# Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 11 (16): 3058-3061, 2011
ISSN 1812-5654 / DOI: 10.3923/jas.2011.3058.3061


# Statistical Modeling of the Resident's Activity Interval in Smart Homes

M.R. Alam, M.B.I. Reaz and M.A.M. Ali
Department of Electrical, Electronic and Systems Engineering,
Universiti Kebangsaan Malaysia, 43600, UKM, Bangi, Selangor, Malaysia

**Abstract:** The activities of residents in smart homes possess temporal information which can be used to classify and model psychological behavior of the resident. In this study, a learning algorithm is proposed to predict the activity interval of smart home inhabitants. The algorithm is based on the hypothesis that residents' activity intervals follow a normal distribution. To predict the starting time of the following activity, it incrementally utilizes mean and standard deviation of previous history which are applied according to the central limit theory of statistical probability. The prediction algorithm exhibits 88.3 to 95.3% prediction accuracies for different ranges of mean and standard deviations when verified by practical smart home data. Further stochastic analyses prove that the time difference between the residents' activities follows normal distribution which was merely an assumption previously.

**Key words:** Smart home, activity interval, normal distribution, pervasive environment, ambient intelligence, temporal prediction

## INTRODUCTION

Smart home is a study of human psychology to create a pervasive environment employing ambient intelligence. The research mostly depends on how efficiently human behavior can be represented into existing computing elements. There are several researches on different aspects of smart home, especially temporal context reasoning and pervasive computing on smart home as an ubiquitous services (Liao and Tu, 2007; Hussain *et al*., 2008; Xiaohu and Guangxi, 2006). User activity is a collection of smaller tasks which occur repeatedly following specific temporal pattern. Classification and modeling of human activity provide fundamental background to develop an ubiquitous environment for smart homes. Human activities are periodic events which occur repeatedly following specific time intervals. Previous researchers try to extract the patterns of resident's activity using statistical prediction methods (Virone *et al*., 2008). Some researchers used Case Based Reasoning (CBR) to represent the problem into a goal based manner (Sormo *et al*., 2005), some has used artificial intelligence (Assim *et al*., 2006). This article investigates the potential of constructing a quantitative temporal relationship between the activities of smart homes' residents.

Although, temporal prediction is a potential problem for smart home implementation, there are only a few research outcomes for this problem. A very early work was done by Allen (1983) which mainly discussed the temporal logic of activity intervals. In this work, Allen argued that time interval is more informative than point of time. The temporal relation between two activities can be classified into thirteen distinct conditions. Based on these relations, a constraint propagation algorithm is presented which incrementally updates its temporal network using predicative logic. However, the method provides only logical interval relationship which is not a numerical value i.e., if X and Y are two activities, it can estimate which relationship is applicable for the activities. It does not provide the numerical time interval of X or Y (in day, h, min or sec).

Gopalratnam and Cook (2007) assumed that the time interval between the activities of smart homes user approximates the normal distribution. Their Active LeZi algorithm incrementally builds a Gaussian that represents the observed normal distribution of the relative time of user activities. Mean and standard deviation of the Gaussian distribution are constructed incrementally by recursively defining the values. The resulting algorithm exhibits 70% probability to get the next activity within the range of Mean±Standard deviation of

**Corresponding Author:** Muhammad Raisul Alam, Department of Electrical, Electronic and Systems Engineering, Universiti Kebangsaan Malaysia, 43600, UKM, Bangi, Selangor, Malaysia
Tel: +603-89216311 Fax: +603-89216146

the predicted time. This algorithm is based on the hypothesis that the intervals follow normal 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) combined above two algorithms for temporal prediction. They simplified Allen's temporal logic which determines the most probable states from thirteen temporal relations. For the purpose, an algorithm is proposed to estimate the most frequent relationship between the activities. The researchers modified Active LeZi to predict between $\mu \pm 2\sigma$ ranges ($\mu$ and $\sigma$ represent mean and standard deviation, respectively). Its functionality is similar to Allen's temporal logic which only approximates the relationship between the activities. It fails to provide methodologies to predict activity intervals of smart homes' inhabitants.

This study analyzes the temporal characteristics between smart home residents' activities. The analysis includes statistical estimation of time difference between smart home user activities. It investigates the probability of finding the time of next user activity between time intervals. Based on the analysis, a temporal model of human activity is proposed which formulates the activity intervals of smart homes' residents.

## MATERIALS AND METHODS

The difference between ending time of an activity and starting time of another activity indicates the temporal interval between those two activities. This article proposed a methodology to find out the time interval between the activities of smart homes' residents.

Suppose, x and y are two user activities.

Let, $t_{e_{x1}}$ is the 1st ending point of the x activity which is followed by, $t_{s_{y1}}$ the next starting point of the y activity.
$t_{e_{x2}}$ is the 2nd ending point of the x activity which is followed by, $t_{s_{y2}}$ the next starting point of the y activity.
$t_{e_{xn}}$ is the n th ending point of the x activity which is followed by, $t_{s_{y2}}$ the next starting point of the y activity.

The problem is to predict the starting time, $t_{s_y}$ of the y activity given the ending time, $t_{e_x}$ of the x activity. It also implies that, the system is aware of all previous ending times $t_{e_{x1}}$, $t_{e_{x2}}$,.......,$t_{e_{xn}}$ of the x activity and starting times $t_{s_{y1}}$, $t_{s_{y2}}$,.......,$t_{s_{yn}}$ of the y activity.

There are m activities. So, total unique m (m-1) activity intervals exist between all user activities.

An incremental learning algorithm is proposed to represent the information into temporal_database. Instead of storing every value of time intervals, it processes the mean and standard deviation using only the previous value which is the incremental output of previous history.

Figure 1 is the pseudocode of the proposed learning algorithm. Here, the temporal_database variable is used to store all the time intervals between the smart home user activities. The window variable is used to temporarily store near past history of the ending time of the activities. Initially, the temporal_database is empty. The algorithm waits for the activity data. If any data arrives, it checks the status of the activity. If activity status is END_TIME, it stores the activity_id and time into the window. If activity status is START_TIME, it extracts all activity_id and time from the window, calculates the interval between activity time and current_time of current activity and updates the corresponding mean and standard deviation in temporal_database as shown in Fig. 1. The algorithm processes only successive activity interval information.

Figure 2 is the pseudocode of the proposed prediction algorithm which uses the temporal_database from the learning algorithm to predict the next activity occurrence time. It can predict a finite range of time of an activity given the current activity, current time and recursive mean, standard deviation between the current and next activity time.

```
initialize temporal_database: = null
initialize window: = null
loop
   wait for data
    if data found
       Grab the activity_id and status
       If status = END_Time
            add the activity_id time in window
        else
          loop until window is empty
          extract activity_id and time from window
          set time_duration: = current_time-time
          set standard_deviation: = |mean-time_duration
          |
          set mean: = (mean + time_duration)/2
          forever
forever
```

Fig. 1: Pseudocode of proposed learning algorithm

```
Given, temporal database
        current_event_id
        next_event_id
        Current_time
search for the corresponding mean and standard deviation in
the temporal_database

next_event_time = current_time + (mean±standard
_deviation)
```

Fig. 2: Pseudocode of proposed prediction algorithm

## STOCHASTIC ANALYSIS

MavLab dataset from MavHome is used to analyze the temporal characteristics of smart home residents' activities (Youngblood and Cook, 2007). MavHome was a smart home testbed at University of Texas in Arlington (mavlab). The dataset consists activity time of MavHome inhabitants in April 2003. It had more than 50 unique user activities which occurred repeatedly about 700 times.

Figure 3 illustrates the prediction accuracy when the starting times of the activities were predicted between several statistical ranges. When the probability of finding the next activity is assumed between $\mu$-$\sigma$ and $\mu$+$\sigma$, prediction accuracy remains between 80-90%. Prediction accuracy increases when the predictor estimates the starting point between $\mu\pm2\sigma$. In this case, it shows persistent prediction accuracy between 90-93%. The performance improves if $\mu\pm3\sigma$ is utilized to verify it. In this case, prediction accuracy remains between 93-95%.

Figure 4 shows how prediction accuracy increases according to the increment of standard deviation multiplier when tested with a fully converged temporal database. The curve shows that, when $\mu\pm\sigma$ is used for starting time prediction, it exhibits 88.3% prediction accuracy. If $\mu\pm2\sigma$ is applied, the accuracy increases to 90.9% which is about 2.6% better than the previous result. It shows about 2.6% improvement when $\mu\pm3\sigma$ is utilized for prediction.

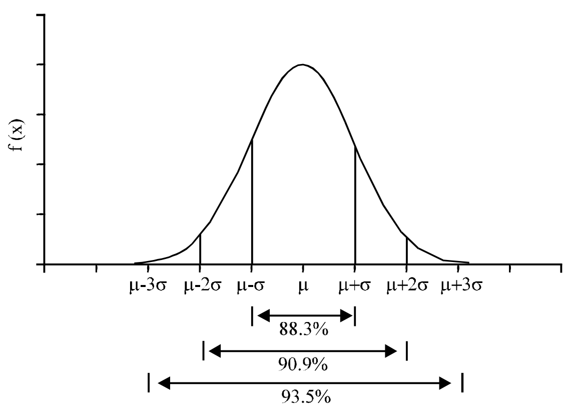


Fig. 3: Relationships between prediction accuracy and $\mu$, $\sigma$

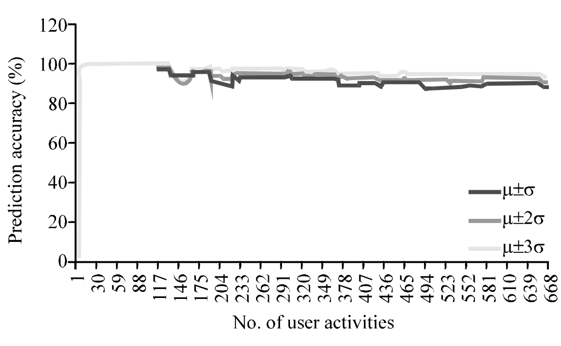


Fig. 4: The bell shaped curve showing that smart home temporal intervals follow normal distribution

The pattern of Fig. 4 resembles the characteristics of normal distribution. A normal distribution is expressed by the following probability density function (Montgomery *et al*., 2010).

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

From Eq. 1, we can estimate that for any normal random variable:

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

These prediction accuracies found from Eq. 1 are almost similar to the proposed temporal prediction algorithm.

## CONCLUSIONS

This study presents statistical analysis of smart home user activities. Several important properties related to the central tendency of the dataset are evaluated to illustrate the actual pattern of temporal intervals. It validates the fact that smart homes residents' activity intervals can be modeled in normal distribution which was only an assumption previously. The solution provides an effective way to represent temporal characteristics of the inhabitants which has a major application in temporal prediction and temporal anomaly detection.

## REFERENCES

Allen, J.F., 1983. Maintaining knowledge about temporal intervals. ACM Commun., 26: 832-843.

Assim, A., M.B.I. Reaz, M.I. Ibrahimy, A.F. Ismail, F. Choong and F. Mohd-Yasin, 2006. An AI based self-moderated smart-home. Informacije MIDEM: J. Microelectronics Electron. Components Mater., 36: 91-94.

Gopalratnam, K. and D.J. Cook, 2007. Online sequential prediction via incremental parsing: The active LeZi algorithm. IEEE Intell. Syst., 22: 52-58.

Hussain, C.S., C.S. Ahmed, A.H. Akbar, A.K. Bashir, K.H. Kim and W.S. Yoon, 2008. Ubiquitous service discovery in pervasive computing environment. Inform. Technol. J., 7: 533-536.

Jakkula, V.R. and D.J. Cook, 2007. Using temporal relations in smart environment data for activity prediction. Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA., June 20-24, Oregon State University, pp: 1-4.

Liao, H.C. and C.C. Tu, 2007. A RDF and owl-based temporal context reasoning model for smart home. Inform. Technol. J., 6: 1130-1138.

Montgomery, D.C., G.C. Runger and N.F. Hubele, 2010. Engineering Statistics. 5th Edn., John Wiley and Sons Inc., USA., ISBN-13: 978-0-470-63147-8.

Sormo, F., J. Cassens and A. Aamodt, 2005. Explanation in case-based reasoning-perspectives and goals. Artifi. Intell. Rev., 24: 109-143.

Virone, G., M. Alwan, S. Dalal, S.W. Kell, B. Turner, J.A. Stankovic and R. Felder, 2008. Behavioural patterns of older adults in assisted living. IEEE Trans. Inform. Tech. Biomed., 12: 387-398.

Xiaohu, G. and Z. Guangxi, 2006. Empowering ubiquitous services in next-generation smart homes. Inform. Technol. J., 5: 64-69.

Youngblood, G.M. and D.J. Cook, 2007. Data mining for hierarchical model creation. IEEE Trans. Syst. Man Cybernetics Part C: Appl. Rev., 37: 561-572.