-572.