A probabilistic approach to automatic health monitoring for elderly

Tim van Kasteren Ben Kröse

Intelligent Systems Lab,
University of Amsterdam ,
Kruislaan 403, 1098 SJ, Amsterdam , The Netherlands
tlmkaste@science.uva.nl krose@science.uva.nl

Keywords: Activity Recognition, Temporal Sensor Patterns, Dynamic Bayesian Networks, Context Awareness

Abstract

The growing population of elders in our society calls for a new approach in caregiving. By inferring what activities elderly are performing in their houses it is possible to determine their physical and cognitive capabilities. In this paper we describe probabilistic models for performing activity recognition from sensor patterns. We introduce a new observation model which takes the history of sensor readings into account. Results show that the new observation model improves accuracy, but a description using less parameters is likely to give even better results.

1 Introduction

The population of people aged 80 or over is projected to increase more than five times by the year 2050 [1]. Because the healthcare infrastructure is most likely not able to handle this growth, it is suggested to let elders live independently at home longer. However, to make it possible for the elders to safely continue to live their lives at home, some form of automatic health monitoring is required.

Several approaches have been attempted by various researchers. Ohta et al. [2] installed infrared sensors in each room of a house and performed health monitoring by comparing the duration of stays in specific rooms with previously recorded data. Aipperspach et al. [3] used various binary sensors and describe a system capable of predicting which sensor will fire next. These predictions can be used to recognize anomalous behavior and deviations from routine. However, both these approaches operate purely on low level sensor data. Incorporating high level information is intuitively sensible and has been shown to give more reliable results [4].

The type of high level information caregivers use themselves are activities of daily living (ADLs). These ADLs are routine activities that people tend to do everyday, such as eating, bathing, dressing and toileting. By giving each ADL a performance score, a caregiver is able to determine a persons' mental and physical abilities [5].

Using activities of daily living (ADLs) as a basis for an automatic health monitoring application has already been attempted by several people (e.g. [6, 7]) and wil also be the focus of this paper. Generally a large number of simple sensors, such as contact switches and pressure mats, is installed throughout a house. Sensors used are cheap, unobtrusive and easy to install, criteria which are very important for the acceptance of such technology. The application infers what ADLs are taking place based on the sensor data.

When a person performs an activity inside the house, sensors will fire accordingly as each step of the activity is taking place. For example, a person performing the activity of toiletting will first enter the toilet, turn the light on inside, after some time will flush the toilet, turn off the light and leave the toilet. Depending on what sensors have been installed, a particular sensor pattern will represent these steps of the activity. Unfortunately several issues make the activity recognition task a lot less straightforward than the previous example might suggest. First, the sensor pattern generated by the activity might contain noise. Somebody could, for example, accidentally cause sensors to fire that are irrelevant to the activity. Second, activities are performed in a non-deterministic fashion. That is, people are not likely to perform the same activity using the same steps every time. Third, there can be ambiguity among sensor patterns. Meaning, different activities might generate similar sensor output.

These issues can be taken into account by using a probabilistic model. Tapia [6] describes a naive Bayes classifier which models sensor readings independently in a probabilistic fashion. This model will account for the first two issues discussed earlier. To also take the third issue into account the model should acquire some sense of time. This is done in work by Wilson [7] where a Dynamic Bayesian Network is used to perform the inference. Both approaches left a lot of parameters open for testing and in this paper we describe the most interesting ones and how they affect the accuracy of the models. Next to the parameter testing we also introduce a new idea for an observation model and test how it influences the accuracy.

The rest of this paper is organized as follows. In section 2 we introduce a notation for describing data. In section 3 we describe the models used for activity recognition. Then, in section 4, we describe the results of experiments done on a dataset provided by MIT. Finally, in section 5 we end with a conclusion.

2 Activities of daily life

Our goal is to recognize activities of daily living (ADLs) from sensor readings in a house. In order to discretize the time space we use intervals of a constant length Δt; we will state the chosen value for Δt in the experiments section. We denote a sensor reading as y_t^i, indicating whether sensor i fired at least once between time t and time $t + \Delta t$, with $y_t^i \in \{0, 1\}$. In a house with N sensors installed, we define a binary vector $\vec{y}_t = (y_t^1, y_t^2, \ldots, y_t^N)^T$. The activity associated with the sensor readings between time t and time $t + \Delta t$ is denoted as a_t.

The activity instances stored in the dataset that we will use have variable duration. With the formalization defined above every activity instance can be divided into a number of pieces, each of size Δt (fig. 1).

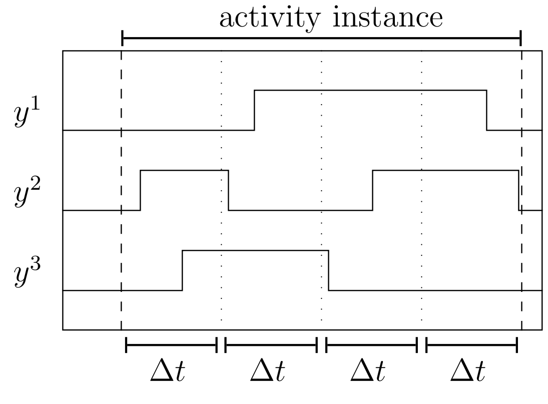

Figure 1: Showing the relation between sensor readings y_i, time intervals Δt and activity instances

3 Probabilistic models

In this section we will describe models for performing activity recognition. We will first describe a naive Bayes classifier, as presented in the work of Tapia [6]. Then we will describe a Dynamic Bayesian Network (DBN) as discussed in the thesis by Wilson [7]. Finally we will introduce a new approach which uses multiple observations.

3.1 Naive Bayes

The naive Bayes classifier assumes independence among its input nodes. In our case this means we assume independence among the probability of individual sensor readings. Although this independence assumption is not quite intuitively correct, it yields a more tractable approach. Without the independence assumption we would have to estimate too many model parameters. The resulting model is shown in figure 2.

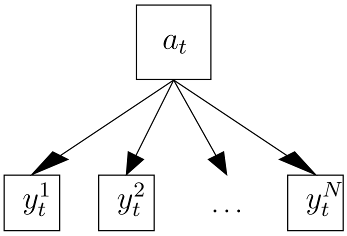

Figure 2: Naive Bayes model for Activity recognition, a_t denotes activity interval, y_t^i denotes sensor i

3.2 Dynamic Bayesian Network

One disadvantage of the naive Bayes model, described in the previous section, is that it disregards time. As described in the introduction some activities might generate ambiguous sensor patterns, that is different activities giving similar sensor readings. The naive Bayes model is unable to distinguish between these activities. A solution to this is to incorporate dynamics into the model, this means the model uses some form of time dependence [8, 7].

A Dynamic Bayesian Network (DBN) is typically described by a hidden state and its observations. The hidden state contains the information we are after, but is not directly observable (hence hidden). In our case the hidden state is the activity. Our observations are the sensor readings, as we can directly measure those from the sensors. When describing the relation between the hidden state and the observations we speak of the observation model. In our case the observation model is similar to the naive Bayes classifier. Now to incorporate a time dependence we also use a transition model. The transition model represents the relation between the hidden state at time t with the hidden state at time $t + 1$.

In terms of activity recognition we could expand our previous model to incorporate a transition model, as shown in figure 3.

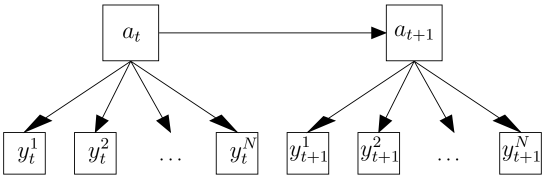

Figure 3:

3.3 Observation history

So far we have been discussing existing models used in previous publications. In this section we suggest a totally novel idea with regards to the observation model to use. The idea is to use not only sensor readings from the current timestep, but to maintain a history of the sensor activity in the previous timesteps. This history should capture some of the correlations that exist in sensor patterns over time, which allows for a more specific discrimination among sensor patterns. Although the previous state provides information about the past, it is not likely to provide as much level of information as previous sensor readings would, as our state consists only of the activity.

We define Y_t to be the observation history matrix.

$$Y_t = \begin{pmatrix} \cdots & y_{t-3}^1 & y_{t-2}^1 & y_{t-1}^1 & y_t^1 \\ \cdots & y_{t-3}^2 & y_{t-2}^2 & y_{t-1}^2 & y_t^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & y_{t-3}^N & y_{t-2}^N & y_{t-1}^N & y_t^N \end{pmatrix}$$

4 Experiments

Our goal is to recognize activities of daily living (ADLs) from sensor data. Although these ADLs can eventually be used for health monitoring purposes, in this paper we focus purely on the classification of the ADLs. To this extend we introduced some models for classification of these ADLs. To evaluate the performance of the different models described we performed various experiments using a dataset made available by MIT. The dataset consists of sensor readings annotated with activities [6]. In this section we will first describe the contents of the dataset and then discuss the various experiments and results.

4.1 Dataset

The dataset consists of sensor readings recorded in the house of a 30-year-old woman who spent free time at home. She lives alone in a one-bedroom apartment where 77 state-change sensors were installed. The sensors were left unattended, collecting data for 14 days in the apartment. This resulted in a total of 2989 sensor firings.

During the study the subject was prompted by a personal digital assistant (PDA) asking what activity she is performing at a 15 minute interval. After the experiment had taken place, a researcher looked at the sensor readings together with the subject to fill in any gaps. A total of 13 different activities were annotated

Activities in the dataset are stored as instances of variable duration. An example of an activity with short duration is toiletting, lasting an average of 7 minutes. While, for example the activity of preparing dinner lasts an average of 20 minutes. An overview of the activities, including the number of instances in the dataset, is shown in table 1.

Table 1: Overview of activities and the number of times they occur in the dataset

Activity	**Instances**
Toileting	85
Grooming	37
Dressing	24
Doing laundry	19
Bathing	18
Preparing lunch	17
Preparing a beverage	15
Preparing breakfast	14
Preparing a snack	14
Going out to work	12
Preparing dinner	8
Cleaning	8
Washing dishes	7

4.2 Experimental setup

In all the experiments we classify the three activities with the largest number of instances, namely: toileting, grooming and dressing. Each instance is divided in data segments of length Δt, after which it is split in a test and training set. The set is split by cutting it in two, rather than selecting samples randomly. This is required to keep the transitional relations in tact. In all experiments we used 50% of the set for training and the other 50% for testing.

To determine the accuracy of a particular model we classify all the segments in the testset and calculate the percentage that was correctly classified. We do 10 runs for every parameter setting we test, then we average the calculated accuracy and plot it together with the variation in an errorplot.

4.3 Performance as a function of Δt

The parameter that is probably most difficult to set intuitively is the time interval Δt. This parameter divides every activity instance into data segments of equal size. Using a very small value means sensor firing can be very precisely distinguished, but might not capture a broad enough picture. On the other hand, using a large value would provide a good summary of sensor firings within the interval, however, too large a value would cause higher ambiguity among sensor patterns.

The error plot shown in figure 4 shows the accuracy of the naive Bayes model for various values of the time interval Δt.

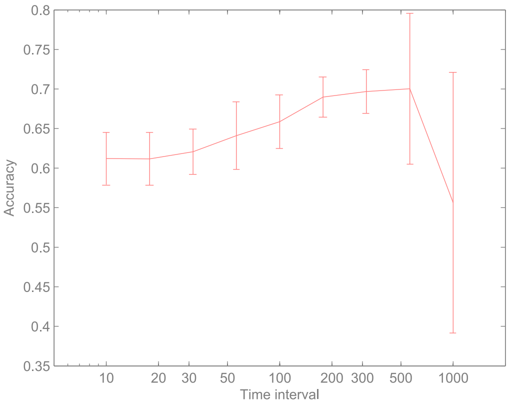

Figure 4: Plot showing the average accuracy of the naive Bayes against time interval Δt

As this figure shows the accuracy increases significantly at large values for Δt, but collapses at too large values. This effect is as expected, but at the very large values another factor comes into play. Dividing the activity instances in very large values results in very little data segments, which means less training and testing segments. This results in a larger variance of the accuracy for larger values of Δt.

4.4 Performance as a function of number of sensors

Another parameter that is not easily set intuitively is the number of sensors to use for classification. Clearly using a small amount of sensors makes discrimination among classes difficult. However, using a large number of sensors results in large conditional probability tables while the amount of information added might be minimal.

In this experiment we sorted the sensors by the number of times they fired in the entire dataset. We start by using the sensors that fired most and add those that fired less. The plot in figure 5 shows the effect of the number of sensors on the accuracy.

This plot shows two things. First, it shows that the DBN model outperforms the naive Bayes, although not significantly. Second, we see that adding more sensors does not necessarily increase the accuracy. Running this experiment on smaller dataset sizes gives a similar picture.

The insignificant performance gain by the DBN model can be explained by the average quality of transitional information in the experimental dataset. Because only a number of activities from the dataset were used for these experiments, some transitional information is lost.

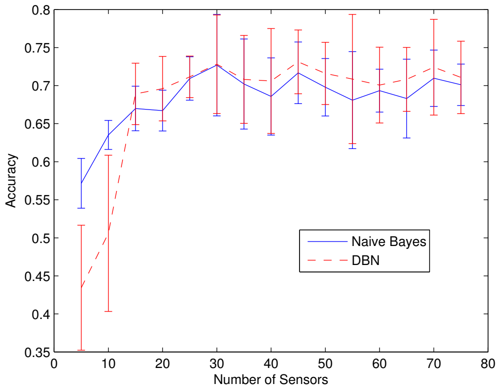

Figure 5: Plot showing the average accuracy of the naive Bayes and DBN model as a function of the number of sensors

With regard to the number of sensors. The sensors were sorted by the number of times they fired. This means that for sensors at the end of the list more training samples are required to get an accurate parameter estimation. As these samples are not available the accuracy does not improve.

4.5 Performance as a function of observation history size

The idea to use an observation history makes sense as it allows the model to incorporate more information about the past. Sensors that fired in the past can provide a context for the classification of the activity. The plot in figure 6 shows the accuracy as a function of the observation history size for two different dataset sizes.

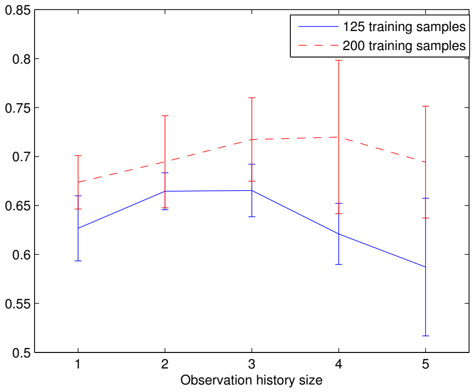

Figure 6: Plot showing the average accuracy of the naive Bayes model as a function of the size of the observation history for two different dataset sizes.

For both dataset sizes the plot clearly shows an increase in accuracy as the size of the observation history increases. However, we see a turning point in this increase as the size of the observation history becomes too large. This can be explained as follows. A larger observation history results in a larger conditional probability table and thus in more values to estimate. As the observation history size becomes too large, the estimated values in the table become too inaccurate and therefore the accuracy drops. This effect appears earlier with the smaller dataset (125 training samples) as less training samples means we have less accurate estimations in the first place. Overall we can say that the idea of the observation history works, but it requires enough training data to get accurate estimates of the parameters.

5 Conclusion

Previous work by other researchers suggested two models for activity recognition. In this paper we investigated the effect of a number of parameters on these models. Furthermore, we introduced an approach that uses an observation history. The experiments we did tell us multiple things.

By trying various time interval values we learned that this parameter greatly affects the accuracy and should therefore be chosen carefully. Although certain values clearly showed a better performance, it remains uncertain if this value holds for every type of activity in the dataset. Future work will have to show whether a golden value can be found or whether the model will have to learn the value from training.

In testing the number of sensors used, we see that adding more sensors does not necessarily increase the accuracy. Sensors placed at very rarely used locations hardly ever fire, therefore it requires very large datasets to get an accurate parameter estimation for these sensors. On the one hand the use of large datasets is discouraging, on the other hand sensors placed in rarely used locations might be very descriptive for one particular activity. We will have to investigate whether a model can be created that can be trained with less samples while maintaining the descriptive quality of such sensors.

Finally, we showed how the use of an observation history increases the accuracy. However, we also noticed how too large a history causes the accuracy to decrease, because too many parameters need to be estimated. In future work we will investigate whether the observation history can be described by less parameters, for example by some sort of distribution. This way we will need less training samples to obtain accurate estimates of the parameter values. This will in turn result in better performance.

Acknowledgements

This work is part of the Context Awareness in Residence for Elders (CARE) project. The CARE project is partly funded by the Centre for Intelligent Observation Systems (CIOS) which is a collaboration between UvA and TNO, and partly by the EU Integrated Project COGNIRON (The Cognitive Robot Companion).

References

[1] United Nations Population Division. World population prospects, the 2002 revision, highlights, February 2003.

[2] S. Ohta, H. Nakamoto, Y. Shinagawa, and T. Tanikawa. A health monitoring system for elderly people living alone. *Journal of telemedicine and telecare*, 8:151–156, 2002.

[3] R. Aipperspach, E. Cohen, and J. Canny. Modeling human behavior from simple sensors in the home. In Kenneth P. Fishkin, Bernt Schiele, Paddy Nixon, and Aaron J. Quigley, editors, *Pervasive*, volume 3968 of *Lecture Notes in Computer Science*, pages 337–348. Springer, 2006.

[4] L. Liao, D. Fox, and H. Kautz. Learning and inferring transportation routines. In *Proceedings of AAAI-04*, 2004.

[5] C. Collin, DT. Wade, S. Davies, and V. Horne. The barthel adl index: a reliability study. *Int Disability Study*, 10:61–63, 1988.

[6] Emmanuel Munguia Tapia. Activity recognition in the home setting using simple and ubiquitous sensors. Master's thesis, Massachusetts Institute of Technology, 2003.

[7] Daniel H. Wilson. *Assistive Intelligent Environments for Automatic Health Monitoring*. PhD thesis, Carnegie Mellon University, 2005.

[8] Donald J. Patterson, Dieter Fox, Henry A. Kautz, and Matthai Philipose. Fine-grained activity recognition by aggregating abstract object usage. In *ISWC*, pages 44–51. IEEE Computer Society, 2005.