AAAI Technical Report FS-12-01
Artificial Intelligence for Gerontechnology

How Is Grandma Doing? Predicting Functional Health Status from Binary Ambient Sensor Data

Saskia Robben
s.m.b.robben@hva.nl
Amsterdam University
of Applied Sciences
Duivendrechtsekade 36-38
1096AH Amsterdam
The Netherlands

Gwenn Englebienne
g.englebienne@uva.nl
University of Amsterdam
Science Park 904
1098XH Amsterdam
The Netherlands

Margriet Pol
m.c.pol@hva.nl
Amsterdam University
of Applied Sciences
Tafelbergweg 51
1105BD Amsterdam
The Netherlands

Ben Kröse
b.j.a.krose@uva.nl
University of Amsterdam
Science Park 904
1098XH Amsterdam
The Netherlands

Abstract

Ambient activity monitoring systems produce large amounts of data, which can be used for health monitoring. The problem is that patterns in this data reflecting health status are not identified yet. In this paper the possibility is explored of predicting the functional health status (the motor score of AMPS = Assessment of Motor and Process Skills) of a person from data of binary ambient sensors. Data is collected of five independently living elderly people. Based on expert knowledge, features are extracted from the sensor data and several subsets are selected. We use standard linear regression and Gaussian processes for mapping the features to the functional status and predict the status of a test person using a leave-one-person-out cross validation. The results show that Gaussian processes perform better than the linear regression model, and that both models perform better with the basic feature set than with location or transition based features. Some suggestions are provided for better feature extraction and selection for the purpose of health monitoring. These results indicate that automated functional health assessment is possible, but some challenges lie ahead. The most important challenge is eliciting expert knowledge and translating that into quantifiable features.

Introduction

Western societies are faced with an aging population and a lot of research initiatives exist to assist the elderly in living independently at home. Particularly activity monitoring with ambient sensors is a fast developing field, with several subtopics. Monitoring the activities of daily living (ADL) is useful as Katz' Index of ADL has been developed as an instrument to evaluate the functioning of an elderly person (Katz et al. 1963). An inquiry as to what ADL are most relevant for monitoring with ambient sensors (Alizadeh et al. 2011) indicated functional transfers and food consumption as the two most important activities. Activities can be monitored in a home environment with ambient binary sensors in combination with supervised methods such as a HMM (van Kasteren 2010). Analyzing deviations in the daily patterns can be done by defining circadian rhythms based on the average time spent in a room and calculate the difference (Virone et al. 2008). Deviations could indicate changes in health status. Another study focusses on the evaluation of the quality of specific activities. In a lab setting subjects were instructed to perform an activity such as meal preparation, either correctly or with a certain error. An algorithm was developed to detect these errors (Cook and Schmitter-Edgecombe 2009).

The issues described above are only a small part of the field, but some issues still are open. A recent review of lifestyle telemonitoring technologies (Brownsell et al. 2011) noted that little attention goes to the question of how to use the detection of changes in activity patterns for follow-up, though there are some case studies described to indicate possible use of a telemonitoring system (Glascock and Kutzik 2006). Several research groups are now focusing on intelligent alert systems, but an ambient monitoring system which could be used for functional health assessments does not exist to our knowledge.

In this paper we address this void by examining the use of an ambient monitoring system for functional health assessment. The activity patterns derived from the sensors are mapped to the AMPS (Assessment of Motor and Process Skills) (Fisher 1999). This is a common validated metric of functional health which is actually used in practice, and was suggested for this use in a review article on intelligent technology for an aging population (Pollack 2005). The challenge is finding a relation between patterns in sensor data and the functional health status.

The contribution of this paper is modeling this relation between binary ambient sensor data and the AMPS. The main research question is: Can intelligent monitoring systems based on these models be used to assess the functional health status of a person? In figure 1 our approach is depicted, the models will be used to predict the AMPS score, and this has to be compared with the assessment of a medical specialist.

To this end we first explain a little more about the nature of our data, both the sensor data and the AMPS as functional health status data. After collecting this data, several questions need to be addressed: How to derive features from the sensor data? After constructing features, how to select the best ones? And finally, can the functional status be learned from the subsets of features?

Feature extraction and selection methods will be explored and expert knowledge will be used for answering these questions. Then a linear mapping will be made from selected feature sets to the functional status of a person.

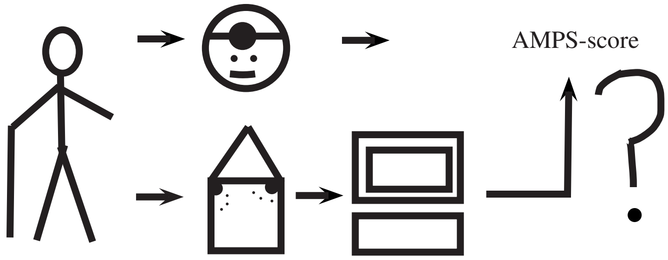

Figure 1: The upper part depicts a medical specialist who makes an assessment of the functional health status of the elder person (AMPS). The lower part depicts an intelligent machine (the models described in this paper) which uses data from a sensor equipped elderly home for predicting the AMPS. The question is whether such an intelligent system can perform an assessment similar as a medical specialist.

Related work

Though a lot of research related to activity recognition or health assessment with ambient sensor technology is grounded in lab settings, an increasing number of studies involves elderly people in their own home. A lot of these report case studies on the use of telemonitoring systems in which typically medical specialists are consulted to review deviations found in the sensor data.

An interesting methodology is comparing health-events such as falls or hospitalizations with sensor data using web-based visualization software (Rantz, Skubic, and Miller 2009). On the possible use of a telemonitoring system a lot of case studies are described (Glascock and Kutzik 2006).

Besides case studies there is also work towards actual health assessment and developing an alert system (Rantz et al. 2012). The main goal of their study is supporting health care providers and evaluating an early warning system by comparing a group with a telemonitoring system with a baseline group. However, the data which is collected provides possibilities to develop a health care assessment system similar as described in this paper.

Approach

In this section some background is given on the data and methods used in the experiment.

Table 1: AMPS score during the first and second measurement. Last column provides information on the number of sensors in the apartment of the subject.

Subject	AMPS-1	AMPS-2	# Sensors
1	1.3	1.3	12
2	2.21	1.8	13
3	2.05	2.54	14
4	0.57	0.97	13
5	3.09	2.43	13

Data Collection

Data was collected of five volunteers living in an independent assisted living environment. The binary ambient sensor network which continuously collects data consists of motion sensors (passive infrared), magnetic sensors, a floating sensor for the toilet and a bed mat. We use the same technology and design principles as van Kasteren et al. (2008) for setting up the sensor network.

The apartments were equipped with an average of 13 binary sensors. As the apartments are furnished differently, the topology of the sensor network differs slightly for each apartment.

The behavior of a person triggers sensors and these are stored in a database as a set of sensor events E, where each sensor event $e_x \in E$ reflects a single sensor event. For using this high dimensional set, a feature set F has to be extracted. This process is described in the section about feature sets.

The participants engage in several research activities related to the sensor network, such as interviews where they are asked to give their opinion on the development of sensor systems or experimenting with tablets which display their activities (Kanis et al. 2011).

The subjects are also periodically visited by an occupational therapist to assess their functional status. Every three months multiple instruments are used to assess the health status of the subjects. From two of these assessments the AMPS is selected for use in this paper.

Two measurements of five subjects results in a total of 10 data points, of which details are given in table 1.

Functional status: AMPS

Both physical and cognitive functioning contribute to the level of which a person can perform activities of daily living. The Assessment of Motor and Process Skills (AMPS) (Fisher 1999) instrument differentiates between motor and process skills. The process score contributes to functioning as it is related to cognitive skills, while the motor skills represents a direct physical decline.

The AMPS is used to assess skills in performing daily activities. The AMPS comprises 12 items of motor skills and 20 items of process skills. The skills were observed in two of 56 standardized daily activities. Scores are linked to a continuous scale of ability in motor or process functioning (range from -3 to 4). Scores above the cut-off point in motor skills (2.0) or in process skills (1.0) indicate that persons are able to functioning independently in the community.

Though normally a specialist will consider both the motor and process part of the AMPS, in this paper we focus on the physical part of the AMPS score, which ranges from -3 to 4.

Models

Modeling the relation between the sensor data and the AMPS is a regression problem, and the basic method for such problems is linear regression. Other (non-linear) methods are often better, but considering the low amount of data no advanced modeling techniques are explored. However, one modeling technique specifically performs well on regression problems with a low amount of data, and that is the use of Gaussian processes (Rasmussen and Williams 2006).

Both models are trained with a subset of the data and subsequently used for predicting the AMPS score of another set. The input for the models consists of a set of features which are extracted from the sensor data.

Linear regression This linear method basically learns a weight for each of the features in a particular set. Below a formalization of a linear regression model.

$$AMPS = \sum_{i=1}^{|F|} w_i \cdot f_i \tag{1}$$

where F is the set of selected features, with $f_i \in F$ and W is the set of weights, with $w_i \in W$. The weights are learned from the training data and can be used to predict the AMPS score of a test person.

Gaussian processes Linear regression uses a single function, which is learned by minimizing the error on the training set: this form of maximum likelihood learning is prone to overfitting and consequently to suboptimal performance on the unseen data. In contrast, rather than computing the most likely model given the training set, Gaussian processes consider infinitely many models, weighed by their probability given the training set, and marginalise out the individual models. As a consequence, with suitable priors they are not prone to overfitting and are therefore particularly well-suited for complex and noisy datasets which arise when human behavior is analyzed (Englebienne 2011).

Formally, a Gaussian process defines a distribution over functions by considering that a function is fully defined by the infinite vector of its outputs. Rather than considering a functional mapping between inputs and outputs, the GP specifies a joint Gaussian distribution over any finite subset of the output values. This is tractable, because in practice a function need not be evaluated for every possible input: it is sufficient to evaluate the function for the finite number of points for which we have an input. The process is fully specified by the mean and covariance matrix of the Gaussian distribution, where the mean is typically chosen to be zero and the covariance matrix is specified by a kernel function.

The prior (kernel function) or other hyperparameters were not optimalized for this particular data set. We used the default settings of the algorithm as implemented in the software package WEKA for predicting the AMPS on the test data (Witten, Frank, and Hall 2011), this comprises a radial basis function (RBF) kernel.

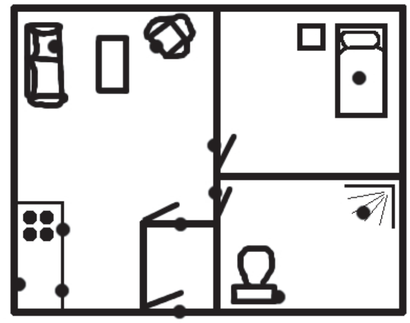

Figure 2: Baseline features. The number of sensor events is stored for each sensor.

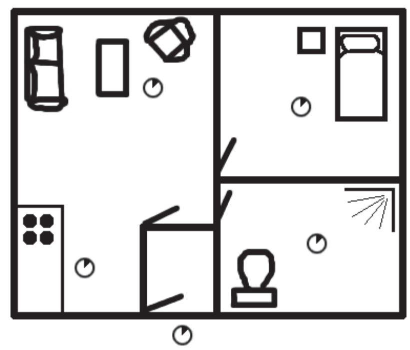

Figure 3: Concept of location based features. The features represent the time spent in a certain area.

Feature sets

The two different modeling methods are applied to several projections of the data. Though numerous methods exist for feature extraction of data of this nature, in this paper the focus lies on using expert knowledge. An occupational therapist was consulted as a domain expert to obtain more information on which features might be relevant in assessing functional status.

Domain experts (occupational therapists) typically look at the general pattern of activities. Also global patterns based on the location of the resident provide useful information. They are more interested in shifts which develop over weeks than in small daily deviations.

When specific activities are considered three characteristics are identified as relevant. These are: the frequency of an activity, the duration of an activity and the time of day the activity is carried out. These characteristics can be transformed into features, after which experts can further guide the feature selection process as some characteristics are more relevant to some activities than to others. E.g. time of day for meals are important, but for toilet visits the frequency is more important (possibly only differentiated by daily or nightly visits).

Finally specialists are interested in transitions between activities as well as transitions between locations, as a measure for indoor mobility.

To summarize, specialists are interested in global location-based patterns, activities with specific characteristics and transitions. In this paper we address the location-based patterns and transitions by translating this information into quantifiable features. A description of the resulting three different feature sets used for the experiment is given below. Figures 2 and 3 are displaying the differences between the baseline set and the location set.

Baseline feature set This feature set is based on the number of firings of specific sensors recorded during one week. The set of sensors is indicated with S, the set of sensor events with E, and F_b is used to denote a set of features. The first set of features is the raw count of sensor events. For each sensor $s_a \in S$:

$$f_{ba} = \sum_{e_i \in E} \begin{cases} 1 \text{ if } e_i = s_a \\ 0 \end{cases} \tag{2}$$

This results in 14 features while there are only 10 data points available. For avoiding that the model is too complex in relation to the low number of training data points and thus prone to overfitting some feature selection has to be done. To find the best set a model selection method with a penalty for complexity could be applied, but the main goal is to have a baseline for comparison with the other two sets. Because the location based set contains five features and the transition set six, this set is limited to six features also.

We used Pearson's correlation coefficient R for each of the features in this set (with the AMPS score). Its square R^2 is used to rank the features (Guyon and Elisseeff 2003) and subsequently the top six features are selected.

Location feature set This set consist of features which represent time spent in a certain location in the house. The apartment of an elderly person consists of several locations where the person can be: 'Bedroom', 'Bathroom', 'Kitchen' and 'Sofa'. A person can also leave the house (i.e. 'Outside'). Based on the sensor data the location of the person is inferred, and it is possible to continuously track the location of a resident.

Below a description of the algorithm.

1. Select: Extract data from the database (all sensor events from selected period).
2. Label: Label data based on sensor location (i.e. label sensor events as 'Bedroom', 'Bathroom', 'Kitchen', 'Outside', 'Sofa'.)
3. Segment: Replace consecutive series of events with the same label with two instances: an item marking the start and an item marking the end of the series.
4. Label time slots: Determine a label for each minute on each day (by means of the following procedure):
   - Set n=1440 (adjustable to allow for a different granularity than minutes)
   - Use the start and end marking of a series as boundaries. Each time slot in between gets the associated label.
   - This results in an matrix sized n*#days.
5. Feature extraction: Calculate the time spent in each location. The set of locations is indicated with L, the set of labeled time slots with T, and F_l is used to denote a set of features. For each location $l_a \in L$:

$$f_{la} = \sum_{t_i \in T} \begin{cases} 1 \text{ if } L(t_i) = l_a \\ 0 \end{cases} \tag{3}$$

The algorithm to extract activity categories is based on the following assumption: *a person is always somewhere*. This is a useful but violated assumption. There are a few places in the apartment where no sensor data is obtained. Either because we did not install a sensor in that location (e.g. a walk-in closet which is not used often). But also when sensors are broken we miss information.

Consider the following scenario: a person is sitting in his living room, when he hears noise at the back door. He stands up and goes outside for a few minutes to check and returns to the living room. However, the sensor on the back door is broken. The tracker will mark all this time as being spent in the living room.

From a machine learning perspective however, it is not a problem that the features possibly represent a slightly different concept than intended. The methods we use will reveal whether this is a proper feature set or not, regardless of the underlying concepts.

Transition feature set The locations are used to calculate the total number of transitions between locations, which is used as a feature.

The set of locations is indicated with L, the set of labeled time slots with T, and f_t is used to denote the transition feature.

$$f_t = \sum_{t_i \in T} \begin{cases} 0 \text{ if } L(t_i) = L(t_{i+1}) \\ 1 \end{cases} \tag{4}$$

In combination with the location feature set this feature forms the transition feature set:

$$F_t = \{F_l, f_t\} \tag{5}$$

Experiment

The goal of this experiment is testing the models developed for predicting the functional health status. To this end the previously described feature sets and models are used.

Methods

Sensor and functional status data is selected from five subjects who are each visited twice. The sensor data of exactly one week following these assessments is selected. This results in a total of 10 data points.

For predicting the functional health status from sensor data all combinations of models and feature sets are explored. As listed in table 2 the three different subsets are compared in combination with linear regression and Gaussian processes. Leave-one-person-out cross validation is

Table 2: Summary of experimental setup.

Setup	Features	$\|F\|$	Model
1	F_b	6	Linear regression
2	F_l	5	Linear regression
3	F_t	6	Linear regression
4	F_b	6	Gaussian processes
5	F_l	5	Gaussian processes
6	F_t	6	Gaussian processes

Table 3: The results with the different setups. Different feature sets in combination with linear regression (LR) and Gaussian processes (GP).

Setup		E_{RMS}
1	LR F_b	1,27
2	LR F_l	1,72
3	LR F_t	2,26
4	GP F_b	0,83
5	GP F_l	0,85
6	GP F_t	0,87

used and thus the process of feature selection and validation is repeated five times. The error measure we use is root mean square.

$$E_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - AMPS_i)^2} \tag{6}$$

where P_i is the prediction of the AMPS score and n the number of data points.

Results and Conclusion

In table 3 the error scores are provided for each combination of model and feature set. From these results it can be concluded that Gaussian processes perform better on this data than the linear regression model and for both models the baseline set performs better than both the location set as the transition set. The aim of this study was to model the relation of binary ambient sensor data from a telemonitoring system to functional health status (AMPS). The main question was whether an intelligent monitoring system based on such models could be used to assess functional health of a person. Though this study showed that some models perform better than others, and some feature sets perform better than others, based on this error scores this question can not be answered conclusively with a yes, as will be discussed below.

Discussion

The Gaussian processes outperforming the linear regression model was as expected, but that the baseline feature set yielded the best results was not as expected. This is probably caused by the feature selection process of the baseline set, which was not independent of the learning process.

Regarding the performance, for an honest comparison with the ground truth the typical error of occupational therapists should also be taken into account. But though the error compared with the assessment of the occupational therapist is substantial, the results presented in this paper can function as a baseline.

The challenges which lie ahead to achieve a better performing intelligent functional health assessment system can be grouped into different areas.

The first is applying more advanced modeling techniques for achieving better performance on this data set. The field of machine learning provides enough ideas for applying more advanced feature selection methods (Guyon and Elisseeff 2003) and more advanced models. It can be expected that fine tuning the best configuration yield a better performance. Examples include using functions on the input variables or in the case of Gaussian processes include hyper parameter learning.

An issue which has to be taken into account here is the noisy nature of the sensor data. Noise could be a result of temporary hardware failure, but also visitors can change the pattern. Even seasonal related alternative behavior might influence the performance of an intelligent functional health assessment system. A model should be robust to this type of noise but sensitive in detecting and processing relevant changes.

Besides learning from the machine learning community, a lot could also be gained from knowledge extraction of health care specialists. It is a real challenge to create common ground between analysts and medical specialists. Parallel to this study another study of our group addresses the question what is the best way for visualizing deviations in daily activity patterns, as well as the question of which deviations are problematic. The results of this study are expected to contribute to the domain knowledge and leads to further refined features.

Using the current knowledge for improving the feature extraction process is also a challenge. For example domain experts suggested taking the transitions between locations into account besides the plain amount of time spent in a location. This was translated into a feature which counted the total number of transitions, but this could be improved by adding weights for different transitions to reflect actual distance. For example a transition from the kitchen to the living room should get a higher weight because they are further apart than a transition from the bedroom to the adjacent bathroom.

Another possible area for improvement is a critical evaluation of the metric for assessing health. Our belief is that the ambient sensor network is best suited for a functional health assessment, such as the AMPS. But perhaps the network is better in predicting related concepts such as frailty or walking speed. The assessment of our subjects consists of several more metrics, so this is a possibility for future experiments.

The last challenge is related to data collection. Subjects not only have to embrace ambient sensor technology in their homes, but also health assessments have to be made regularly. Ideally more data-points are collected, for building models which are more reliable. Among the target audience for an intelligent health system as proposed in this paper are independent living elderly. Typically these elderly function quite well and therefore may not be in sight of the health care professionals yet. As there is not an immediate urgency to install a sensor network for reassuring purposes, recruiting volunteers for this type of research is quite a challenge. Then convergence towards an intelligent system for functional health assessment becomes real, and it could assist the specialists around elderly people.

Acknowledgments

This research was supported by the 'Health-lab' project in the Dutch Pieken in de Delta program and the 'Smart Systems for Smart Systems' project in the SIARAAK program. We thank the residents from 'The Flank' for participating and Zorggroep Vivium Naarderheem, a communal care group in the Netherlands, for facilitating the contact with our volunteers. Also we thank Bas Terwijn for providing technical support.

References

Alizadeh, S.; Bakkes, S.; Kanis, M.; Rijken, M.; and Kröse, B. 2011. Telemonitoring for assisted living residences: The medical specialists' view.

Brownsell, S.; Bradley, D.; Blackburn, S.; Cardinaux, F.; and Hawley, M. 2011. A systematic review of lifestyle monitoring technologies. *Journal of Telemedicine and Telecare* jtt–2010.

Cook, D., and Schmitter-Edgecombe, M. 2009. Assessing the quality of activities in a smart environment. *Methods of information in medicine* 48(5):480–485.

Englebienne, G. 2011. Bayesian methods for the analysis of human behaviour. *Computer Analysis of Human Behavior* 320.

Fisher, A. 1999. *Assessment of motor and process skills*. Three Star Press Fort Collins, CO.

Glascock, A., and Kutzik, D. 2006. The impact of behavioral monitoring technology on the provision of health care in the home. *Journal of Universal Computer Science* 12(1):59–79.

Guyon, I., and Elisseeff, A. 2003. An introduction to variable and feature selection. *The Journal of Machine Learning Research* 3:1157–1182.

Kanis, M.; Alizadeh, S.; Groen, J.; Khalili, M.; Robben, S.; Bakkes, S.; and Kröse, B. 2011. Ambient monitoring from an elderly-centred design perspective: What, who and how. *Ambient Intelligence* 330–334.

Katz, S.; Ford, A.; Moskowitz, R.; Jackson, B.; and Jaffe, M. 1963. Studies of illness in the aged. *JAMA: the journal of the American Medical Association* 185(12):914–919.

Pollack, M. 2005. Intelligent technology for an aging population: The use of ai to assist elders with cognitive impairment. *AI magazine* 26(2):9.

Rantz, M.; Skubic, M.; Koopman, R.; Alexander, G.; Phillips, L.; Musterman, K.; Back, J.; Aud, M.; Galambos, C.; Guevara, R.; et al. 2012. Automated technology to speed recognition of signs of illness in older adults. *Journal of Gerontological Nursing* 38(4):18–23.

Rantz, M.; Skubic, M.; and Miller, S. 2009. Using sensor technology to augment traditional healthcare. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, 6159–6162. IEEE.

Rasmussen, C. E., and Williams, C. K. I. 2006. Gaussian processes for machine learning.

van Kasteren, T.; Noulas, A.; Englebienne, G.; and Kröse, B. 2008. Accurate activity recognition in a home setting. In *Proceedings of the 10th international conference on Ubiquitous computing*, 1–9. ACM.

van Kasteren, T. 2010. *Activity Recognition for Health Monitoring Elderly using Temporal Probabilistic Models*. Ph.D. Dissertation, Universiteit van Amsterdam.

Virone, G.; Alwan, M.; Dalal, S.; Kell, S.; Turner, B.; Stankovic, J.; and Felder, R. 2008. Behavioral patterns of older adults in assisted living. *Information Technology in Biomedicine, IEEE Transactions on* 12(3):387–398.

Witten, I.; Frank, E.; and Hall, M. 2011. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.