Behaviour Recognition in Smart Homes

Sook-Ling Chua and **Stephen Marsland** and **Hans W. Guesgen**
School of Engineering and Advanced Technology
Massey University
Palmerston North, New Zealand
{s.l.chua, s.r.marsland, h.w.guesgen}@massey.ac.nz

1 Problem Addressed

Behaviour recognition aims to infer the particular behaviours of the inhabitant in a smart home from a series of sensor readings from around the house. There are many reasons to recognise human behaviours; one being to monitor the elderly or cognitively impaired and detect potentially dangerous behaviours. We view the behaviour recognition problem as the task of mapping the sensory outputs to a sequence of recognised activities. This research focuses on the development of machine learning methods to find an approximation to the mapping between sensor outputs and behaviours. However, learning the mapping raises an important issue, which is that the training data is not necessarily annotated with exemplar behaviours of the inhabitant. This doctoral study takes several steps towards addressing the problem of finding an approximation to this mapping, beginning with separate investigations on current methods proposed in the literature, identifying useful sensory outputs for behaviour recognition, and concluding by proposing two directions: one using supervised learning on annotated sensory stream and one using unsupervised learning on unannotated ones.

1.1 Supervised approach for behaviour recognition

Since sensor observations from the home in some way represent behaviours of the human inhabitants, our interest begins at the question of 'can we infer behaviours from these series of sensor observations?' The challenges in this task are that behaviours are rarely identical on each use; the order in which the individual components happen can change and components can be present or absent at different times (for example, making a cup of tea may involve milk, or may not, and the milk could be added before or after the water). One common approach is to build a probabilistic model (e.g. the hidden Markov model (HMM)) of how behaviours arise from these observations. However, there are a few problems that have to be solved. One is to break the sensor sequence into appropriate pieces that represent individual behaviours (i.e., segmentation) and another is to classify the behaviours.

Most current approaches assume that the activities have been segmented, and use a fixed window length to partition the input stream [Govindaraju and Veloso, 2005; Kim *et al.*, 2007]. However, different activities have different numbers of sensors activations and it can vary between presentations, so it is inappropriate to rely on a fixed window length. This doctoral study addresses this problem by using a variable window length and a set of HMMs for behaviour recognition and segmentation. The general idea was to train a set of HMMs, each representing one behaviour (e.g. one HMM to represent toileting behaviour, another HMM to represent doing the laundry, etc.). The HMMs are trained using the Expectation-Maximisation algorithm, where the observations are the sensor observations and the hidden states are the events that caused the observations. For example, the sensor observation could be that the shower faucet is turned on and a possible state that caused this token to be output is that somebody is showering.

An initial window of length 10 was slid across the sensor stream, presenting the 10 observations in the window to the sets of trained HMMs for competition. A winning HMM was chosen based on the HMM that maximises the likelihood of the 10 observations in the window. However, it is unlikely that all of the sequences in the window belong to one behaviour, and so the HMM chosen to represent it will, at best, represent only some of the activities in the window. We want to ensure that other behaviours in the window are also recognised. To do this, we used the forward algorithm [Rabiner, 1989] for re-segmentation by calculating the likelihood of each observation in the window according to the winning HMM. The activity was segmented when the likelihood of the observation dropped below a threshold value and the competition process of behaviour recognition iterates from that point onwards. To demonstrate the effectiveness of the system, we used a real smart home dataset from the MIT PlaceLab [Tapia *et al.*, 2004]. They collected the data using a set of 77 state-change sensors that were installed in an apartment. Our results showed that using a variable window length outperforms a fixed window length and that our approach of re-segmentation improves the recognition results, achieving an accuracy of more than 90% [Chua *et al.*, 2009].

Although many behaviour recognition systems have been proposed, they are based on supervised learning algorithms, and so training them requires that a dataset is prepared and annotated [van Kasteren *et al.*, 2008; Chua *et al.*, 2009]. One problem with supervised learning approaches is that they require a sufficient number of labelled data for training, which is often done by a person. This is a rather time consuming process and one that is prone to error. This therefore leads to the next research question on 'can we learn from unlabelled data without any human labelling and yet achieve an accuracy comparable to the one labelled by human?'

1.2 Unsupervised learning of activities

The next research problem is to automatically find the mapping between sensor information and behaviours in an unsupervised manner. The approach is based on compression and text analysis. The main reason why a set of activities form a behaviour is because they are repeated, albeit with variations, over time. To illustrate, making a hot drink is a behaviour, since it might well be repeated several times a day, and showering is a behaviour, because it is probably repeated daily. However, for most people, receiving a phone call while cooking dinner is not a behaviour, since it does not happen frequently. Based on this reasoning, it seems clear that we can identify behaviours from a set of sensors that are seen repeatedly in data, which can be considered as redundant in the representational sense and therefore detectable. Lossless compression can be used to exploit the redundancy in the sensory stream without any prior human labelling.

In this study, we represent the sensory data as a set of tokens (here, English characters), where a token could be the direct representation of the current sensor states being triggered (i.e., bathroom light is turned off, oven is switched on, etc.). We used the Lempel-Ziv-Welch (LZW) [Welch, 1984] algorithm to build a codebook of potential patterns (also referred to as 'words'), which is then processed to produce the prototype vectors of our clusters. However, patterns (e.g. 'hello') often do not repeat perfectly each time they are seen, such as the ordering of certain tokens being additionally present (e.g. 'heYllo') or absent (e.g. 'helo'), that tokens could be in different order (e.g. 'helol') or that there is minor variations in a token (e.g., 'hfllo). We hence want to recognise variations in the patterns. Unfortunately, LZW does not generalise to variations of the input. To allow for variability, a lossy compression is more suited to our problem. We do this by extending the LZW encoding to perform lossy compression using edit distance [Levenshtein, 1966], which measures the similarity between pairs of strings. It works by computing the minimum number of actions required to transfer one string into the other, where an action is a *substitution*, *deletion*, or *insertion* of a character into the string.

Once the prototype 'words' for the dictionary have been identified, we parse the data stream (test set) to recognise dictionary exemplars. The challenges of segmentation are that the number of tokens that define a behaviour is not fixed and the presentation of the tokens almost always vary. For these reasons, we used the edit distance to identify the matches between the 'words' in the dictionary and the data stream. A threshold is used to control the variations between word samples. We demonstrated this method on the MIT PlaceLab dataset [Tapia *et al.*, 2004]. The results showed that the discovered patterns from the data stream correspond with the inhabitant's activities and achieved an accuracy of 83%, which is comparable to the supervised method (implemented in the first year of this doctoral study) with 91% recognition accuracy.

In a separate study, we demonstrated that the proposed approach can be implemented in a semi-supervised method where the unsupervised learning approach can provide labels to training data for a supervised algorithm in a bootstrap approach to learning. The supervised algorithm can then be used to recognise behaviours from that point onwards. The results showed that the output of unsupervised method can be used to train a supervised classifier, achieving an accuracy of 89%.

2 Ongoing Work

To date this research project consisted of two key areas: (1) a supervised approach to recognising behaviours from labelled sensor data and (2) an unsupervised learning of unannotated ones. The remaining work consists of identifying informative sensors where we consider the following research question 'can we identify informative sensors in the home that could help identify inhabitant's behaviour?' This work involves investigation of information-theoretic approaches. Although it is still in its infancy, we hope the results will help us in choosing which sensors to install to identify the inhabitant's behaviours. One way to make this work efficiently may be to set the problem in the framework of Minimum Description Length (MDL), which is a model selection framework that favours the shortest possible encodings [Rissanen, 1983].

References

[Chua *et al.*, 2009] Sook-Ling Chua, Stephen Marsland, and Hans W. Guesgen. Behaviour recognition from sensory streams in smart environments. In *Australasian Conference on Artificial Intelligence*, pages 666–675, 2009.

[Govindaraju and Veloso, 2005] Dinesh Govindaraju and Manuela Veloso. Learning and recognizing activities in streams of video. In *Proceedings of the AAAI Workshop on Learning in Computer Vision*, 2005.

[Kim *et al.*, 2007] Daehwan Kim, Jinyoung Song, and Daijin Kim. Simultaneous gesture segmentation and recognition based on forward spotting accumulative HMMs. *Pattern Recognition*, 40(11):3012–3026, 2007.

[Levenshtein, 1966] Vladimir Levenshtein. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10:707, 1966.

[Rabiner, 1989] Lawrence R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[Rissanen, 1983] Jorma Rissanen. A universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, 11:416–431, 1983.

[Tapia *et al.*, 2004] Emmanuel M. Tapia, Stephen S. Intille, and Kent Larson. Activity recognition in the home using simple and ubiquitous sensors. In *Pervasive*, pages 158–175, 2004.

[van Kasteren *et al.*, 2008] Tim van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. Accurate activity recognition in a home setting. In *Proceedings of the 10th international conference on Ubiquitous computing*, UbiComp '08, pages 1–9, New York, NY, USA, 2008. ACM.

[Welch, 1984] Terry A. Welch. A technique for high-performance data compression. *Computer*, 17(6):8–19, 1984.