



AUTHOR(S):

TITLE:

YEAR:

Publisher citation:

OpenAIR citation:

Publisher copyright statement:

This is the _____ version of proceedings originally published by _____
and presented at _____
(ISBN _____; eISBN _____; ISSN _____).

OpenAIR takedown statement:

Section 6 of the "Repository policy for OpenAIR @ RGU" (available from <http://www.rgu.ac.uk/staff-and-current-students/library/library-policies/repository-policies>) provides guidance on the criteria under which RGU will consider withdrawing material from OpenAIR. If you believe that this item is subject to any of these criteria, or for any other reason should not be held on OpenAIR, then please contact openair-help@rgu.ac.uk with the details of the item and the nature of your complaint.

This publication is distributed under a CC _____ license.

FITsense: Employing Multi-Modal Sensors in Smart Homes to Predict Falls

Stewart Massie¹, Glenn Forbes¹, Susan Craw¹, Lucy Fraser², and Graeme Hamilton²

¹ School of Computing Science & Digital Media
Robert Gordon University, Aberdeen, UK
`{s.massie,g.forbes6,s.craw}@rgu.ac.uk`
`www.rgu.ac.uk/dmstaff/lastname-firstname`

² Albyn Housing Society Ltd, Invergordon, UK

Abstract. As people live longer, the increasing average age of the population places additional strains on our health and social services. There are widely recognised benefits to both the individual and society from supporting people to live independently for longer in their own homes. However, falls in particular have been found to be a leading cause of the elderly moving into care, and yet surprisingly preventative approaches are not in place; fall detection and rehabilitation are too late. In this paper we present FITsense, which is building a Smart Home environment to identify increased risk of falls for residents, and so allow timely interventions before falls occurs. An ambient sensor network, installed in the Smart Home, identifies low level events taking place which is analysed to generate a resident's profile of activities of daily living (ADLs). These ADL profiles are compared to both the resident's typical profile and to known "risky" profiles to allow evidence-driven intervention recommendations. Human activity recognition to identify ADLs from sensor data is a key challenge. Here we compare a windowing-based and a sequence-based event representation on four existing datasets. We find that windowing works well, giving consistent performance but may lack sufficient granularity for more complex multi-part activities.

Key words: Human Activity Recognition-Smart Homes-Sensors

1 Introduction

In this work we examine the opportunities to support assisted living environments with ambient sensors in a Smart House solution to monitor health trends in the home. A Case-Based Reasoning (CBR) approach is proposed which exploits the pattern of activities identified by the sensors to infer information about the health of the resident. In particular, the initial solution aims to predict increased risk of falls for residents of 16 Smart Homes being built near Inverness in Scotland.

The country is facing an ageing population with many people living longer. 10 million people in the UK are currently over 65 with a further increase of 5.5

million projected over the next 20 years. 3 million people are aged over 80 which is expected to double by 2030 [1]. An ageing population puts additional strains on health and social services with both a smaller proportion of working population available to support services, and with the elderly having more complex medical needs. Furthermore, with modern lifestyles, carers from within the family are less available. More people are tending to live alone as families live further apart with increased levels of relocation for work. In this changing scenario it is important that we help people with mobility or social needs to live independently for longer, and so reduce their reliance on more expensive health care solutions.

In particular, falls are an ongoing problem accounting for over 4 million bed days a year in the UK [2]. They are the most common cause of death for over 65s [3] with on average 35,848 fall-related deaths occurring annually in the EU between 2010 and 2012 [4]. One study performed in Torbay found 28% of falls proved to either be fatal or became so within 12 months, highlighting that research into preventative measures may be a more promising approach than rehabilitation after falls have occurred. Identifying preventative measures to be taken against falls could reduce morbidity, while also reducing costs and workload on health services. In addition to direct physical health concerns, falls have a lasting psychological effect which can reduce a person's confidence in their independent abilities, leading to an increase in sedentary behaviour and depression [5]. Effective methodologies for anticipating falls would be invaluable and the benefits associated with prevention appear to outweigh those of rehabilitation.

Recent developments in a number of technologies (sensors, Internet of Things, Cloud Computing, and increased computational power), along with reduced costs have resulted in substantial interest in the development of Smart Homes for automation, security and to a lesser extent health. Smart Homes offer a ubiquitous computing solution, in which a house utilises many sensors to deliver a safer environment. The core design of the common devices (e.g. IR, magnets, temperature) have remained unchanged over the last decade although size, cost and efficiency have all improved. Newer technologies are also beginning to be relevant, such as Wi-Fi and radar. Ambient sensors are practical for continuous monitoring in health application; lacking the overhead associated with wearables, or privacy concerns with video.

In this paper, we explore the use of everyday, low-cost ambient sensors installed in new-build Smart Homes with the aim of supporting tenants to live independently for longer. Specifically, we identify and discuss the main challenges in designing and deploying a real-world health monitoring system that senses and predicts the level of risk of falling attributed to Smart Home residents. Data is captured by a range of sensors installed in specially-designed, technology-enabled "FitHomes". Targeting specific activities identified as precursors to falls, the system analyses data derived from these sensors to identify patterns of activity, and changes in these patterns which could be linked to an increased risk of falling. It is hoped that evidence-based alerts will enable families and agencies to intervene with preventative measures before incidents occur. An outline solution is developed and initial experiments are undertaken to evalu-

ate alternative approaches to classifying activity with low level, raw data inputs from multiple multi-modal sensors. The key contributions of the work are:

- to outline a novel CBR solution for identifying the risk of falls for residents in Smart Homes; and
- to evaluate alternative representations for activity recognition from the low-level, data inputs delivered from sensors.

The remainder of this paper describes our approach to employing ambient, non-intrusive sensors for monitoring and predicting risk of falls in purpose-built assisted homes, and presents experiments that evaluate alternative representations for activity recognition. In Section 2 we review existing research on the use of Smart Homes for health monitoring and risk of fall prediction. Section 3 discusses, in more detail, the specific scenario being faced in this real world development along with the associated challenges that we plan to address. Our proposed 2-part solution, which first employs Machine Learning (ML) to generate a more abstract case-structure on which we can then build an effective CBR system is outlined in Section 4. In Section 5 we introduce four datasets that we use to evaluate alternative feature representations for our initial activity recognition task. Finally, we draw our conclusions in Section 6.

2 Related Work

Activities of Daily Living (ADLs) are events in daily life which would be considered intrinsic to a person’s ability to live independently. Typical ADLs include being able to dress oneself, get out of bed, and feed oneself. Katz [6] originally proposed the term along with a scale for rating a person’s independent ability using their performance in simple ADLs. He suggests that there may be a procedural decline in ADL capability. While this was not proven, the concept of losing ADLs as we age has influenced future research in the field by identifying that specific ADLs are more indicative of reduced capability than others. Observing variances in ADL performance, such as missing a key step in an activity or performing steps out of order, can aid the identification of degenerative mental and physical capability which in turn may contribute to an increased risk of falling.

Physiological expressions, such as movement, can also be used to identify an increased risk of falling. Vestergaard [7] performed a study in which a relationship between performance in a 400-meter walk test and subsequent mortality in older adults was observed. This test (and other shorter variants) is usually performed in laboratory or hospital conditions, in which a physician would also be able to consider the patients condition and other metrics from the test. These include but are not limited to, whether or not a break was taken, variation in lap times and existing health conditions. However, lab-based testing is time consuming, costly and impractical for many patients, especially those with mobility issues. In addition, some studies have been able to identify risk of falling, and other potential health issues, in the elderly using gait velocity alone [8–10]. So

while gait and other expressions of movement are indicative of many underlying conditions, measuring all aspects of gait such as swing and stride length requires specialist equipment, e.g. vision-based sensors. Gait velocity, however, can be measured using simpler equipment and still provides excellent insight into subject movement. For instance, Rana [11] performed a study in which gait velocity was estimated using simple infra-red motion sensors. We plan to adopt this approach and, while lab-based testing can provide higher accuracy, we hope accessible in-home testing can contribute to early detection of health problems.

Housing installations with ubiquitous simple sensors offer an opportunity to provide continuous behavioural and physiological monitoring of residents. These simple sensors can range from binary magnetic switches [12], to IoT-monitored motion sensors [13], all of which can provide insight into behavioural and physiological expressions. ADLs can then be reconstructed and modelled by identifying temporal patterns in these sensor outputs.

Modelling and classifying activities from sensor data typically involves applying ML techniques best adapted for pattern recognition. Several manually annotated datasets taken from Smart Home installations have been produced for the purpose of activity recognition [12, 14, 15]. Tapia made use of an extended Naive Bayes classifier to identify activities in their labelled dataset, whereas Kasteren made use of Hidden Markov Models (HMM) and Conditional Random Fields (CRF). All these techniques have demonstrable strengths in activity recognition, however the use of generative methods, such as HMMs and CRFs allow for the use of sequential data to train a model based on successive activities [16]. We use these existing datasets in our experiments to explore the effectiveness of alternative modelling and ML approaches.

3 FitHomes & Predicting Falls

FitHomes is an initiative, lead by Albyn Housing Society Ltd (AHS) in partnership with Carbon Dynamic, that aims to support independent living with the supply of custom-built Smart Homes fitted with integrated non-invasive sensors. 16 houses are being built and near completion at Alness near Invergordon. These houses are part of a development cycle with a further 32 FitHomes, funded by the Inverness City Deal, planned to be built within the Inverness area within 3 years. FITsense is a one year Data Lab³ funded project that aims to exploit the sensor data to develop a prototype fall prediction system for these FitHomes.

3.1 Sensors

One of the first considerations in designing a Smart Home focused on health monitoring is the choice of what type and mix of sensors to use in order to provide a cost-effective solution that is also acceptable to residents. AHS have conducted initial research and it was clear that their tenants wanted an unobtrusive system that supported them in their homes, but did not take over. Both

³ The Data Lab, Scotland. <https://www.thedatalab.com/>

video and wearables were considered too intrusive for continuous in-home use; video due to privacy issues and wearables due to the ongoing overhead associated with 24 hour operation. As a result, the focus in this project has been on simple everyday sensors, many of which are already widely used in security and automation applications. FITsense is an applied project and with this approach we can establish the limits of existing technology now, rather than developing new solutions for the future. A plug and play design will be adopted such that new sensors can be easily integrated as new technologies become main stream. A further benefit is provision of a low cost solution from the hardware perspective but with additional challenges for the data analysis.

FITsense aims to identify increased risk of falls and so a key focus for monitoring is to identify activity levels, patterns and speeds. However, the monitoring can go beyond just movement to consider other factors that have been shown to be related to falls, including dehydration, tiredness and mental health. Gaining information on these additional factors requires monitoring to also capture data on more general activities such as eating & drinking behaviours, sleep patterns, and toileting & grooming habits. With these criteria in mind a range of sensors have been selected for the FitHomes, that include:

- IR motion sensors that capture movement in each room;
- contact sensors to capture room, cupboard, and fridge door opening/closing;
- pressure sensors that identify use of the bed and chairs;
- IR beam break sensor to identify gait speed;
- electricity smart meters to identify power usage pattern;
- float sensors identifying toilet flushing;
- humidity sensors to identify shower use; and
- temperature sensors integrated with the humidity sensors.

Figure 1 shows typical sensors being used, including from left to right: a motion sensor; a presence sensor, being considered to identify presence of carers; a contact sensor; and electricity usage sensor, being considered for specific electrical items.



Fig. 1. Example of the sensors used in FITsense

Most of the sensors chosen have a binary output that simply activate when the event they are monitoring takes place e.g. a door opening; however others

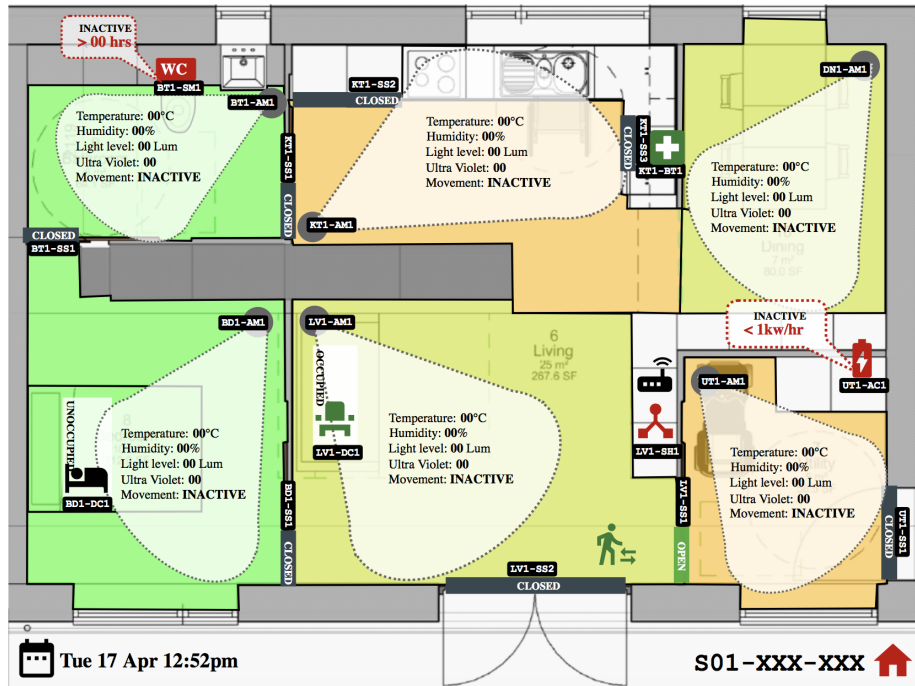


Fig. 2. Annotated floorplan of a FitHome

output continuous readings provided at fixed polling rates. The data fusion task across multiple sensors with different output modes becomes one of the main challenges in employing large numbers of sensors.

3.2 Smart Homes

The FitHomes are factory manufactured and supplied on-site ready to connect to site services, providing a cost effective build method for multiple properties. Sensor installation currently takes place on-site. Positioning and orientation of the sensors is important to give as much information as possible but also to consider building constraints, for example to give access to an electricity connection and remove reliance on battery usage. Figure 2 gives a plan layout view of the FitHomes with positioning identified for many of the sensors, including 6 motion sensors (one in each room), contact & pressure sensors, humidity & temperature sensors, electricity meter, and a float sensor in the toilet.

A Samsung SmartThings hub is used as the data centre to collect output from the sensors via ZigBee and pass the data on over the internet to cloud storage that allows API access for data analysis at a later date as required. The key challenge in employing multiple ambient sensors of varying types is to transform the low-level largely event-driven individual sensor activations (e.g. movement in kitchen) into meaningful activities on which to reason (e.g. food preparation).

4 Case-Based Approach

The elements of the CBR solution are first to identify patterns in the data that allow us to create representations from the low-level, raw sensor data that capture the residents activities and behaviours of daily living, e.g. sleeping, dressing, showering, cooking etc., and then to assemble these activities into personalised daily and weekly profiles. The second stage is the analysis of these activity profiles to enable both the identification of changing trends in the residents activities over time and to make comparisons with data collected from other similar residents. Changes in the Smart Home residents own activity patterns over time can then be used to detect deterioration in health linked to falls, while comparisons with the patterns of other Smart Home residents can provide benchmark measures of health. The data thus supports evidence-driven intervention tailored to the resident and their specific circumstances.

4.1 Classifying ADLs

Human Activity Recognition (HAR) to identify ADLs is challenging in Smart Home scenarios because large volumes of data is generated from multi-modal sensors in real time making patterns associated with specific activities difficult to identify. Simple sensors (e.g. door open/closed sensors) are binary and record events, while more complex sensors (e.g. electricity consumption meters) poll data at fixed intervals to produce single or multi-dimensional time series outputs.

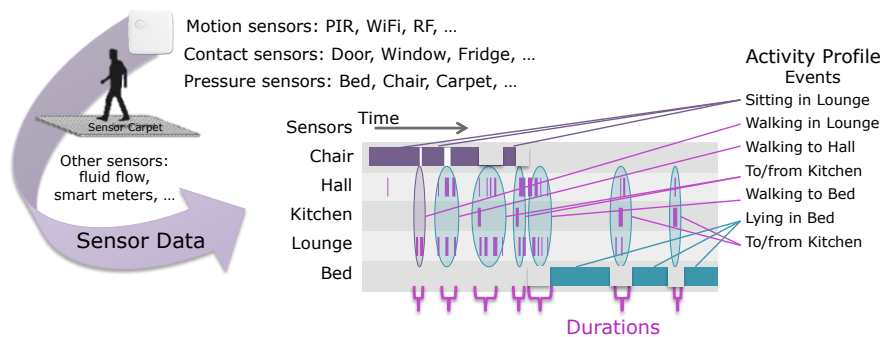


Fig. 3. Identifying activities from sensor activations

Figure 3 shows a diagram with examples of sensor activations for motion sensors in a hall, kitchen and lounge together with pressure sensors on the chair and bed. Simple events can be inferred from this data to generate activities. A mix of approaches will be adopted to identify activities and to then generate the residents daily activity profile. For the simple activities shown (e.g. time sitting, time in bed, number of toilet visits, number of room transitions) only one or two

sensor activations are required to identify the activity; a rule-based approach with simple human generated rules is sufficient to identify the activity. Where effective this approach will be adopted.

However, more complex activities can only be recognised by the interaction of several sensors e.g. food preparation, showering, grooming, disturbed sleep. For these more complex activities a ML approach will be adopted. HAR typically employs a windowing approach to create a single aggregated vector representation on which ML (e.g. kNN, Support Vector Machines or Naive Bayes) can be applied for classification. These approaches can work well but are perhaps less able to handle the data fusion scenarios from Smart Homes because of difficulties in selecting appropriate time windows for different activities; and due to the loss of information when the sequence of events is not maintained, by aggregating within a window. In this paper we investigate using a sequence-based representation, in which the events are placed in order based on their time stamp.

4.2 Reasoning with ADLs

Identifying ADLs in themselves does not give an indication of health. However, it has been shown that one of the best ways to evaluate the health status of older adults is through functional assessment [17]; ADLs are lost as we age and in FITsense the plan is to monitor changes in ADL activity as an indicator of deteriorating health and increased risk of falls. To do this a CBR approach is adopted. With CBR, new problems are solved by retrieving similar, previously solved problems and reusing their solutions. In our scenario, a set of ADL templates (together with contextual information) will be used as the problem representation to retrieve similar profiles from a case base of existing profiles. Solutions will identify interventions, where required, and their previous outcomes.

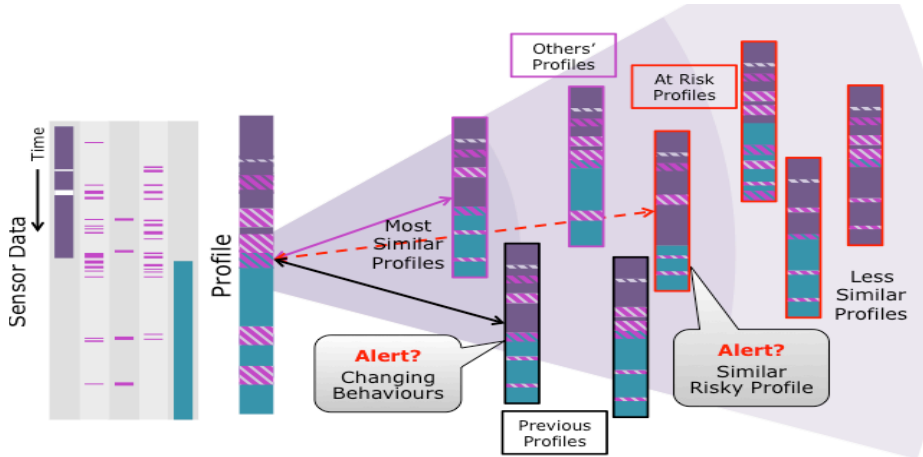


Fig. 4. CBR Approach to Identifying 'Risky' Behaviours

Figure 4 presents an overview of our approach. Low-level, time-stamped events identified by the sensors are transformed into a daily user profile. The profiles are a set of ADLs with mixed data types: some ADLs are binary, e.g. disturbed sleep; some ADLs are counts, e.g. number of room transitions or stand up from seat count; some are cumulative daily time spans, e.g. time sitting, or time in bed; while others are numeric, e.g. average gait speed. Whatever the data type a similarity measure is associated with each ADL so that comparison can be made between them. A set of daily ADL profiles for a resident can then be compared with those in the case base, on the right of Figure 4. Retrieval of similar profiles labelled as at risk identifies the need to recommend intervention, and falling similarity with the user’s own previous profiles identifies changing behaviours. Importance in determining similarity for FITsense is given to ADLs known to correlate with falls. For other health conditions the similarity knowledge could be refined to reflect specific conditions e.g. gait for falls, erratic behaviour for Dementia, general physical activity level for obesity, etc.

A key challenge is to identify risky or deteriorating behaviour. Labelled positive cases (identifying a fall is likely) are rare because people don’t fall that often. The initial approach is to generate template solutions with guidance from health care professionals. Then, as real data becomes available, we can learn/refine/supplement these hand-crafted templates with the addition of real experiences as they occur in the data generated both by the user and by others.

5 Evaluation

The initial task is to assess our effectiveness at classifying ADLs from raw sensor data. We do not yet have live data being generated by tenants from the FitHomes, so for this evaluation we use existing datasets. The aim of the evaluation is to compare the performance of different ML algorithms when applied to the window-based and sequence-based representations.

5.1 Datasets

Four publicly available datasets are used in our experiments: CASAS⁴ (adlnormal), Van Kasteren⁵ (kasteren) and two from the Massachusetts Institute of Technology⁶ (tapia1/2). These datasets share similar properties to that expected from the FitHomes with a focus on activity recognition using simple sensors in Smart Home installations. They all capture binary sensor activation data from the homes and have been labelled with class information, i.e. the ADL identified during the specified time period. The activities are of varying length.

Table 1 gives an overview of the structure of the datasets. These are relatively small datasets with between 120 and 295 instances, reflecting the high cost of

⁴ <http://casas.wsu.edu/datasets/adlnormal.zip>

⁵ <https://sites.google.com/site/tim0306/kasterenDataset.zip>

⁶ http://courses.media.mit.edu/2004fall/mas622j/04.projects/home/thesis_data_txt.zip

Table 1. Overview of the datasets used.

Dataset	Classes	Attributes	Instances
adlnormal	5	39	120
kasteren	7	14	242
tapia1	22	76	295
tapia2	24	70	208

manual labelling. The number of attributes varies between 14 and 76 reflecting differences in the number of sensors present in different installation set ups. Likewise, there are differences in the number of activities being monitored (i.e. classes) depending on the focus of the particular study; tapia in particular has a large number of different activity labels, some of which would not be relevant for predicting falls. Some activities are more popular than others and as a result most datasets do not have balanced class distributions. In table 2, the activity classes present in each dataset are shown, along with a count of the number of times the activity is recorded in the dataset.

The average sequence length of activities identified in the datasets varies between 4.7 in kasteren and 34.4 in adlnormal, as can be seen in table 3. The datasets feature complete representations of sensor activations, including timestamps and durations, which allows us to build both a window-based and sequence-based representation.

The window-based representation is a fixed-length vector which does not change with varying activity lengths. If we count the number of sensors in the installation there will be one problem-side attribute for each sensor. The attribute value being a count of the number of times the sensor is activated during an activity timespan. The solution is a single class label, namely the labelled activity.

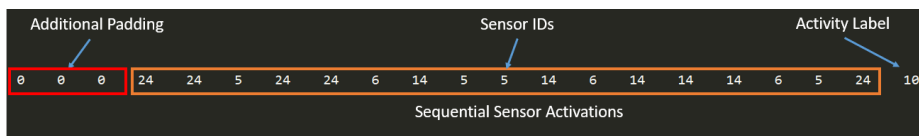


Fig. 5. Example Sequence-based representation for a shortened kasteren dataset.

The sequence-based representation captures temporal relationships between attributes. The intuition is that this additional information will aid activity classification performance, with the ordered sequences of sensor activations allowing more detailed understanding of activities and the underlying sensor network in

the installation. A fixed length representation set to the length of the longest activity sequence in the dataset is used. Hence, as shown in figure 5, each problem-side attribute in a case is a sensor activation identified by its unique id, or a null padding value. As with the window-based representation, the solution is a single class label, identifying the activity. The longest activity length in the shortened “kasteren” example, is 20, and a sequence of 17 sensor activations is recorded in this activity. Hence, the first 3 attributes are null. As the maximum activity length increases, the number of null attribute values in shorter activities (which form the majority of datasets) will increase.

5.2 Experiment Set-Up

Popular ML algorithms that delivered good performance on these datasets were selected from the default Weka library to run on the window-based representation of each dataset [18]. These were compared to CRFs run on both the window-based and sequential-based representation. CRFs were selected for use with sequence-based representation as they can train based on the probability of previous sequences reoccurring. By modelling state-to-state dependencies the context of a sequence within a meta-sequence can be considered during training. Weka does not natively support learning with CRFs, and so for CRF learning,

Table 2. Details on the distribution of activities across datasets.

Dataset	Activities (in order of expression)
adlnormal	24 x Phone.Call, 24 x Wash_hands, 24 x Cook, 24 x Eat, 24 x Clean
kasteren	34 x Leave_House, 113 x Use_Toilet, 23 x Take_Shower, 23 x Go_to_Bed, 20 x Prepare_Breakfast, 10 x Prepare_Dinner, 19 x Get_Drink
tapia1	1 x Going_out_for_entertainment, 15 x Preparing_a_snack, 19 x Doing_laundry, 4 x Dressing, 1 x Washing_hands, 8 x Washing_dishes, 3 x Watching_TV, 14 x Preparing_breakfast, 12 x Going_out_to_work, 2 x Putting_away_dishes, 37 x Grooming, 9 x Cleaning, 2 x Putting_away_groceries, 18 x Bathing, 8 x Preparing_dinner, 17 x Preparing_lunch, 1 x Other, 2 x Putting_away_laundry, 2 x Going_out_for_shopping, 1 x Lawnwork, 15 x Preparing_a_beverage, 84 x Toileting
tapia2	4 x Talking_on_telephone, 1 x Lawnwork, 3 x Cleaning, 5 x Dressing, 16 x Preparing a snack, 2 x Home_education, 17 x Listening_to_music, 2 x Grooming, 37 x Toileting, 15 x Watching_TV, 2 x Other, 14 x Taking_medication, 13 x Preparing_breakfast, 5 x Working_at_computer, 3 x Going_out_for_shopping, 20 x Preparing_lunch, 20 x Washing_dishes, 1 x Preparing_a_beverage, 1 x Putting_away_groceries, 1 x Going_out_for_entertainment, 3 x Bathing, 3 x Putting_away_dishes, 1 x Putting_away_laundry, 14 x Preparing_dinner

Table 3. Average and maximum length of activities.

Dataset	Sequence Length		Temporal Length	
	Avg (cnt)	Max (cnt)	Avg Time (sec)	Max Time (sec)
adlnormal	34.4	127	203.3	658
kasteren	4.7	92	8588.4	38193
tapia1	9.4	156	732.5	8132
tapia2	9.4	184	1824.5	14936

the CRF++ toolkit was used. Both tools make use of different data formats, so each dataset was converted to ARFF (for use in Weka), and CSV (for use with CRF++).

- Bayes Network: Using the BayesNet bayes classifier.
- k-NN: Using the IBk lazy classifier (with k=3).
- SVM: Using the SMO function classifier.
- J48: Using the J48 tree classifier.
- CRF-Win: Using CRFs on the window-based representation.
- CRF-Seq: Using CRFs on sequenced-based representation.

Given the limited data available, Leave-One-Out cross validation was applied on all experiments. In addition to recording average accuracy results, confusion matrices were plotted for each dataset and ML algorithm combination using Matplotlib⁷ (for CRFs), and Weka (for other algorithms).

Table 4. Experiment results (in accuracy %).

Dataset	BayesNet	k-NN	SVM	J48	CRF-Win	CRF-Seq
adlnormal	98.3	91.6	92.5	92.5	95.0	96.7
kasteren	92.6	94.2	81.0	93.4	80.6	93.0
tapia1	50.8	54.2	56.3	54.2	61.0	55.6
tapia2	28.3	34.6	35.1	47.1	47.1	42.0

⁷ <https://matplotlib.org/>

5.3 Results and Discussion

The performance of BayesNet, k-NN, SVM, J48 and CRFs when used with Windowed data, and CRFs when used with Sequenced data are compared. The results can be seen in table 4 with the highest accuracy achieved on each dataset in bold.

On the window-based representation, high accuracies, generally in excess of 90%, are achieved on adlnormal and kasteren compared to highs of 61% and 47% on tapia1 and 2 respectively. The differences reflect that both tapia datasets present a much harder classification task with over 20 fine grained activities, many of which are hard to distinguish even with over 70 sensors. adlnormal and kasteren have fewer activities being identified (5 and 7 respectively) and fewer sensors (39 and 14 respectively). kasteren in particular is more in line with the type of activities and sensor network we plan for FITsense.

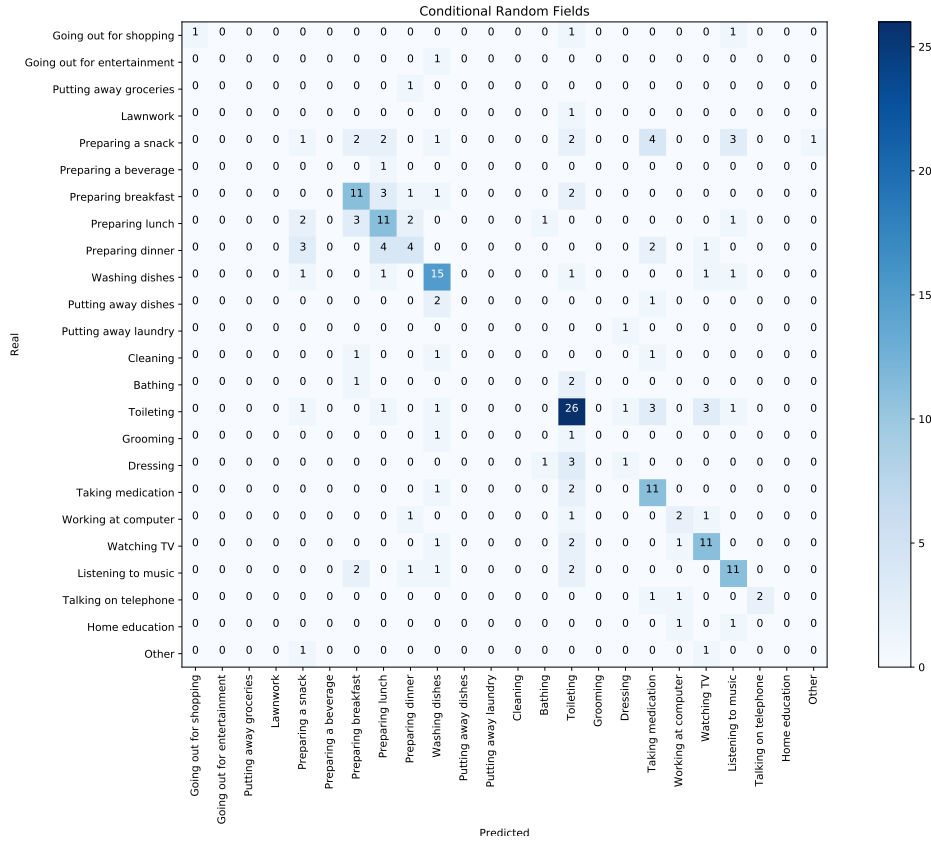


Fig. 6. Confusion matrix for CRFs on tapia1 windowed data.

With the algorithms applied to the window-based representation, there is not a clear winner. BayesNet, k-NN and J48 all provide good performance on the simpler datasets (adlnormal and kasteren); k-NN gives highest accuracy on kasteren which having the fewest sensors and shortest activity sequences is likely to have few noisy attributes. BayesNet gives highest accuracy on adlnormal which is distinguished by having long sensor sequences associated with activities. CRF-Win gives highest accuracies on the more complex tapia datasets, which seems to indicate that the relationship between sensor activations becomes more important for distinguishing similar activities from each other.

On the sequence-based dataset representation, CRF-Seq outperformed CRF-Win on the simpler datasets, although it was beaten by BayesNet on adlnormal and k-NN on kasteren. On the tapia datasets CRF-Seq did not perform as well as CRF-Win, although its performance was competitive with the other algorithms. These results are slightly surprising as we anticipated that knowledge of sensor activation sequence would improve classification.

Figure 6 shows an example confusion matrix (CRF-Win on tapia1). This view of the results identifies specific activities that get miss-classified and interestingly the activity they get miss-classified as. There are errors that might be expected, for example confusing preparing breakfast with preparing lunch and vice-versa. Generally, activities associated with specific sensors such as Taking Medication or Toileting tend to classify better than activities performed in shared spaces with several sensors used across many activities e.g. preparing dinner.

6 Conclusions

In this paper we have presented a Smart Home approach to predicting increased risk of falls for residents in 16 assisted living houses being built in Scotland. Simple ambient sensors are employed to monitor activities of daily living. We propose a two stage approach in which activities are first classified based on low level sensor data inputs. Daily/weekly activity profiles are then assembled for each resident and compared to their own past data and known risky profiles.

Overall, the initial experiment results on activity classification are promising and we can expect accurate identification of activities in FITsense, providing that the classes are not too fine-grained. It appears that the window-based representation is sufficient for effective classification, although the results are not clear and additional comparisons will be made when data becomes available from FitHomes. It may be that a hybrid approach is required with the assumption that attributes are independent being an effective simplification for simple activities; but for more complex activities, methods that take advantage of attribute interaction and event sequences may be more effective.

Acknowledgements This work was part funded by The Scottish Funding Council via The Data Lab innovation centre. Thanks also to Matt Stevenson at Carbon Dynamic and Angus Watson at NHS Highland, Inverness for their support of the FITsense project.

References

1. Population Division, United Nations Department of Economic and Social Affairs: World population ageing (2015)
2. NHS Improvement: The incidence and costs of inpatient falls in hospitals (2017)
3. Tian, Y., Thompson, J., Buck, D., Sonola, L.: Exploring the system-wide costs of falls in older people in Torbay. London: The King's Fund (2013) 1–12
4. Turner, S., Kisser, R., Rogmans, W.: Falls among older adults in the EU-28 : Key facts from the available statistics. Technical report, Swansea University, EuroSafe (2015)
5. Arfken, C.L., Lach, H.W., MSN, Birge, S.J., Miller, J.P.: The prevalence and correlates of fear of falling in elderly persons living in the community. *American Journal of Public Health* **84**(4) (1994) 565–570
6. Katz, S., Ford, A.B., Moskowitz, R.W., Jackson, B.a., Jaffe, M.W.: Studies of illness in the aged. the index of ADL: A standardized measure of biological and psychosocial function. *Journal of the American Medical Association* **185** (1963) 914–919
7. Vestergaard, S., Patel, K.V., Bandinelli, S., Ferrucci, L., Guralnik, J.M.: Characteristics of 400-meter walk test performance and subsequent mortality in older adults. *Rejuvenation Research* **12**(3) (2009) 177–184
8. Stone, E.E., Skubic, M.: Unobtrusive, continuous, in-home gait measurement using the Microsoft Kinect. *IEEE Transactions on Biomedical Engineering* **60**(10) (2013) 2925–2932
9. Jiang, S., Zhang, B., Wei, D.: The elderly fall risk assessment and prediction based on gait analysis. 2011 IEEE 11th International Conference on Computer and Information Technology (2011) 176–180
10. Montero-Odasso, M., Schapira, M., Soriano, E.R., Varela, M., Kaplan, R., Camera, L.A., Mayorga, L.M.: Gait velocity as a single predictor of adverse events in healthy seniors aged 75 years and older. *Journal of Gerontology* **60A** (2005) 1304–1309
11. Rana, R., Austin, D., Jacobs, P.G., Karunanithi, M., Kaye, J.: Gait velocity estimation using time-interleaved between consecutive passive IR sensor activations. *IEEE Sensors Journal* **16**(16) (2016) 6351–6358
12. Tapia, E.M., Intille, S.S., Larson, K.: Activity recognition in the home using simple and ubiquitous sensors. *International Conference on Pervasive Computing* (2004) 158–175
13. Suryadevara, N.K., Mukhopadhyay, S.C.: Wireless sensor network based home monitoring system for wellness determination of elderly. *IEEE Sensors Journal* **12**(6) (2012) 1965–1972
14. Van Kasteren, T., Noulas, A., Englebienne, G., Kröse, B.: Accurate activity recognition in a home setting. In: *Proceedings of the 10th international conference on Ubiquitous computing*, ACM (2008) 1–9
15. Cook, D.J., Schmitter-Edgecombe, M.: Assessing the quality of activities in a smart environment. *Methods of information in medicine* **48**(5) (2009) 480
16. Kim, E., Helal, S., Cook, D.: Human activity recognition and pattern discovery. *IEEE Pervasive Computing* **9**(1) (2010) 1–14
17. Cook, D.J., Schmitter-edgecombe, M., Cook, D.J., Schmitter-edgecombe, M., Dawadi, P.: Analyzing Activity Behavior and Movement in a Naturalistic Environment Using Smart Home Techniques. **19**(August) (2015) 1882–1892
18. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. *SIGKDD Explorations* **11**(1) (2009) 10–18