Pervasive informatics and persistent actimetric information in health smart homes

Yannick Fouquet, Nicolas Vuillerme, and Jacques Demongeot*

TIMC-IMAG Laboratory,
Domaine de la Merci, 38710 La Tronche,
*Jacques.Demongeot@imag.fr
http://www-timc.imag.fr

Abstract. This paper discuss the ability to obtain a reliable pervasive information at home from a network of localizing sensors allowing to follow the different locations at which a dependent (elderly or handicapped) person can be detected. The data recorded can be treated as the sequence of color coding numbers of balls (symbolizing activity-stations) taken in a Polya's urn, in which the persistence of the presence in an activity-station is taken into account by adding a number of balls of the same color as the ball just drawn. We discuss the pertinency of such a procedure to early detect sudden or chronic changes in the parameters values of the random process made of the succession of ball numbers and we use it to trigger alarms.

1 Introduction

Since about 12 years [1]-[3], many experiments have been achieved for watching dependent people at home, in particular elderly and handicapped persons. In order to acquire data necessary to permit the alarms triggering, numerous sensors have been invented, in particular for localizing the person at home or in the surroundings. These localizers are on the body (like the GPS or the accelerometers), in the flat rooms (on the walls, like the infrared or radar detectors or on the ground, the bed or the chairs, like the pressure sensors, *cf.* Figure 1), on the doors (like the magnetic switches) or in gardens and streets (like the video-cameras). The sensors network is very important to follow up the dependent people during their walk trajectories inside home or outside. If the space/time data are acquired on healthy elderly people or on patients suffering from a neuro-degenerative disease, the sensors recording must be very well calibrated, to give birth to specific profiles of the time series corresponding to the successive locations in rooms inside the flat or in specific places inside a room [8]. A big hope comes from this ambient information to be able to detect a progressive stereotyped behavior (for the early diagnosis of a chronic disease like the Alzheimer one) or a sudden fall on the ground. The optimal use of this pervasive information implies fusion and scoring from the primary data, in order to detect minimal changes in time profiles: a way to do that is to considerably simplify the information by giving a color coding number to the different locations (pertinent for the

watching), and to follow up the succession of these numbers, *e.g.* by interpreting them as the succession of colors of balls drawn from a Polya's urn: in this kind of urn, the persistence (or *a contrario* the unstability) of an action in a location is represented by adding $n_i(k(i))$ balls of color $k(i)$, when a ball of color $k(i)$ has been obtained at time i. If $n_i(k(i))$ depends only on i through k, the random process constituted by the succession of the $n(k(i))$'s is called homogeneous and a change in homogeneity can be detected by estimating the $n(k)$'s and testing their significant consecutive differences. We will give in the following elements of material and methods in order to describe more precisely the data collection and treatment procedures, and then we will discuss the pertinence of such a research protocol and its implementation in the current life of dependent people at home.

Fig. 1. infrared (arrows) for localizing dependent people in a health smart home (left) & pressure sensors (right: FSA Seat 32/63 pressure mapping system, Vista Medical Ltd.)

2 Material and methods

A private apartment of an older woman, aged 80, at the Institution Notre-Dame (Grenoble, France) is equipped with a health integrated smart home (HsH). In general, the underlying principle of the HsH consists in continuously collecting data regarding her individual activity within her home environment and sending them to a telemedicine center via electronic mails (SMTP). As illustrated in Figure 2, our experimental health smart home includes 7 presence infra-red sensors (DP8111X, ATRAL), allowing the detection of the infrared radiations emitted by body surfaces (face, hands), and hence the monitoring of individuals successive activity phases within her home environment [10]-[13].

These different detections are timestamped and stored in a database (SQL) and then transmitted by e-mail through an attached file (XML). They permit
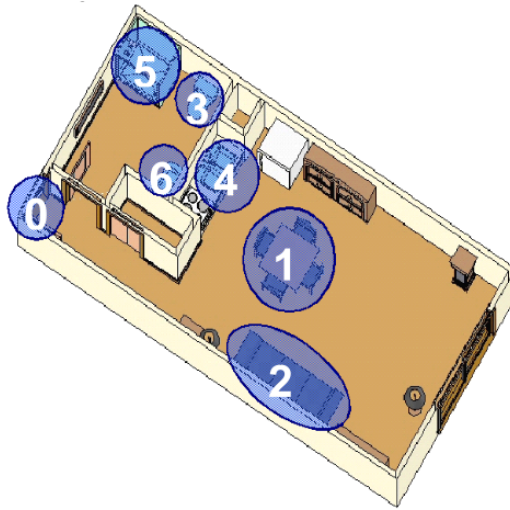
Fig. 2. Architecture of the experimental health smart home. Location sensors are placed at different places in the apartment, allowing the monitoring of individuals successive activity phases within his/her home environment: **0.** Entry hall - **1.** Living room - **2.** Bedroom - **3.** WC - **4.** Kitchen - **5.** Shower - **6.** Washbasin.

the continuous real-time surveillance on the screen of a dedicated workstation at the Hospital at Home (HaH) service which possesses nurses and doctors ready to visit the person at home in case of an acute pathologic problem or to transmit to a chronic disease service the information about the occurrence of a problematic change in the physiologic variables recorded at home (cf. Figure 3 left and center).

The data analysis of the records at home is primarily done through real-time updated descriptive statistics like presence histograms (Figure 3 right) but it is also achieved by using more sophisticated random processes techniques like time series or Polya's urns. The random processes made of the succession of the recorded localization data have been indeed already modelled by using classical time series techniques like Box-Jenkins auto-regressive processes in which were extracted the entropy [2] or the coefficients of the linear auto-regression [15]-[19]. In this paper,the information to be treated is reduced at the minimum and the only thing retained is the succession of the activity-station-codes corresponding to the successive locations of the elderly people at home. An important feature to extract from the random process made of the succession of these activity-station-codes is the breaking times at which a specific model of Polya's urn is no more avalaible, obliging to change the values of the parameters $n(k(i))$'s corresponding to the (algebraic, possibly negative) number of balls which must be added after obtaining a ball of color $k(i)$ at time i. It is supposed that if there is no pathologic change (sudden due to a fall or chronic due to the entrance in a neuro-degenerative disease), $n(k(i))$ is not depending explicitly on the time i, but only on the activity-station-code $k(i)$.
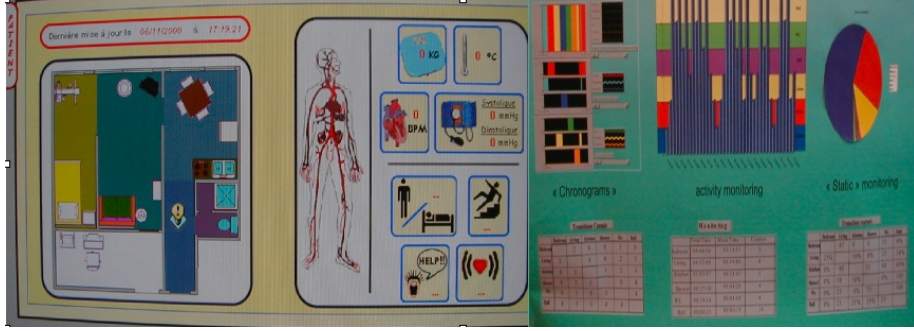
Fig. 3. Watching workstation at the HaH (Hospital at Home) service for the surveillance of dependent people at home

The first use of Polya's urns to represent persistence in a succession of qualitative data has been done since 25 years by climatologists for the sequence of dry and wet days [4,23], and a lot of fundamental [5,7,22], or more applied [6,9,14], papers have been after published for studying the theoretical properties of the corresponding random process, or for estimating its parameters or its thermodynamical variables (like the entropy of its stationary distribution).

The Polya's urn scheme is a Markov chain in which the balls are sequentially drawn from an urn initially containing a given number $a_0(j)$ of balls of the j-th color, $j = 1, \ldots, N$, and after each draw the ball is returned into the urn together with $n(j)$ new balls of the same color j. It is assumed that we observe at time i the $a_i(k)$'s (corresponding to the number of balls of color k drawn from the Polya's urn at time i) and $b_i = \sum_{c=1}^{N} a_i(c)$ balls and that we estimate the parameters $n(1), \ldots, n(N)$ supposed to be positive, by observing the frequencies in m trials of occurrence of balls of corresponding colors. For processing the estimation of $n(j)$'s, we consider the integer-valued random vector denoted $a_i = (a_i(1), \ldots, a_i(N))$ and distributed, if $n(j) \geq 0$, in the set:

$K_N = \{k = (k_1, \ldots, k_N) : k_i \text{ integers } / s = \sum_{c=1}^{N} k_c \geq b_0 = \sum_{c=1}^{N} a_0(c)\}$

according to:

$$\forall i \in \{1, \ldots, m\}, \\ P(\{a_i = k\} \mid a_{i-1}) = \begin{cases} \frac{a_{i-1}(j)}{b_{i-1}}, & \text{if } k_j = a_{i-1}(j) + n_i(j) \\ & \text{and } \forall r \neq j, k_r = a_{i-1}(r) \\ 0, & \text{if not,} \end{cases}$$

where $n_i(j)$ is the number of balls of color j added at time i. Let us denote $n_i = \sum_{c=1}^{N} n_i(c)$ and suppose that $n_i = Np > 0$, where p is the mean persistence rate (supposed to be independent of the time). When the n_i's are equal, the probability above reduces to:

$$\forall i \in \{1, \ldots, m\}, P(\{a_i = k\} \mid a_{i-1}) = \frac{a_{i-2}(j) + p}{b_{i-2} + p},$$

where j is the index in $\{1, \ldots, N\}$ for which we have: $k_j = a_{i-1}(j) + p$ and $\forall r \neq j, k_r = a_{i-1}(r)$.

Then we can calculate the probability $P(\{a_i = k\})$, when $k_j = a_0(j) + s_j p$, by using the formula:

$\forall i \in \{1, \ldots, m\}$,
$$P(\{a_i = k\}) = C_i^{k_1, \ldots, k_N} \frac{\left[\prod_{j=1}^{N} \prod_{s_j=0}^{\frac{k(j)-1}{p}} (a_0(j) + s_j p)\right]}{\prod_{s=0}^{i-1}(N + sp)},$$

where the k_i's verify: $k_i \geq 0$ and $\sum_{c=1}^{N} k_c = i$.

Let us now consider possible strategies of estimating the persistence p and the initial distribution a_0:

1) If we know the initial distribution of balls a_0, observing the empirical frequency $f(\{a_i = k\})$, we can estimate p by calculating the likelihood function and using the maximum likelihood estimator

2) If we do not know the initial distribution nor the persistence, we can:

- either estimate it by deciding that b_0 is fixed to a multiple of the number of activity-stations (*e.g.* twice this number) and by using a procedure similar to those proposed in [6], by supposing p known, and after deriving this initial estimation as function of p, finally trying to fix p at the integer value maximizing the likelihood function

- or, if the attempt above is not succeeding, to assume the uniformity of the initial distribution (*i.e.* decide that the initial number of balls was the same for each color/activity-station).

In the case where: $\forall j \in \{1, \ldots, N\}, a_0(j) = 1$, the probability of having the balls vector equal to k at time i becomes:

$\forall i \in \{1, \ldots, m\}$,
$$P(\{a_i = k\}) = C_i^{k_1, \ldots, k_N} \frac{\left[\prod_{j=1}^{N} (\prod_{s=0}^{\frac{k(j)-1}{p}} (1 + sp))\right]}{\prod_{s=0}^{i-1}(N + sp)},$$

where the k_i's verify: $k_i \geq 0$ and $\sum_{c=1}^{N} k_c = i$.

Then by replacing $P(\{a_i = k\})$ by $f(\{a_i = k\})$, we can estimate p. The empirical frequency $f(\{a_i = k\})$ is calculated from observations done at different days supposed to be independent (the initial distribution a_0 is suppposed to remain the same at the beginning of each day and the days are supposed to be independent). If there is 2 persistence parameters to estimate, *e.g.* p for the living (activity-station number 1) and p' for the other activity-stations, we can use a sequential probability ratio test (SPRT) procedure [21] by considering that there are only 2 super activity-stations codes, 1 for the living and 2 for the other activity-stations and by trying to estimate the best sampling size threshold allowing a significant decision in testing the hypothesis $H_0 \equiv \{p = p'\}$ (*i.e.* the persistence is the same in the two super activity-stations) against $H_1 \equiv \{p \neq p'\}$ (*i.e.* the persistence is different in the two super activity-stations).

3) If we have no information about the initial distribution of balls a_0 (even concerning the initial total number of balls b_0), but if we suppose that there is the same persistence in each activity-station, we can follow during a sufficient time the activity of the dependent person at home and estimate the conditional probability:
$$P(\{a_{i+1}(j) - a_i(j) = 1\} \mid \{a_i(j) = k\}) = \frac{a_0(j) + kp}{b_0 + ip}$$

By replacing the conditional probability above by the corresponding conditional empirical frequency (estimated from series of independent activity days for different activity-stations), we can get an estimation of p. We can also perform a test of $H_0 \equiv \{p = 1\}$ against $H_1 \equiv \{p > 1\}$, by comparing the empirical frequency of the event $\{d_{i+1}(j) = 1\} \cap \{d_i(j) = 1\}$ (i.e. the frequency to have consecutively the same color j, if $d_i(j)$ is the number (0 or 1) of balls of color j drawn from the Polyas's urn at time i) to its theoretical probability, which is binomial under H_0, with the probability to draw a ball j at time i equal to $\frac{a_0(j)}{b_0}$. When i increases, the estimation of $\frac{a_0(j)}{b_0}$ becomes rapidly very precise and allows the use of a classical test of comparison between an empirical and a theoretical frequency.

In the present case of persistence in activity-stations, we can assume that after a series of presence in an activity-station equal to or more than 2 recording intervals, if the activity-station changes, that involves a reset and we return to the distribution a_0. Then, we can use the following sequential procedure for tests

: i) initially as above $H_0 \equiv \{p = 1\}$ against $H_1 \equiv \{p > 1\}$,

ii) if H_0 is rejected, $H_1 \equiv \{p = 2\}$ against $H_2 \equiv \{p > 2\}$,

iii) if H_1 is rejected, $H_2 \equiv \{p = 3\}$ against $H_3 \equiv \{p > 3\}$, ...

until we reach, with the value of $p = k$ at the step k, a probability of activity-station changing (rejection of H_{k-1}) in k steps more than the threshold value 0,95.

3 Data and Results

The files treated bring together the data recorded in the flat of the elderly people in a period of 8 months from the 24th of March 2005 until the 25th of November 2005. The file follow the structure:

Table 1. The times and locations of the records

Day	Month	Year	Hour	Minute	Second	activity-station-code
24	03	2005	17	37	36	1

...

The columns represent successively the time (with the day, month, hour, minute and second of the recording) and the activity-station-code corresponding to the location of the watched person at this time.

From these records, we have tested different hypotheses about the persistence following the procedure given in the previous Section. We will give below on a short example a sketch of our testing strategy. The data for 200 record times were used for performing the two following steps:

i) - we calculated the empirical frequency $\frac{a_0(1)}{b_0} = \frac{58}{200} = 0.29$

- we use it for testing H_0 against H_1. The probability to observe 2 consecutive stays in the living (activity-station 1) is equal, under the hypothesis H_0 to $0.29 \times 0.29 = 0.084 \pm 0.06$ (the variance of an empirical frequency observed on a records sample of size i being estimated by $\frac{f(1-f)}{i}$). Then the hypothesis H_0 is rejected

with a significativity less than 1/1000: large deviations (with probability less than 1/1000) of the number of pairs of consecutive stays in living start at the value 29, and there are 31 such pairs in the records .

ii) by pursuing the sequence of tests, we found that $p = 3$ is the best estimation of the parameter of persistence in the living, because it is the first integer giving probabilities 6/10, 6/13, 6/16, 6/19 and 6/22 of exiting from the living after respectively 1, 2,..., 6 stages in this activity-station. These probabilities have been estimated by the corresponding empirical frequencies of exit from the living after 1, 2,..., 6 stages. These empirical frequencies were respectively equal to $14/24 \pm 0.06, 4/9 \pm 0.06, \ldots, 1/3 \pm 0.07$ in the experimental records of 200 sampling times.

4 Discussion

We have assumed in the previous calculations 5 important hypothesis we can now discuss:

1) the activity is homogeneous in time and space inside a day, *i.e.* we have the same persistence for each activity-station sojourn and a reset of the persistence memory at the end of a sojourn.

We have surely a persistence more important in activity-stations in which several tasks can be done involving a long time investment, compared to stereotyped and standardized tasks done in other activity-stations.

2) the activity records sequence is a Markovian process, for which the future depends on the past only through the present.

There are surely some breaks of the Markovian character, specially during activities asking for more attention (like cooking or reading), in which a time series approach would be more convenient than the Polya's urn modelling (the classical time series treatment involves the extraction of a tendency through a mobile time window, and then the calculation of a time linear regression order [2]).

3) the role of the activity-stations is symmetrical, *i.e.* each activity-station generates the same initial conditions in the initial distribution of balls (representing activity-stations) in the Polya's urn.

Because of many differences of surface, functionality, illumination, the activity-stations are not playing the same role and have different weights after resetting, depending on the time in a day (certain tasks being executed only once at a fixed hour of the morning or afternoon).

4) the persistence is non depending on the time

In fact, there are nycthemeral variations of activity ([15,16,18]) as well as seasonal effects we have to take into account for making more precise the statistical structure of the persistence. A remanence of the persistence surely exists, especially at the end of day where the level of awakeness and attentiveness is diminished.

5) successive days can participate to the same independent identically distributed (iid) sampling.

In fact, there is certainly a dependency linked to the place of the days in the week (Saturday being for example used for recapitulating the working days activity and for anticipating the leisure organization of Sunday).

5   Conclusion

The monitoring of dependent people at home allows the recording of their activity, especially the activity-station changing sequences, what is very useful to detect deviances with respect to the normal behavior. The detection of large deviations from the "normal" individual distribution of the random process retained for the ordinary walking of a dependent person inside his flat, permits to anticipate the fall, whose risk is high and renders it ineluctable a day, after 80 years. The fixed or embarked localizing sensors give sufficient indications to trigger an alarm at the level of the patient (for a real-time correction of a desequilibrium, in case of vestibular pathologies [20]) or at the level of the Hospital at Home service (for an emergency sending nurses or doctors, depending on the gravity of the detected dysfunctioning). The body sensors are incorporated in ordinary clothes rendering the surveillance ergonomically acceptable. We are now developping techniques for studying (like for a drug), the "toxicity" of the watching system, toxic here meaning unaesthetic, intrusive, invasive and/or pathogenic, the level of toxicity depending on the "compliance" of the recorded subject. We present the beginning of such studies in a companion paper in the present issue [20] and we will develop further psycho-physic studies for the determination of the liminal level of sensitivity/specificity and of the level of rejection, necessary for quantifying the degree of acceptability of the sensors network studied in this paper.

References

1. P. Couturier, A. Franco, J.F. Piquart, J. Mansotte, C. Montani, C. Suarez, A. Mollier, C. Gucher, M. Frossard, L. Nicolas, G. Jasso Mosqueda, M.C. Mouchet, A.M. Argentier, J.L. Bosson, P. Carpentier, J. Demongeot, "Telegerontology": from teleassistance to teleconsultation of elderly people. Mythe or reality?, Rev. Griatrie 21 (1996) 23-31.
2. S.K. Das, N. Roy, A. Roy, Context-aware resource management in multi-inhabitant smart homes: A framework based on Nash H-learning, Pervasive and Mobile Computing 2 (2006) 372404.
3. J. Demongeot, G. Virone, F. Duchne, G. Benchetrit, T. Herv, N. Noury & V. Rialle, Multi-sensors acquisition, data fusion, knowledge mining and alarm triggering in health smart homes for elderly people, Comptes Rendus Biologies 325 (2002) 673-682.
4. E. Galloy, A. Le Breton, S. Martin, A model for weather cycles based on daily rainfall occurence, in: Rhythms in Biology and other Fields of Application, J. Demongeot et al. eds, Lecture Notes in Biomathematics 49 Springer-Verlag, Berlin, 1983, pp. 303-318.

5. K. Inoue, S. Aki, Polya urn models under general replacement schemes, J. Japan Statist. Soc. 31 (2001) 193-205.

6. G. I. Ivchenko, Estimation in the Markov-Polya Scheme, Mathematical Notes 64 (1998) 322-329.

7. S. Kotz, H. Mahmoud, P. Robert, On generalized Polya urn models, Statistics & Probability Letters 49 (2000) 163-173.

8. G. Le Bellego, N. Noury, G. Virone, M. Mousseau, J. Demongeot, A Model for the Measurement of Patient Activity in a Hospital Suite, IEEE Transactions ITB 10 (2006) 92-99.

9. C.F. Martin, Y.C. Ho, Value of information in the Polya urn process, Information Sciences 147 (2002) 6590

10. N. Noury, G. Virone, J. Ye, V. Rialle, J. Demongeot, New trends in Heath smart homes, ITBM-RBM 24 (2003) 122-135.

11. N. Noury, G. Virone, P. Barralon, J. Ye, V. Rialle, J. Demongeot, New trends in Health Smart Homes, in: IEEE Healthcom 2003, IEEE Proceedings, Piscataway, 2003, pp. 118-127.

12. V. Rialle, N. Noury, L. Bajolle, J-P. Lamy, G. Virone, F. Duchne, N. Moha, J. Demongeot, Le concept d'Habitat Intelligent pour la Sant : considrations techno-scientifiques pour un service mdico-social, Revue de Griatrie 28 (2003) 403-416.

13. V. Rialle, F. Duchne, N. Noury, L. Bajolle, J. Demongeot, Health 'smart' home: Information technology for patients at home, Telemedicine Journal and E-Health 8 (2003) 395-409.

14. W.F. Rosenberger, T.N. Sriram, Estimation for an adaptive allocation design, Journal of Statistical Planning and Inference 59 (1997) 309-319.

15. G. Virone, N. Noury, J. Demongeot, A system for automatic measurement of circadian activity deviations in telemedicine, IEEE Trans. Biomed. Eng. 49 (2002) 1463-1469.

16. G. Virone, N. Noury, J. Demongeot, The health integrated integrated smart home information system (HIS2) : monitoring the nycthemeral circadian rhythms divergences, in : IEEE Healthcom 2002, J.P. Thomesse et al. eds, LORIA, Nancy, 2002, pp. 57-62.

17. G. Virone, D. Istrate, M. Vacher, N. Noury, J.F. Srignat, J. Demongeot, First Steps in Data Fusion Between a Multichannel Audio Acquisition and an Information System for Home Healthcare, in: IEEE EMBC 2003, IEEE Proceedings, Piscataway, 2004, pp. 1364-1367.

18. G. Virone, B. Lefebvre, N. Noury, J. Demongeot, Modeling and computer simulation of physiological rhythms and behaviors at home for data fusion programs in a telecare system, in: IEEE Healthcom 2003, IEEE Proceedings, Piscataway, 2003, pp. 111-117.

19. G. Virone, N. Noury, J.P. Thomesse, V. Rialle, J. Demongeot, A home health information system based on the can fieldbus, In : FeT2003 Aveiro, 5th IFAC Int. Conf. On Fieldbus Systems and their applications, in: IFAC Proceedings, Elsevier, Amsterdam, 2003, pp. 270-281.

20. N. Vuillerme, N. Pinsault, O. Chenu, A. Fleury, Y. Payan and J. Demongeot, A wireless embedded tongue tactile biofeedback system for balance control, IPAL09.

21. A. Wald, Sequential analysis , J. Wiley, 1947.

22. S. Walker, P. Muliere, Beta-Stacy processes and a generalization of the Polya-urn scheme, The Annals of Statistics 25 (1997) 1762-1780.

23. L.L. Wilson, D.P. Lettenmaier, E. Skyllingstad, E., A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation. J. Geophys. Res. 97 (1992) 27912809.