Monitoring Quality of Life Indicators at Home from Sparse and Low-Cost Sensor Data

Dympna O'Sullivan[1] Rilwan Basaru[2] Simone Stumpf[2] and Neil Maiden[2]

[1] ASCNet Reseach Group, School of Computer Science, TU Dublin, Dublin, Ireland
[2] City, University of London, London, UK
dympna.osullivan@tudublin.ie, remilekun.basaru.1@city.ac.uk,
simone.stumpf.1@city.ac.uk, neil.maiden.1@city.ac.uk

Abstract. Supporting older people, many of whom live with chronic conditions, cognitive and physical impairments to live independently at home is of increasing importance due to ageing demographics. To aid independent living at home, much effort is being directed at reliably detecting activities from sensor data to monitor people's quality of life or to enhance self-management of their own health. Current efforts typically leverage large numbers of sensors to overcome challenges in the accurate detection of activities. In this work, we report on the results of machine learning models based on data collected with a small number of low-cost, off-the-shelf passive sensors that were retrofitted in real homes, some with more than a single occupant. Models were developed from sensor data to recognize activities of daily living, such as eating and dressing as well as meaningful activities, such as reading a book and socializing. We found that a Recurrent Neural Network was most accurate in recognizing activities. However, many activities remain difficult to detect, in particular meaningful activities, which are characterized by high levels of individual personalization.

Keywords: Activity Recognition. Sensors, Machine Learning, Independent Living.

1 Introduction

An understanding of a person's activities and the extent to which activities are being achieved or not can be used to improve self-monitoring and self-care at home, including their quality of life [1]. There are two main challenges to implementing activity recognition at home. First, there is the challenge of retrofitting residences with sensors. Typically, smart home solutions have hundreds of sensors with the aim of collecting data to recognize a range of different activities. The cost and complexity of such installations often prevents their take-up in real-world applications. Second, even with large amounts of sensor data, there are challenges to developing machine learning models for activity recognition. These include noisy sensor data, large numbers of false positives, difficulty training activity recognition algorithms on data collected in homes with a different layout and the multiple-occupancy problem. Research has typically focused on detecting activities of daily living (ADLs), which are tasks that people undertake routinely in their everyday lives, for example, eating, sleeping and grooming [2]. There is less research on monitoring meaningful activities, i.e. physical, social, and leisure activities that provide the patient with "emotional, creative, intellectual, and spiritual stimulation" [3], as an important indicator of quality of life.

To address these challenges, we developed and investigated a toolkit composed of a small number of low-cost off-the-shelf passive sensors, typically up to 10, which were retrofitted into real, sometimes multiple-occupancy homes to detect both ADLs and meaningful activities. To measure meaningful activities we employed beacons sensors, small devices that broadcast packets of data over Bluetooth, and are placed on objects in the home that residents interact with frequently. This allows capturing more minute details on a resident's activities. For example, in [4] the authors demonstrated how accelerometer data captured from beacon sensors could detect not only the presence of residents interacting with the objects, but also the way the objects were moved (e.g. placing a knife on the table vs. using the knife to cut food in its preparation). Niu et al. [5] propose a similar approach using BLE (Bluetooth Low Energy) beacons to measure movement and achieved an accuracy of 70% averages across seven activities. There are challenges with the use of such small sensors affixed to objects. In addition to the size constraints of the sensors themselves, energy consumption can be a problem, as analyzing accelerometer data requires a high transmission rate in order to capture the movements effectively with machine learning techniques. However, accuracy of detection using beacons is relatively high and they are suited to multiple occupancy environments as they can provide specific location accuracy allowing to identify who is interacting with a device. We collected data from five users in five different homes, each over a period of one week. Three of the homes were multi occupancy homes. We used this data to train five machine learning algorithms and evaluated their accuracy in recognizing ADLs and meaningful activities. In this paper we present the methods employed in our study, including how we collected data and ground truth labels from human participants, and how we trained and evaluated the machine learning models. We present our results, focusing on the overall accuracy of the machine learning models as well as the accuracy in recognizing individual activities. We conclude by discussing the potential implications of our work, as well as directions for future research.

2 Methods

2.1 Data Collection

We recruited 5 participants (3 males, 2 females), all aged 18 and above, without any cognitive or physical impairments to take part in a pilot study. We received ethics approval prior to commencing the study and obtained informed consent from all participants. Participants were able to choose a set of activities (Table 2), agreed between the researcher and each participant, with a mixture of ADLs and meaningful activities. Participants carried out a set of activities over the course of one week in their own homes. To detect interaction with objects around the home six main sensor types were used - motion, door, power, ambient (temperature and humidity), pressure and beacon sensors. The motion, door, and pressure sensors are binary sensors that can detect motion in an environment, for example, opening of a door and the application of a pressure on a surface such as a bed respectively. The temperature, humidity, and power sensors are continuous sensors that detect changes in temperature, humidity and power surges. Finally, the beacon sensor is a binary sensor that detects the disturbance of any object or surface it is attached to. For example, they were attached to bookmarks and the remote control for the TV. Based on the selected set of activities, the appropriate set of sensors was provided and installed by the researcher, who noted down the location on a rough sketch of the floor plan of the participant's home. During the study, data collected from the sensors was stored in a database on a Raspberry Pi. Because of the time-dependent nature of the data being stored, we used InfluxDB [6], an open-source time series database framework, optimized for fast, storage and retrieval of time series data. The motion, door and ambient sensors were from the same manufacturer, Xiaomi. The pressure and power sensors interfaced with the Raspberry Pi, using a z-wave communication protocol. We used the Home Assistant open-source framework [7] as a service for asynchronously listening for sensor readings and updating the InfluxDB database. A typical kit was composed of 25 sensors and cost on average £412 including the hub components.

Data collection took place over February and March 2019. Participants recorded a log of activities using a journaling app called ATracker [8] on an Android tablet to record the start and end time of activities as they were completed. These logs were used as ground truth labels for the sensor data. We collected data for 14 activities. There was high variation in the frequency and the duration of completing each task. Sleeping, was recorded the most frequently (11 times) and recorded the most (95.34 hours), followed by Going Out (10 times, 30.44 hours). Food preparation was the most frequently recorded activity (24 times) but on average took much less time (0.29 hours). On the other end of activity frequency and duration were Vacuuming (3 times), Nail Care (2 times), Grooming (2 times), Laundry (3 times) and Playing Board Games (1 times); these activities only happened infrequently and also recorded the least amount of time overall. To reduce bias in the prediction models (such that models developed would not be biased towards classes with higher frequency or duration), we removed infrequent activities where there are not enough training and testing data (playing board games) and we applied a class weight to "boost" activities with lower frequencies.

2.2 Model Development and Evaluation

We used the ScikitLearn Python machine learning library to implement SVM, Naïve Bayes, Logistic Regression, and Perceptron models. The Naïve Bayes was multinomial and trained with an adaptive smoothing parameter (alpha) of 0.01. The SVM model was trained with 5 maximum epochs. The Perceptron model was trained with a stopping criterion of 1e-3. The RNN was implemented with the TensorFlow framework and trained with a learning rate of 0.001, weight decay of 0.005 and under 2 epochs. Data was split into training and validation sets by a 75:25 ratio. Model performance was measured by comparing predicted activities with ground truth gathered via the Atracker app. We calculated accuracy for each algorithm as a ratio of all correctly labelled data point to all test data points. To take into consideration the imbalanced nature of the data, we also computed micro and macro averages for precision, recall, and F1-score.

3 Results

3.1 Model Accuracy

RNN achieved the highest average accuracy, correctly recognizing 65.59% of the activities from the dataset followed closely by Perceptron on 65.09%. The other models performed as follows - SVM (59.3), Logistic Regression (58%) and Naïve Bayes (53.95%). The micro-average, macro-average and F1 scores for each classifier and shown in Table 1. The RNN yields the highest scores, with macro-average precision and recall scores of 0.88 and 0.41 respectively, and an F1-score of 0.46. It significantly outperforms the other classifiers at correctly recognizing a range of activities, achieving a macro average F1-score that is 228.5% higher than the Perceptron. A McNemar test with alpha = 0.05.indicated that the prediction performance of the RNN was statistically significant compared with the other four models. We hypothesize that the superiority of RNN is owed to its inherent feedback architecture, which allows it to hold latent information about the previous state of the model in memory. For example, the Sleeping activity is typically completed in 8 hours, hence a suitable time window for tracking the Sleeping activity will be too large for tracking Laundry (which typically takes 40 minutes). The RNN model is better able to adjust its weight (during the training step) to adaptively retain information.

Table 1. Micro-averaged and macro-averaged precision, recall and F1-scores

Algorithms		Precision	Recall	F1-Score
SVM	Micro Average	0.59	0.59	0.59
	Macro Average	0.18	0.10	0.09
Naïve Bayes	Micro Average	0.54	0.54	0.54
	Macro Average	0.04	0.07	0.05
Logistic Regression	Micro Average	0.58	0.58	0.58
	Macro Average	0.11	0.10	0.09
Perceptron	Micro Average	0.65	0.65	0.65
	Macro Average	0.16	0.14	0.14
RNN	Micro Average	0.56	0.56	0.56
	Macro Average	**0.88**	**0.41**	**0.46**

3.2 Activity Accuracy

We explored the performance of the models across the different activities (Table 2). The Perceptron had high overall accuracy and a high micro-average accuracy, however, the macro-average accuracy showed that it is not very good at recognizing a variety of activities. In comparison, RNN can recognize a much wider range of activities reliably than the Perceptron. Seven activities had low F1-scores across all algorithms - Washing Dishes, Mealtime, Food Prep, Watching TV, Sleeping, Reading and Grooming.

Table 2. F1-scores per activity, decreasing order of RNN's F1 score

	SVM	Naïve Bayes	Logistic Regression	Perceptron	RNN
Nailcare	0.00	0.00	0.00	0.00	1.00
Laundry	0.00	0.00	0.00	0.00	0.98
Housekeeping	0.00	0.00	0.00	0.00	0.85
Bathing	0.00	0.00	0.00	0.00	0.82
Mealtime	0.00	0.00	0.00	0.00	0.22
Dressing	0.00	0.00	0.00	0.00	0.75

No Activity	0.72	0.70	0.72	0.76	0.71
Wash Dishes	0.00	0.00	0.00	0.00	0.46
Food prep	0.00	0.00	0.00	0.00	0.14
Watching TV	0.00	0.00	0.00	0.31	0.11
Sleeping	0.16	0.00	0.00	0.00	0.04
Going Out	0.40	0.00	0.51	0.88	0.04
Reading	0.00	0.00	0.00	0.00	0.00
Grooming	0.00	0.00	0.00	0.00	0.00

4 Discussion and Conclusions

Our results demonstrate that an RNN model shows promise given a limited number of cheap off-the-shelf sensors and a low number of training examples. Activities that involved a number of distinct subtasks were difficult to detect, e.g. meal times may involve laying a table with cutlery or plates and sitting at a table. Furthermore, real users may have different routines, for example, breakfast may be a faster event and involve fewer tasks than eating dinner. This suggests careful consideration needs to be given to the set and combination of sensors to capture activities. Furthermore, high levels of personalization are likely to be necessary for detecting meaningful activities, which can be learned from datasets collected over longer periods to analyze user habits. In future work we are interested in addressing our limitations in using BLE sensors. Rather we propose the use of conventional Bluetooth. Although this will consume more energy, they can be detected by sensors on mobile devices more consistently. This approach can detect location and therefore activity recognition in multi-occupancy scenarios and may also help recognizing more personalized meaningful activities.

References

1. S. Majumder, E. Aghayi, M. Noferesti, H. Memarzadeh-Tehran, T. Mondal, Z. Pang, M.J. Deen. Smart Homes for Elderly Healthcare—Recent Advances and Research Challenges. Sensors 2017, 17, 2496.
2. M. Wallace and S. Mary, “Katz Index of Independence in Activities of Daily Living (ADL),” Urologic Nursing Journal, vol. 27, no. 1, pp. 93-94, 2007.
3. NICE, “Mental wellbeing of older people in care homes,” 2013. [Online].
4. G. Civitarese, S. Belfiore and C. Bettini, “Let the objects tell what you are doing,” in The ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg,Germany, 2016.
5. L. Niu, S. Saiki, L. D. Bousquet and M. Nakamura, “Recognizing ADLs Based on Non-intrusive Environmental Sensing and BLE Beacons,” in 8th International Conference on Indoor Positioning and Indoor Navigation, Sapporo, Japan, 2017.
6. InfluxDB, “InfluxDB: Purpose-Built Open Source Time Series Database,” [Online]. Available: https://www.influxdata.com/. [Accessed 12 April 2020].
7. Home-Assistant, “Home Assistant,” [Online]. Available: https://home-assistant.io. [Accessed 21 September 2019].
8. Wonderapps, “ATracker - Daily Task and Time Tracking,” [Online]. Available: http://www.wonderapps.se/atracker/home.html. [Accessed 19 March 2020].