Correlation Between Real and Simulated Data of the Activity of the Elderly Person Living Independently in a Health Smart Home

T. Hadidi, N. Noury, *Senior Member, IEEE*

Abstract— **Modeling the behavior of a system from observation is the first step in understanding a phenomenon. The model then opens the way to simulation in order to reproduce the behavior of the real system with the flexibility to modify, adjust and play with various scenarios. In this paper, we present a simulator of the activity of elderly persons in a Health Smart Home using the HMM "Hidden Markov Model". This simulator aims at modeling the effects of the loss of autonomy for the elderly living independently. We test several correlation methods to evaluate the similarity between real and simulated data. The experimental data was obtained in the HIS "Habitat Intelligent pour la Santé" within the framework of the French national "AILISA" project.**

I. INTRODUCTION

Actimetry is an interesting measure of the homeostasis of an individual in his environment. It proves of benefit in the field of remote health monitoring as it can be performed at home, within the so called Health Smart Home (HSH) [1-2]. The underlying idea is to build a model of the "normal" activity of the subject and to detect and interpret any significant deviations. Collecting information in the HSH however suffers from technical failures (imperfection of sensors, loss of data) and Modeling problems (insufficient learning period, lack of cases). Therefore, we decided to create a simulator in order to provide a regular HSH data using an ideal sensor, an adjustable duration and a configurable Profile.

In this paper, we present a simulator of activities in a Health Smart Home, based on the HMM "Hidden Markov Model" [3]. Obviously, we need methods to measure the similarity between the simulated and the real data. We assessed 4 methods, based on the concept of correlation, which we adapted to our context. The results of simulation and similarity methods are presented in the present paper. It is important to note that there are many published studies in the field of "behaviourism" most of them however are psychologically rather than mathematically oriented [4-6].

The experimental data was obtained in the HIS [7] "Habitat Intelligent pour la Santé" within the French national project AILISA [8]. Within this framework, our system was developed in order to produce data on activity which is similar to real life data, through a simulation process. While the displacements of a person could be considered as a stochastic process, the simulation is reproducible using the HMM. We started by modeling the experimental HIS data, then we performed the analysis with the AILISA software [9]. Eventually a database was created to be used instead of the real detected data which usually involves long periods of observation (months), with several contingencies (network logout, loss of information) and is limited to a reduced number of volunteers.

II. MATERIAL AND METHODS

The HIS (Habitat Intelligent pour la Santé) is a system based on the concept of "Health Smart Home". It was built in our laboratory several years ago [10] and further tested in hospital suites. Our HIS uses passive infrared (PIR) sensors to detect and record data on the daily activities of the inhabitant. The current data was obtained in an experimental HIS set up, installed in a hospital suite at Charles Foix hospital-Paris, equipped with 6 sensors. The Infrared Presence Sensors were placed on the walls (height 1.80m) so as to detect the presence of the subject in the entrance door, bed, armchair, sink, shower and toilet (Fig 1).

Manuscript received 14 March 2011.
Tareq Hadidi is with the TIMC-IMAG Lab, Grenoble, France; e-mail : Tareq.Hadidi@imag.fr
N.Noury is with the University of Lyon, lab INL-INSA, Lyon, France (phone: +33472438986; e-mail: Norbert.Noury@insa-lyon.fr) and lab TIMC-IMAG, Grenoble, France.

978-1-61284-696-5/11/$26.00 ©2011 IEEE

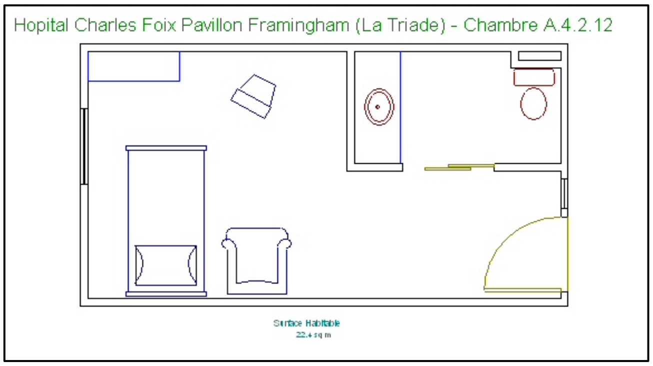

Figure 1: Architecture of the hospital suite at Charles Foix hospital-Paris.

A. Raw Data Processing

The real data collected by sensors is stored as time series. Every line consists of a list of events of detection [Date-Time-Localization] automatically written in a log file in XML format. The file is transmitted once a day by Email to a central data base [11]. Upon reception, the file is loaded in the Matlab™ environment, where data is stored in a preliminary matrix built on the events of detection. After transformation of data (time discretization and rectangularization) we obtain the signal $S_j^*(i)$ in final form. The daily graphical representation of the signal (1) is termed an « ambulatogram » (Figure 2). It enables immediate interpretation of the localization, duration and behaviour of daily activities and occupations [12].

$$S_j^*(i) = S^*(j,i) \qquad (1)$$

With i = identifier of the indicator and j=day

For example, with a simple visual reading of this ambulatogram, we can observe a period of inactivity from 10H P.M to 12H P.M.

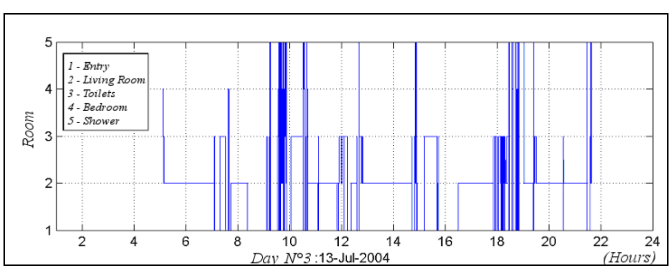

Figure 2: The "ambulatogram" Sj*(i) represents the detected data in each room for 24 Hours.

From this raw data, several parameters were elaborated, with the main goal to produce relevant indicators of non visible trends in the activity taking place in the suite, so as to inform the Health team, in charge of the subject, of potentially abnormal trends. We therefore decided to focus on the following indicators [13]: the mean activities occurring in front of each sensor (movements), the mean displacements from one sensor to another (displacements), the mean activity during the day (diurnal activity) and the mean activity at night time (nocturnal activity). Each indicator is a normalized value for the whole period (3),

$$Ind_i(j) = value_i(j) / Max\{value_i\} \qquad (3)$$

With i = identifier of the indicator and j=day.

These indicators were intensively studied in order to find the most pertinent parameters [13-15]. Then the parameters will be used to characterize the simulator.

B. Simulator

In this section, we present our work in three chronological steps. Firstly, we analyze data in order to understand the further steps of reasoning required. Secondly, we justify the selection of the data modeling methods, either for extraction or simulation processes, and how these methods are employed. Finally, we present the simulation, the simulator being implemented within the Matlab™ environment.

1) HIS data:

As seen above (paragraph A), the data collected from HIS is in the form of time series and presented as a matrix [Date, Time, Localization]. A monthly matrix can contain up to 60,000 lines of detected data. We first applied a non-linear median filter to remove detection artifacts (detections shorter than 1 second are considered as false events). Considering the nature of human behavior, and the random values of the time and the localization detections – the system variables - we decided to model the displacements using a stochastic system[1]. We so ended up with a stochastic process which produces random values controlled by the time. Therefore, we selected common methods from Artificial Intelligence to extract, then to simulate the HIS data. The extraction process starts with a classification phase using a clustering method coming from the Data Mining techniques. Then the data is processed under the predefined ***HMM*** function in Matlab™ to estimate the Markov transition and emission matrixes for each period of the day. Then we simulated the same data with another predefined function ***HMM*** which uses the former Markov matrixes as input parameters.

a. Data extraction

To prepare the process of classification, we first segmented the raw data into time periods. The clustering method was applied in order to create the daily matrix from the whole month. These daily matrixes then were cut into 4 time slots of equal length (six hours) for each day (figure 3).

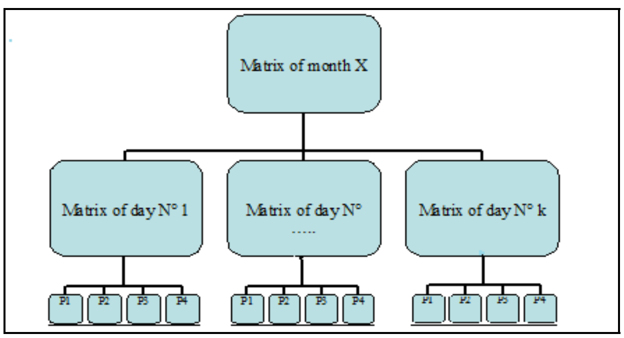

Figure 3: Pre-process of data extraction before classification.

The idea behind this reasoning is based on the circadian rhythm, with a slight modification of the number of the most pertinent periods to take in consideration [16]. We believe that sleep and meals are the most significant activities. Therefore we looked for the slots in the day where these activities usually happen, then we obtained the periods (P1, P2, P3 and P4).

b. Hidden Markov Model (**HMM**)

Through the data processing phase, we showed that the data matrix is characterized by two columns. One is the column of time in seconds then, if (i) is the time element, a one day finite time space is:

i ϵ [1,86400] : seconds.

The other column is the localization (n) and a one day finite states space is:

n ϵ [0,8] : 1 to 8 = sensor number.

From this set of data, we can establish all the statistics of mean and standard deviation of occupations and transitions. The latter are commonly represented using Markov Models, therefore the use of ***HMM*** seems a suitable probabilistic approach to model our system.

2) Markov Matrixes:

We used 2 Markov's matrixes when handling this kind of system. One is the ***transition*** matrix which links the real (experimental) data to the simulated data (Figure 4). It's an estimate of the transitions probabilities between the different states to the system according to the following formulas:

Stat probability

$$P(X_{n+1} = j / X_n = i) = P(X_1 = j / X_0 = i)$$

Transition Matrix

$$T(i,j) = P(X_t = j / X_{t-1} = i),\ 1 < t \leq N$$

With $T(i,j) \geq 0,\ \forall (i,j) \in N^2$ And

$$\sum_{j=1}^{N} T(i,j) = 1, \forall i \in N$$

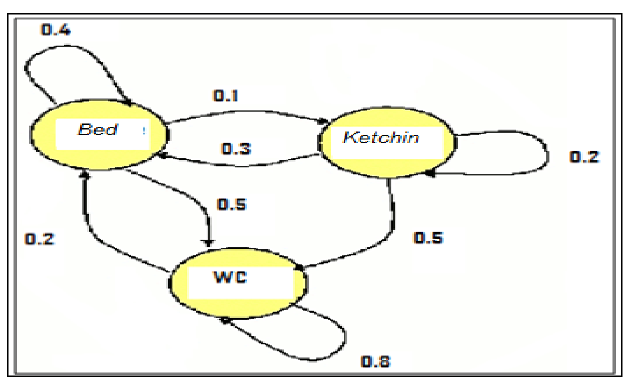

Figure 4: An example of the Markov model with 3 states.

[1] A stochastic system is a collection of random variables *{X(t), t∈T}* defined in a common probability space indexed by the index set T which describes the evolution of some system.

Consequently, we have a system where the future state only depends on the actual state (no memory effect).

The second one is the ***emission*** matrix which rules the temporal dimension through synchronizing the probability values in the matrix of transition. At this stage of data processing, we obtained the transition elements to build a profile (Figure 5) before entering the simulation phase.

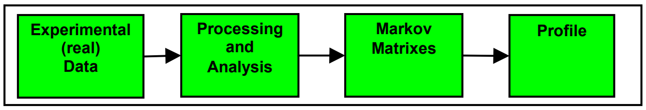

Figure 5: Diagram of the successive steps of the extraction phase.

3) Simulation

In this phase, we continue the simulation process but in reverse way (Figure 6). Thus, the production of profile parameters is done to be used in the reproduction (or regeneration) of similar data.

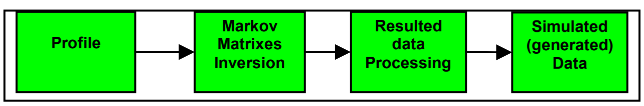

Figure 6: Diagram of successive steps of simulation phase.

These parameters are loaded into a function which generates correctly the two columns of data. One column is for the time with a simple format (in seconds), and the other is for the states (localizations). We note that the time column needs to be converted to standard time format, so a time converter is integrated within the simulator software.

4) Similarity

The next question concerns the similarity of the simulated data to the real data. This is a problem of distance between 2 time series. We used the concept of correlation and implemented the following 4 methods of calculation:

a. Mathematical correlation

In this method, we use the covariance to calculate the correlation coefficient between the two vectors of real and simulated data

$$\rho = \frac{COVrs}{\sigma r \times \sigma s}$$

b. Shifting correlation

This method is based on the standard Euclidean norm (|| ||) and the standard scalar product (|*R.S*|). Note that the notion of shifting is to fix the real vector and do a sweep to the vector of simulated data to obtain a good convergence.

$$\rho = \frac{\sqrt{|R.S|}}{\|R\|}$$

Where R: Vector of real data.

S: Vector of simulated data.

c. Euclidian correlation

In this method, the calculation of the correlation coefficient uses the Euclidean norm and the distance measurement between the two data vectors ($\| \|_2$).

$$\rho = \frac{\sqrt{\|R - S\|}}{\|R\|}$$

Note: in both methods of Euclidean correlation (b and c) the normalization is imposed because we work on a circle of radius 1. Moreover, the results obtained by both methods are different because they were interpreted differently.

d. Surface correlation

This method takes into account the temporal dimension of our data (variables). In brief, it uses a standard (or standards of Manhattan) as the sum of the areas. In fact, we calculate the integral or the discrete sum of areas of rectangles formed by vectors over time. Unlike previous methods, this one takes into account the time of residence in each state.

$$\rho = 1 - \frac{|R - S|}{|R|}$$

If the simulated values converge to the real values, then the term |***R-S***| $\xrightarrow{p.s} 0$. To ensure a good correlation, this coefficient is constructed to approach the term 1.

III. Results

A. Experimentation

The data were collected over a one month period (Feb 2007) in a hospital suite at Charles Foix hospital - Paris. The subject was an elderly person aged 85, hospitalized after a fall.

B. Daily data simulation

An example of visualisation of resulted data involves the real and the simulated data of the 6th day (Fig. 7) . It is used to show the most important steps in data processing and simulation. The first diagram (top) shows the real data for 24. The two Markovian matrixes are then used to generate the simulated data (bottom).

Correlation	%
Mathematic	81.28%
Shifting	87.87%
Euclidian	34.94%
Surface	94.06 %

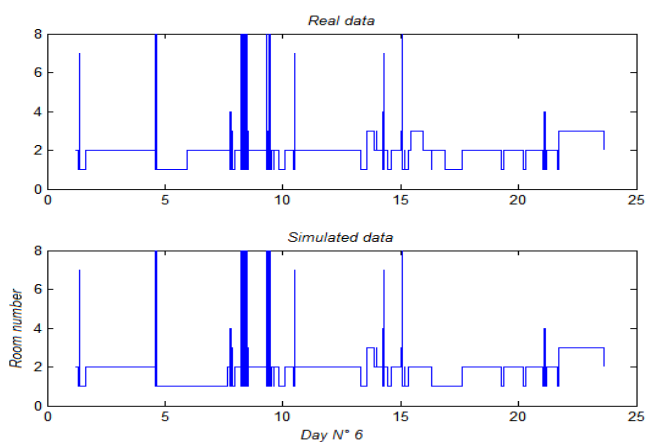

Figure 7: Ambulatogram of the 6th day with the table of the 4 correlations.

C. Results discussion

We found that the simulated data is similar to the real data, while the degree depends on the number of detections. To illustrate this point, we show in figure 8 the ambulatogram of the 16th day.

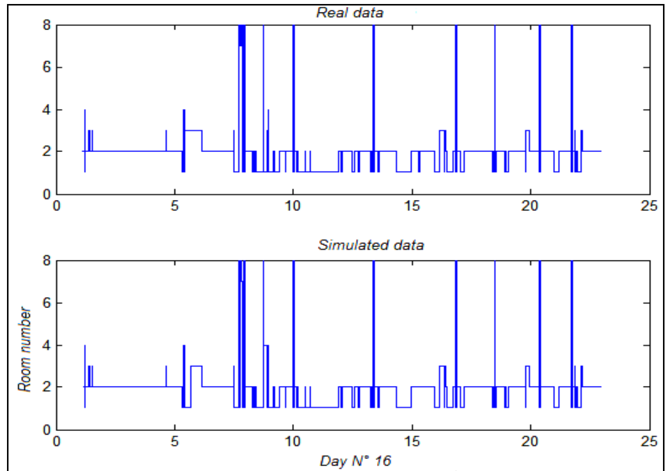

Figure 8: Ambulatogram of the 16th day, surface correlation is more than that on the 6th day.

Correlation	%
Mathematic	77.89%
Shifting	94.16%
Euclidian	45.92%
Surface	99.96%

A higher number of detections gave better similarity, due to the probabilistic method used. The 16th day has more detections than the 6th day with 99.96 % for the surface correlation which was the best methods, compared to other three. It appears that the simulated data of Markov model follows very closely the real data model, confirmed by an excellent surface correlation.

IV. Conclusion

Actimetric measurement within the Health Smart Home is an interesting expression of an individual's homeostasis in his environment, and it can be performed non invasively at home.

Simulation is an important tool in the field of engineering sciences, and it becomes even more important in the case of health applications. The methods of Artificial Intelligence proved their usefulness in the compartmental simulations. Despite the difficulty of modelling and simulation of human behaviour, the use of Hidden Markov Model gave encouraging results in our Health Smart Home application.

The concept of using correlation as a distance is an efficient method study the similarity between real and simulated data. The surface correlation is the more adaptable method among the four methods we tested.

References

[1] N. Noury, A. Tarmizi A, D. Savall D, P. Boissy, P. Barralon, G. Virone, P. Rumeau, “A smart sensor for the fall detection in daily routine”, SICICA2003, Aveiro, July 2003.

[2] J. L. Baldinger, J. Boudy, B. Dorizzi, J.-P. Levrey , R. Andreao, C. Perpère, F. Delavault, F. Rocaries, C. Dietrich, A. Lacombe, “Tele-surveillance System for Patient at Home: the MEDIVILLE system”, Congrès ICCHP 2004, Paris, Juillet 2004.

[3] T. Hadidi, N. Noury, “Model and Simulator of the activity of the elderly person in a Health Smart Home”. 12th IEEE International Conference on e-Health Networking Applications and Services (Healthcom), 2010, Lyon, France. Pp. 7-10

[4] AJ. Gordon, “The human patient simulator: acceptance and efficacy at a teaching tool for students”, *Acad Med 2000* May; 75(5):522.

[5] RL. Marshall, JS. Smith and al, “Use of a human patient simulator in the development of resident trauma management skills”, *J Trauma* July, 51(1):17-21, 2001

[6] S. Abrahamson, « Sim One – A patient simulator Ahead of Its Time », *Caduseus*, Autom 1997, Vol. 13, No. 2.

[7] N. Noury, G. Virone, P. Barralon, J. Ye, V. Rialle, J.Demongeot, “New Trends in Health Smart Homes”, in Healthcom2003, Santa-Monica-Califonia, 6-7 Jun 2003, pp 118-127

[8] N. Noury “AILISA : plateformes d’évaluations pour des technologies de télésurveillance médicale et d’assistance en gérontologie”, Journal Gérontologie et Société, n°113, Juin 2005, pp. 97-119.

[9] G. Le Bellego, N. Noury, G. Virone, M. Mousseau, J. Demongeot, “Measurement and Model of the Activity of a Patient in his Hospital Suite”, IEEE Trans. on Information Technology in Biomedicine, vol.10, Jan.2006, pp.92-9

[10] N. Noury, T. Herve, V. Rialle, G. Virone, E. Mercier, G. Morey, A. Moro, and T. Porcheron. “Monitoring behavior in home using a smart fall sensor and position sensors”. In Proc. 1st annual Conf on Microtechnologies in Medicine and Biology, pp. 607–610, (2000).

[11] N. Noury, C. Villemazet, A. Fleury, P. Barralon, P. Rumeau, N. Vuillerme, R. Baghai, “Ambient Multi-Perceptive System with Electronic Mails for a Residential Health Monitoring System”, in 28th IEEE-EMBC2006, New-York, 31 Aout - 3 sept. 2006, pp. 3612-15

[12] G C. Franco, F. Gallay, M. Berenguer, C. Mourrain, P. Couturier, “Non-invasive monitoring of the activities of daily living of elderly people at home – a pilot study of the usage of domestic appliances”, in Journal of Telemedicine and Telecare 2008, vol.14, pp. 231–235

[13] N. Noury, T.Hadidi, M.Laila, C. Villemazet, V. Rialle, A. Franco, “Level of Activity, Night and Day Alternation, and well being measured in a Smart Hospital Suite”, in Proc. IEEE–EMBC’08, Vancouver, Aug. 20-24, 2008, pp. 3328-31

[14] Noury N, Hadidi T, A relationship between the night-day levels of activities of elder patient in a health smart home, in proc. IAGG2009, the 19th Congress of Gerontology and Geriatrics, 5-9 July 2009, Paris.

[15] Hadidi T, Noury N, “A Predictive Analysis of the Night-Day Activities Level of older patient in a Health Smart Home”, proc. of the 7th International conference on smart homes and health telematics, ICOST2009, Tours, France, July 2009, pp. 290-293

[16] G. Virone, “Architecture et simulation locales du système d’information domotique-santé intégré à domicile (SID) pour la détection de situations a risque et l’aide a la décision”, University doctorate, Specialty Models and Instruments in Medicine and Biology, Université Joseph Fourier, Grenoble. 260 pages. 26 Novembre 2003.