Pervasive and Mobile Computing 34 (2017) 157–167

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Unsupervised visit detection in smart homes

Ahmed Nait Aicha [a,*], Gwenn Englebienne [b], Ben Kröse [b,a]

[a] Department of Computer Science, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands
[b] Department of Computer Science, University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Available online 3 June 2016

Keywords:
Ambient Assisted Living (AAL)
Markov Modulated Poisson Process (MMPP)
Smart homes
Gerontechnology

ABSTRACT

Assistive technologies for elderly often use ambient sensor systems to infer activities of daily living (ADL). In general such systems assume that only a single person (the resident) is present in the home. However, in real world environments, it is common to have visits and it is crucial to know when the resident is alone or not. We deal with this challenge by presenting a novel method that models regular activity patterns and detects visits. Our method is based on the Markov modulated Poisson process (MMPP), but is extended to allow the incorporation of multiple feature streams. The results from the experiments on nine months of sensor data collected in two apartments show that our model significantly outperforms the standard MMPP. We validate the generalisation of the model using two new data sets collected from an other sensor network.

1. Introduction

Intelligent technology that supports elderly to live independently needs information about their activities. There is an increasing interest in networks of ambient sensors, such as motion detectors and door switches, for monitoring human activities [1–3]. Often the focus is on Activities of Daily Livings (ADLs), such as sleeping, toileting and cooking. ADLs are considered important indicators for the functional health status of older adults [4].

In our group, researchers from different disciplines (information technology, machine learning and occupational therapy) work together to develop monitoring systems for older adults. The objective of the research is to use pervasive technology to support older adults to independently live longer in their own homes and their familiar and safe environment. Currently our group monitors about 20 elderly living alone using sensor networks. For a correct health assessment it is important that we are sure that the data originates from the activities of the resident only and, therefore, that we know when there are visitors. Furthermore, the type and the frequency of visits are important indicators of the social participation of elderly. One solution to identify the visitors is using RFID tags [5], but this method has some practical disadvantages in real life situations such as the ease of forgetting to scan a visit. It is also possible to use a video sensor to count persons [6], but this is hard to realise in real life situations because of privacy reasons. An unobtrusive supervised method is used in [7]. A more holistic solution, multi-person activity recognition in smart homes, has been presented [8–11]. Hence simple sensors, both wearable and ambient, can be used for the detection and monitoring of multiple persons in smart homes. Unfortunately, these methods rely on large supervised data sets which bring the difficulty of collecting the ground truth data.

Because the number of apartments we are monitoring is large and future projects will involve even more apartments, we cannot rely on supervised learning methods to build accurate individual models for visit detection. In this paper we present an unsupervised method for visitor detection. The method specifically looks at transitions between sensors and models the

* Corresponding author.
E-mail address: a.nait.aicha@gmail.com (A. Nait Aicha).

http://dx.doi.org/10.1016/j.pmcj.2016.05.003
1574-1192/

regular pattern as a normal (non visitor) pattern. An anomaly in this pattern (for example non-modelled transitions between two distant sensors) may indicate that visitors may be present. A model that has been successfully applied for the detection of anomalous events using counts as features is the Markov Modulated non-homogeneous Poisson Process (MMPP) [12], which takes into account both the periodic and non-periodic influences present in the data. This fits our situation, assuming that the resident has periodic (daily and weekly) living patterns and non-periodic patterns such as a visit on an occasional basis. However, MMPPs are univariate, and as such cannot deal with the richer, multidimensional, data sets that are common to AAL.

In [13] we showed, on a single data set, that Markov Modulated Multidimensional non-homogeneous Poisson Process (M3P2) can be used for visit detection. The M3P2 model extends the MMPP by allowing the use of a multidimensional feature stream. In this paper we present a more extensive evaluation on a more elaborate data set. We studied the effect of feature selection, the assumption on periodicity of data and the generalisation to other sensor networks.

The contribution of our research is two-fold: (a) we show that properly designed unsupervised methods can be used to detect visits with networks of a small amount of simple sensors and (b) we show that M3P2, a novel model that deals with multiple data streams, is significantly better at this task than MMPP. Moreover, it allows us to distinguish between regular and non-regular visits automatically, and to have a model of daily and weekly cycles in the person's routine. We evaluate the performance of our model on real-life sensor data. We use two data sets, that we make public, consisting of nine months of sensor data collected in the apartment of two elderly persons. The results show that our model significantly (significance level, $p < 0.05$) outperforms the standard Markov Modulated Poisson Process (MMPP). Two additional data sets are used to evaluate the generalisation of the M3P2 model.

2. Related work

The issue of dealing with multiple persons in a smart home is an important problem that recently became the subject of extensive study. In his survey [14], Teixeira describes the ability to detect the presence, count, track and identify persons using different methods and sensor types. Binary sensors in a network are able to count and localise individuals with an accuracy that depends on the number of the nodes. Using a large number of sensors in a network, Singla et al. [11] described a method that focuses on the recognition of ADLs in multi-user contexts. Using the same dense sensor network, Crandall et al. applied standard supervised classification techniques, such as Naive Bayes and HMM, to identify and track multiple smart home residents [15–17]. Phua et al. [9] noted that standard supervised techniques yield high accuracies only if (a) the number of simple sensors is large, (b) the training data is accurately labelled and (c) the activities are simple and done in habitual way. These assumptions are unrealistic in real life situations. Using a small sensor network, comparable to our situation, Petersen et al. [7], applied Support Vector Machines to detect the presence of a visitor in a smart home during a period of 6 weeks. However, supervised techniques have the weakness that they assume that the activities do not evolve over time [9]. The collection of annotated data to train supervised classifiers is difficult and involves other (invasive) sensors or a strict administration from the elderly. Unsupervised methods for the detection of abnormal behaviour in smart homes have been presented in applications like fall detection or wandering. Clustering methods such as K-means are used for the identification and prediction of abnormal behaviour of elderly dementia sufferers, but these methods are not effective in the presence of visitors or pets [5].

MMPPs [18], as unsupervised methods, are widely applied for the detection of anomalous events in various areas: to detect intrusions in a telephone network, Scott [19] introduced the non-homogeneous MMPPs, which take into account the natural cyclic nature of the variations in telephone traffic. A similar model was used by Hutchins et al. [20] to model the occupancy in a building and by Scott et al. [12] to model web traffic data. These models are all univariate, and as such cannot deal with the richer data sets that are common to AAL. Multivariate MMPPs have been described in detail in [21] for the homogeneous case, but lack the capacity of non-homogeneous MMPPs to model regular variations.

3. Sensor data

We have collected multiple data sets, in several ambient assisted living apartments, for a duration of up to more than a year. Different sensor networks were used to collect data as described in Section 6. These sensor networks consist of off-the-shelf binary sensors that measure motion, pressure on the bed, toilet flush and the opening and closing of cabinets and doors. An overview of the location of the sensors in the apartment of one resident is shown in Fig. 1. The elderly are living their routine life and are not told to modify their behaviour in any way. The location of the sensors is chosen so that all the important rooms in the apartment are covered and so that the network does not affect the elderly's daily life. For instance, the pressure sensor for the bed is installed under the mattress and sensors in the kitchen are installed above the stove, under the freezer, etc. A detailed description is provided in [13,10].

4. Features: description and extraction

The binary sensors generate a continuous stream of sensor-events. Our experience is that sensor-transitions are better than the number of sensor-events in the measurement of ADLs. Two consecutive sensor events are referred to as a sensor-transition. The number of these sensor-transitions during some time slice is likely to be smaller when there is only one