What's cooking and Why? Behaviour Recognition during Unscripted Cooking Tasks for Health Monitoring

Kristina Yordanova[1,2], Samuel Whitehouse[2], Adeline Paiement[2], Majid Mirmehdi[2], Thomas Kirste[1], Ian Craddock[2]
[1] University of Rostock, 18051 Rostock, Germany
[2] University of Bristol, Bristol BS8 1UB, UK

April 3, 2017

Abstract

Nutrition related health conditions can seriously decrease quality of life; a system able to monitor the kitchen activities and eating behaviour of patients could provide clinicians with important indicators for improving a patient's condition. To achieve this, the system has to reason about the person's actions and goals. To address this challenge, we present a behaviour recognition approach that relies on symbolic behaviour representation and probabilistic reasoning to recognise the person's actions, the type of meal being prepared and its potential impact on a patient's health. We test our approach on a cooking dataset containing unscripted kitchen activities recorded with various sensors in a real kitchen. The results show that the approach is able to recognise the sequence of executed actions and the prepared meal, to determine whether it is healthy, and to reason about the possibility of depression based on the type of meal.

1 Introduction

Nutrition affects our health and is an important factor for having a healthy lifespan [7]. Nutrition related diseases thus can impact our well-being and reduce the quality of life. This is particularly true for long term physical conditions, such as diabetes, eating disorders, or mental conditions such as depression, that affect the patient's willingness to prepare and consume healthy food, or people suffering from dementia disorders whose ability to prepare food is hampered by the disease's progression [10]. To reduce the costs associated with hospitalisation and treatment of these conditions, different works have attempted to provide automated home monitoring of the patient that besides reducing the hospitalisation costs, potentially improves the wellbeing of the patient as they can be monitored and treated in home settings [2].

To build such systems, one needs to recognise the user actions and goals as well as the causes behind the observed behaviour [5]. To achieve that, different works make use of knowledge-based models [9]. In contrast to data-driven approaches that rely on large amounts of sensor data and which can learn only situations that are present in the data, knowledge-based approaches have the advantage that they are able to reason beyond the sensor data. Thanks to the rules that define the possible behaviour, they can provide information about the person's situation, e.g. caused by the progression of the disease [10]. The main challenge for rule-based approaches is that they are usually unable to cope with the problems associated with real world scenarios: (a) the variability of the user behaviour results in complex models that are often computationally infeasible, and (b) the presence of imperfect sensors makes the purely symbolic models unable to cope with the ambiguity.

To address these problems, some works propose approaches that combine rules and probabilistic inference, such as [3, 8]. This type of models is also known as computational state space models (CSSMs) [4]. CSSMs combine symbolic representation with probabilistic reasoning to cope with the combination of behaviour variability and sensor noise [3, 4]. One challenge with CSSMs is that so far they have only been applied to scripted scenarios which implies simplified settings that do not address the challenges of complexity and behaviour variability present in real settings. Another challenge is the reconstruction of the user activities and goals from low level sensor data. As [11] point out, "bridging the gap between noisy, low-level data and high-level activity models is a core challenge". In this work, we address the above challenges by presenting first empirical results showing that CSSMs are able to reason about the user's actions and goals in unscripted kitchen scenarios based on low level sensor data.

2 Related Work

There are different rule-based approaches that can reason about the person's actions and goals based on context information. One such approach is where ontology-based behaviour libraries are explicitly provided by human experts [9]. One problem with this type of approaches is that "library-based models are inherently unable to solve the problem of library completeness caused by the inability of a designer to model all possible execution sequences leading to the goal" [13].

A second option for arriving at a suitable model is to mine action sequences from observations of human behaviour. Such approaches manually define an initial library of behaviours. Later, behaviour variations are added or removed based on observations of the user activities [1]. Although this approach provides a solution to the problem of keeping behaviour libraries up-to-date, it still relies on initial manual definitions of behaviour variations.

To address the problem of designing models that represent the behaviour variability without relying on large amounts of sensor data, some works propose computational state space models [3, 8, 4, 13]. CSSMs describe the actions in terms of preconditions and effects, and some of them allow probabilistic reasoning about the user state, goals and context. The manually defined model is very compact as it requires the definition of several action templates that are automatically expanded into different execution sequences based on the causal relations between them. This provides an alternative solution to the problem of manually defining all execution sequences, or relying on large amounts of annotated sensor data to learn them.

So far, CSSMs have been used only in controlled experiments with predefined execution sequences. Such experiments limit the behaviour variability typical for real-world problems, which in turn simplifies the CSSM model needed to recognise this behaviour. To our knowledge, so far there is no empirical evidence that CSSMs are able to cope with the behaviour complexity typical for a real world daily activities. What is more, so far CSSMs have been used for goal recognition only based on simulated data [3, 8]. In a previous work we proposed a CSSM model that is able to recognise the protagonist's activities during unscripted kitchen tasks [12]. The model was tested on simulated sensor data. In this work, we extend this model for activity recognition based on real sensor data by building the appropriate observation model and actions' durations. Furthermore, we extend the model for goal recognition and we show first empirical evidence that it is possible to perform both activity and goal recognition based on noisy sensor data in real-world everyday scenarios.

3 Computational Causal Behaviour Models

The CSSM approach we chose for our problem is called Computational Causal Behaviour Models (CCBM) which has been shown to perform adequate activity recognition in problems with large state spaces and noisy sensor observations [4, 13]. CCBM relies on the idea of Bayesian filtering to recognise the person's actions and goals based on observations. Figure 1 shows the dynamic Bayesian network (DBN) structure of a CCBM model. Informally, CCBM can be divided into two parts: observation model, which provides the probability of having an observation given a certain system state and system model, which provides the probability of the current system model state given the previous state. The observation model is defined through $Y_t = (W_t, Z_t)$, the observation data for time step t, e.g sensor data collected from low level sensors or simulated data. So far CSSMs had either performed only activity recognition using real sensor data [4, 13] or they have used simulated data to perform goal recognition [3, 8]. In this work we use low level sensor data to perform both activity and goal recognition. The system model consists of causal model (expressed through G_t, A_t, S_t), duration model (expressed through D_t and U_t), and action selection heuristics (which are not represented in Figure 1). G_t is the current goal the person is pursuing. In difference to [4, 13] who assume that the current goal is a constant, in this work we follow different goals and the goal can change dynamically over time. S_t is the high-level model state (the state describing the person's behaviour). It is either the result of applying a new action A_t or carrying over the old state S_{t-1}. Actions can last longer than a single time step, i.e. they have durations. U_t denotes the starting time of an action, while the boolean random variable D_t indicated the termination status of the previous action. V_t is the time stamp associated to the DBN slice.

In CCBM the causal model is presented in terms of rules that describe the possible initial and goal states, the conditions that have to hold in order for an action to be executable and the changes to the world after the action is executed. To select a new action CCBM uses action selection heuristics such as goal distance, cognitive heuristic etc. The duration model is expressed through probability distribution indicating the

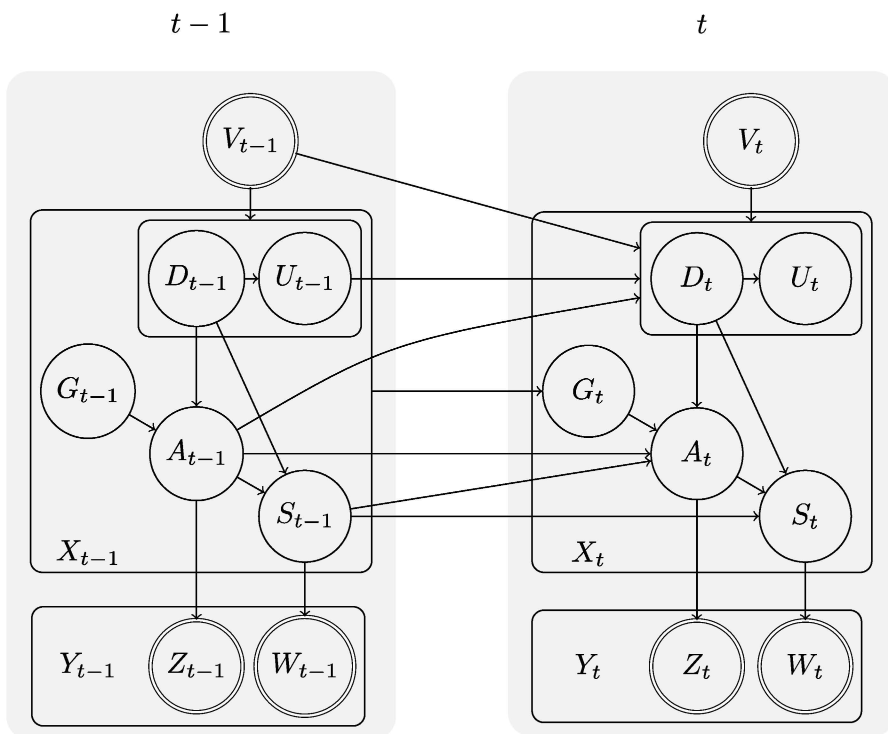

Figure 1: DBN structure of a CCBM model. Adapted from [4].

probability of terminating the action given the starting time of the action. For more details on CCBM see [4, 13].

4 Experimental Setup

Data Collection A sensor dataset consisting of 15 runs of kitchen activities was recorded in the SPHERE House which is part of the SPHERE project (a Sensor Platform for HEalthcare in a Residential Environment) [14]. The SPHERE House in Bristol (UK) is a 2-bedroom house equipped with a variety of environmental sensors. The sensor network in the kitchen of the house collects data on *temperature*, *humidity*, *light levels*, *noise levels*, *dust levels*, *motion within the room*, *cupboard and room door state*, and *water and electricity usage*. A head-mounted camera was used to record the actions of the participants to allow for annotation of the observations. The resulting dataset can be downloaded from [6].

Data Processing The original sensor data is in JSON format. The data was converted so that for each type of sensor there is a separate column. This conversion produced multiple rows with the same timestamp (in milliseconds). Rows with the same timestamp were then combined as long as per sensor type there was only one unique value. As this new format produces NAs for some sensors at a given time (due to the way in which the data is initially collected from the sensors), the NAs between two sensor readings were replaced with the first value. As the state of the most sensors is being read at a certain sampling rate but additionally there is a reading when a change in the state is detected, we believe that this simple replacement of NAs is sufficient. The resulting data contained identical observations for different action labels. To reduce the impact of this artefact on the model performance, a sliding window of 5 time steps with overlapping of 50% was used and the observations in this window were represented by the maximum value for each sensor in the window.

CCBM Models In this work we use extended version of the model proposed in [12] where they use it for activity recognition on simulated data. Here, we extend the model by adding probabilistic action durations and goal recognition and use it with real sensor data for following different goals.

Causal model: 15 specialised models, specifically fitted for the corresponding execution sequence, were developed. Furthermore, a general model was developed which can handle all sequences in the dataset. Each of the models can recognise the following action classes: *clean*, *drink*, *eat*, *get*, *move*, *prepare*, *put*, *unknown*. The model dimensions for the two model implementations can be seen in Table 1. Some additional

Table 1: Parameters for the different models.

Parameters	**General model**	**Specialised model**
Action classes	8	8
Ground actions	92	10 – 28
States	450 144	40 – 1288
Valid plans	21 889 393	162 – 15 689

discussion on the models can be found in [12].

Goals in the model: The model has three types of goals: 1) the type of meal the person is preparing (13 goals); 2) whether the meal / drink is healthy or not (4 goals); 3) whether the person is depressed or not (2 goals). For 3) we rely on the assumption that the person is depressed when they are preparing ready meals instead of cooking.

Duration model: The durations of the model were calculated based on the annotation. Empirical probability was assigned to each action class, indicating how long the model can stay in the same state before transitioning to another state.

Observation model: Two types of observation models were trained with a decision tree: 1) OM_o: All data was used both for training the OM and for testing the CCBM model. 2) OM_p: The first run was used for training the OM and the remaining runs were used for testing the CCBM. The first run was chosen because it is the only run where all action classes appear.

The decision tree for OM_o without any additional underlying model achieved mean accuracy of .52. The decision tree for the OM_p achieved mean accuracy of .39. This is to be expected as the tree was trained only on the first run and in each of the remaining experiments, a different meal was prepared usually by a different person.

Experiments For each of the observation models, the following experiments were conducted: 1) activity recognition of the action classes based on: the specific CCBM model (we call this model $CCBM_s$); the general CCBM model with one goal[1] (we call this model $CCBM_g$); the general CCBM model with multiple goals[2] (we call this model $CCBM_{g1}$). 2) goal recognition on: the different meals and drinks that can be prepared (the goal recognition is done with the $CCBM_{g1}$ model); whether the prepared meal is healthy (we call this model $CCBM_{g2}$); whether the person is depressed (we call this model $CCBM_{g3}$).

5 Results

Figure 2 shows the results from the activity recognition with the different observation and system models. It can be seen that for both OM, the CCBM models performed better than the classification with decision tree. For most models, Shapiro-Wilk normality test did not reject the null hypothesis that the samples come from normal distribution with the exception of $CCBM_s$ and $CCBM_{g1}$ with OM_p (p value $\leq$ 0.03). For that reason, to test whether the results significantly differ from each other we performed both signed t test and Wilcoxon test. Both tests showed that all CCBM models with OM_o do not significantly differ from each other (p $\geq$ 0.84 for t test and p $\geq$ 0.74 for Wilcoxon test). This indicates that the general models do not significantly reduce the recognition rate in comparison to the specialised model. This however showed that all three CCBM models with OM_o significantly differ from the decision tree tested on the training data (p $\leq$ 0.006 for t test and p $\leq$ 0.003 for Wilcoxon test). On the other hand,

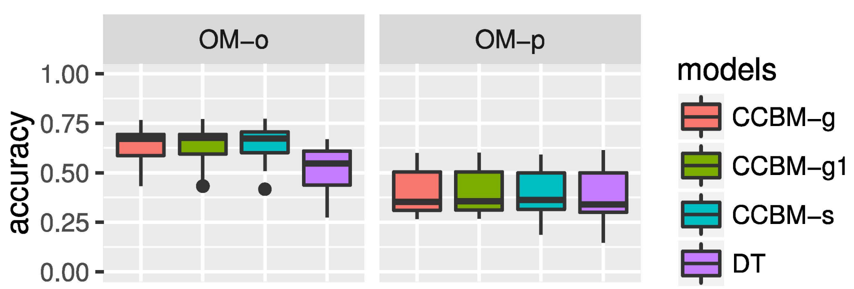

Figure 2: Accuracy for the activity recognition.

both tests illustrated that the CCBM models with OM_p and the decision tree where a test and train dataset was used do not significantly differ (p $\geq$ 0.73 for t test and p $\geq$ 0.86 for Wilcoxon test). This stands to show that the very ambiguous observation

[1]This means that all possible meals are described with "OR" statement.
[2]This means that all possible meals are described as separate goals.

model does not allow for the system model to improve the activity recognition performance. However, it also shows that despite the inaccurate OM, the system models do not reduce the recognition performance.

Figure 3 shows the results from the goal recognition for the different observation models. Here we measure the F score, as for some of the experiments more goals were recognised than were followed during the experiment. Surprisingly, OM_p did not reduce recognition of the goal for healthy meal and the type of meal. On the contrary, despite the ambiguous OM and the low activity recognition results, the models were able to perform better than when using OM_o. On the other hand, the recognition of

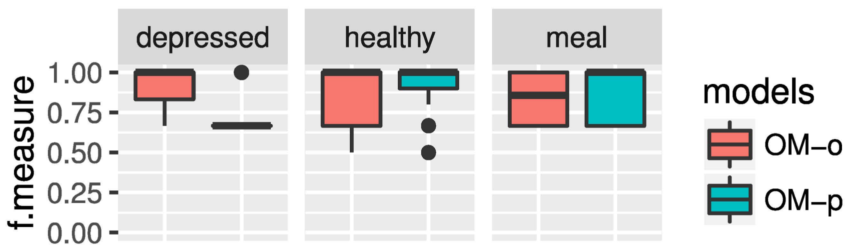

Figure 3: Accuracy for the goal recognition.

whether the person is depressed or not was lower with OM_p than with OM_o. To test whether the results for the two OM significantly differ, we performed a Wilcoxon test (the Shapiro-Wilk normality test rejected the null hypothesis that the samples come from a normal distribution with a p $\leq$ 0.001). The Wilcoxon test showed that results for the two OMs for meal recognition and healthy meal do not significantly differ (p = 0.86 for $CCBM_{g1}$, and p = 0.40 for $CCBM_{g2}$). This means that the models perform comparably and that the inaccurate observation model OM_p does not reduce the accuracy of recognising the person's goal. The results however showed that the OM has influence on the recognition of whether the person is depressed or not (p = 0.004 for $CCBM_{g3}$). In other words the more accurate OM_o significantly improved the recognition of the cause for the prepared meal. Figure 4 shows an example of the probability of preparing a healthy meal / drink for one of the experiments. The true goal is "healthy meal" and it can be seen that after time step 125 the model converges to the real goal. Although in this example, the protagonist pursues only one goal, in 4 of the experiments the goal changes throughout the experiment as the protagonist prepares more than one meal or drink.

6 Conclusion and Future Work

This work investigates the applicability of CSSMs to real world everyday activities. We applied the approach to a sensor dataset containing 15 unscripted meal preparations. The results showed that the approach is able to perform activity recognition without reducing the recognition quality compared to the performance of decision tree. They also showed the approach is able to perform goal recognition and to accurately reason

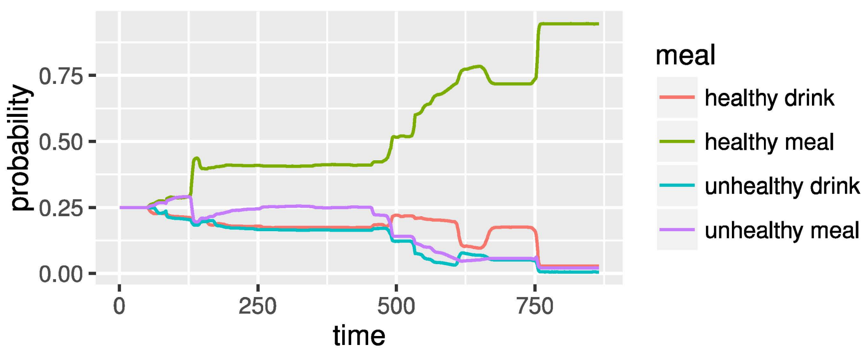

Figure 4: Evolving of the goals' probability with the accumulation of new observations during the experiment execution.

about the type of meal, whether it is healthy and whether the person preparing the meal is depressed even in the case of poor activity recognition results. These first results show that the approach has the potential to reason about the person's behaviour and the causes behind it that could hint at (the progression of) medical conditions.

In the future, we intend to compare the results from the goal recognition to state of the art approaches, such as HMMs. Furthermore, we intend to add the data from depth cameras to the observations and to investigate the influence of the type of sensor on the model performance.

7 Acknowledgments

This work is partially funded by the UK Engineering and Physical Sciences Research Council (EPSRC) within the context of the SPHERE IRC, grant number EP/K031910/1.

References

[1] L. Chen, C. Nugent, and G. Okeyo. An ontology-based hybrid approach to activity modeling for smart homes. *IEEE Transactions on Human-Machine Systems*, 44(1):92–105, Feb 2014.

[2] A. Helal, D. J. Cook, and M. Schmalz. Smart home-based health platform for behavioral monitoring and alteration of diabetes patients. *Journal of Diabetes Science and Technology*, 3(1):141–148, January 2009.

[3] L. M. Hiatt, A. M. Harrison, and J. G. Trafton. Accommodating human variability in human-robot teams through theory of mind. In *Proceedings of IJCAI*, pages 2066–2071. AAAI Press, 2011.

[4] F. Krüger, M. Nyolt, K. Yordanova, A. Hein, and T. Kirste. Computational state space models for activity and intention recognition. a feasibility study. *PLoS ONE*, 9(11):e109381, 11 2014.

[5] F. Krüger, K. Yordanova, C. Burghardt, and T. Kirste. Towards creating assistive software by employing human behavior models. *Journal of Ambient Intelligence and Smart Environments*, 4(3):209–226, May 2012.

[6] M. Mirmehdi, T. Kirste, S. Whitehouse, A. Paiement, and K. Yordanova. Sphere unscripted kitchen activities. University of Bristol, 2016. https://data.bris.ac.uk/data/dataset/raqa2qzai45z15b4n0za94toi.

[7] S. D. Ohlhorst, R. Russell, D. Bier, D. M. Klurfeld, Z. Li, J. R. Mein, J. Milner, A. C. Ross, P. Stover, and E. Konopka. Nutrition research to affect food and a healthy lifespan. *Advances in Nutrition: An International Review Journal*, 4:579–584, September 2013.

[8] M. Ramirez and H. Geffner. Goal recognition over pomdps: Inferring the intention of a pomdp agent. In *Proceedings of IJCAI*, pages 2009–2014, 2011.

[9] P. C. Roy, S. Giroux, B. Bouchard, A. Bouzouane, C. Phua, A. Tolstikov, and J. Biswas. *A Possibilistic Approach for Activity Recognition in Smart Homes for Cognitive Assistance to Alzheimer's Patients*, pages 33–58. Atlantis Press, 2011.

[10] A. Serna, H. Pigot, and V. Rialle. Modeling the progression of alzheimer's disease for cognitive assistance in smart homes. *User Modeling and User-Adapted Interaction*, 17(4):415–438, September 2007.

[11] Gita Sukthankar, Christopher Geib, Hung Hai Bui, David Pynadath, and Robert P. Goldman. *Plan, Activity, and Intent Recognition: Theory and Practice*, chapter Introduction, pages xix–xxxv. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 2014.

[12] S. Whitehouse, K. Yordanova, A. Paiement, and M. Mirmehdi. Recognition of unscripted kitchen activities and eating behaviour for health monitoring. In *Proceedings of the 2nd IET International Conference on Technologies for Active and Assisted Living (TechAAL 2016)*, London, UK, October 2016. INSPEC.

[13] K. Yordanova and T. Kirste. A process for systematic development of symbolic models for activity recognition. *ACM Transactions on Interactive Intelligent Systems*, 5(4):20:1–20:35, December 2015.

[14] N. Zhu, T. Diethe, M. Camplani, L. Tao, A. Burrows, N. Twomey, D. Kaleshi, M. Mirmehdi, P. Flach, and I. Craddock. Bridging e-health and the internet of things: The sphere project. *IEEE Intelligent Systems*, 30(4):39–46, July 2015.