Activity Prediction for Agent-based Home Energy Management

Ngoc Cuong Truong, Long Tran-Thanh, Enrico Costanza and Sarvapali D. Ramchurn
Agent, Interaction and Complexity Group,
School of Electronics and Computer Science
University of Southampton, UK.
{nct1g10 ltt08r ec sdr}@ecs.soton.ac.uk

ABSTRACT

In this paper, we address the problem of predicting the usage of home appliances where a key challenge is to model the everyday routine of homeowners and the inter–dependency between the use of different appliances. In particular, given an efficient day–ahead prediction of electrical usage, home energy management systems can suggest homeowners when is the best time to run appliances in order to save cost, without violating their preferred everyday habits. To this end, we propose an agent based prediction algorithm that captures the everyday habits by exploiting their periodic features. In addition, our algorithm uses a episode generation hidden Markov model (EGH) to model the inter–dependency between appliances. We demonstrate that our approach outperforms existing methods by up to 40% in experiments based on real–world data from a prominent database of home energy usage. We also show that the computational cost of our algorithm is 100 times lower on average, compared to that of the benchmark algorithms.

1. INTRODUCTION

In the face of dwindling fossil fuels, an ageing electricity distribution infrastructure, and the adverse effects of high levels of green house gasses on climate change, the problem of generating affordable and clean electricity reliability is one of the greatest challenges of this century [5]. To make matters worse, energy demand is growing at a fast pace given the electrification of heating and transport [6]. In more detail, figures show that worldwide energy consumption in the domestic sector accounts for approximately 27% of all electricity consumption. In the UK in 2009, domestic electricity use accounted for approximately 24% of the country's overall electricity consumption and approximately 30% to the UK's total CO_2 emissions [7]. Given this, energy demand in housing and domestic appliances is rapidly increasing, hence, improving the efficiency of energy usage in the home can significantly impact on national CO_2 emissions.

Now, to make the use of the electrical devices in the home more efficient, and thus, to reduce both carbon emissions and cost, a set of agent based demand side management techniques have recently been introduced to optimise the schedule of loads [16]. In particular, in these approaches an agent controls the smart meter and takes into account the real time carbon content/cost if electricity in order to optimise the schedule of specific loads. However, these techniques typically do not take into account the homeowner's preferences in their optimisation. Thus, such scheduling methods might not be acceptable to real homeowners as they do not meet the latter's everyday routine. For example, suppose that a homeowner prefers to use her washing machine on weekends when she has time to take the clothes out to dry and iron them. Thus, she would not accept a schedule that would put the use of the washing machine on another day even it is cheaper to do so.

Moreover, demand–side management algorithms ignore inter-dependencies between the usage of different appliances. In particular, the home owner might use the dishwasher and the oven on the same day, or prefers turning on the TV whenever she starts cooking. Given this, schedules that do not take these possible inter–dependencies might not meet the homeowner's preferences either, and thus, would not be accepted.

Within all the aforementioned scenarios, the main challenge is to predict the energy consumption activities of homeowners, so that the agent can design optimal schedules by planning ahead the electricity usage that meets the human's preferences. To date, no research has been done to address both predicting human behaviour and inter-dependency between the usage of appliances within the energy domain. In particular, existing human activity prediction models are typically designed for location prediction [11, 19], and thus, may not be adaptable for modelling complex inter–dependencies between the usage of different appliances within a typical home (as location prediction has to deal with only one data stream). In turn, a number of efficient methods for tackling complex prediction problems with multiple inter–dependent data streams have been developed [18, 12]. As we will show in Section 4, since they are not designed specifically for human activities, they may not be appropriate for this purpose.

Against this background, we propose a novel approach to predicting the energy consumption of different home appliances, that takes into account both the human routine activities and the inter–dependency between appliances. To do so, we rely on the common assumption that human behaviour follows a certain cyclic pattern [11]. Based on this, we build a model that exploits this cyclic behaviour. To handle the inter–dependency between the appliances, we use the episode generation Hidden Markov model (EGH) [18] to efficiently identify the patterns that form the inter–dependency between the usage of the appliances. By putting the two models together, we demonstrate that our approach outperforms the state–of–the–art, that only focus on either human behaviour detection on inter–dependency pattern identification.

Against this background, we contribute to the state–of–the–art as follows.

- We propose the first algorithm that merges techniques for human behaviour prediction and inter–dependency pattern identification in order to efficiently predict the usage of electrical appliances in the home.

- We demonstrate through extensive simulation, using real–world data, that our algorithm outperforms the state–of–the–art by up to 40%, while its running time is typically 100 times faster than that of the benchmarks.

The remainder of the paper is structured this paper as follows. In the Section 2, we review existing models that could be used in our scenario. We then formalise our problem scenario in Section 3. Section 4 evaluates the algorithm and analyses the results that we obtain from our experiments. Finally, Section 5 concludes.

2. RELATED WORK

As our approach lies on the nexus of human behaviour prediction and prediction with multiple inter–dependent data streams, we discuss these two areas in more detail in this section.

Prior work on human behaviour prediction mainly address user location prediction in spatio–temporal domains. In particular, prior work has focused on predicting the user's location using mobile applications [11, 2, 19]. These approaches include, but are not limited to, prediction tasks with eigenvalue decomposition [9], non-linear time series analysis of arrival times [17], and variable order Markov models [2]. More recently, a number of works relied on the use of the Pitman–Yor Process to detect whether the homeowner is away from home [19, 10]. Recently, McInerney *et al.* addressed the problem of predicting human behaviour with sparse data [15]. Although these techniques are efficient at predicting a single user's behaviour, they do not address the challenges of the inter–dependency between different sequences of data (i.e. history of appliance activity).

On the other hand, graphical models are typically used to describe the aforementioned inter–dependencies. In particular, graphical models have been used to represent the structure of conditional independence among random variables [8]. In addition to these models, Bayesian networks [1, 13] are also widely used for cases when missing data entries occur. More recently, Gunawardana *et al.* proposed PCIM, a technique for modelling inter–dependencies among variable [12]. The authors claim that their model is efficient at capturing the dependencies between the data streams. Another popular class of methods is the rule–based approach, where the main goal is to detect dependency patterns (i.e. rules). In particular, these methods are used to generate rules that are based on discovered episodes (i.e. patterns), and the future activities can be predicted based on those rules [3]. A state–of–the–art technique in this domain is the Episode Generation Hidden Markov Model (EGH) [18].

Both rule–based methods and dependency models, however, are not designed to exploit the cyclic behaviour of human users, and thus, might fail in predicting human related data sequences, as is the case in our settings (see Section 4 for more details). To fill the aforementioned gaps, we develop a human routine activities prediction method that can take into account both the cyclic patterns of human routine and the inter–dependencies between the usage of electrical appliances.

3. PREDICTING THE USAGE ACTIVITIES OF APPLIANCES

In this section we describe the problem of predicting appliance usage activities in more detail. To do so, we first describe the formalisation of our problem in Section 3.1. We then introduce our algorithm in Section 3.2.

3.1 Model Description

In this paper we aim to design an agent based approach that predicts whether a particular appliance is used (and for how many times) in the next day, in order to estimate the future energy consumption of the home. To do so, we assume that we have a finite set of consumer activities, where different types of activities are distinguished by $labels\ l \in L$. An activity profile of label l $a_{l,t}$ is a tuple $\langle t, l, n \rangle$, composed of a time step t (measured in days), a label l and number of usage n, that denotes the number of occurrences of label l on day t. For example, such usage profiles are "Washing Machine was used on Tuesday" (i.e. $a =$ (Tue, washing machine, 1)), or "Oven was used on Thursday" (i.e. $a =$ (Thu, oven, 1)). For the sake of simplicity, we only consider the binary case of occurrence. That is, we assume that $n \in \{0, 1\}$. Let $x_t = \langle a_{1,t}, a_{2,t}, \ldots, a_{L,t} \rangle$ denote the usage profile of day t that contains the information about the usage of each label $l \in L$ on day t. The appliance usage history h_t of time slot t is the sequence $h_t = \{x_1, x_2, \ldots, x_t\}$. Our goal is to estimate x_{t+1} for any $t > 0$ with high accuracy, given h_t. To solve this problem, the agent has to take into account two main dependencies that underlie a consumer's activities:

- Time dependencies: Consumers can trigger their activities at different time slots which satisfy their needs and daily routine. For example, if we use a washing machine at time t, it is unlikely we will use it again at time $(t + 1)$) due to lack of dirty clothes to wash.

- Activity inter–dependencies: Some types of activities may depend on other activities. For example, when one cooks, one might need to use the oven and microwave, then one might need to use the dishwasher to clean the dirty dishes. Therefore, the activities of using the oven, microwave, and dishwasher are dependent on cooking.

In addition, we have to take into account the cyclic behaviour of users as well. Following the work of Gonzalez *et al.* [11], we assume that the human behaviour in terms of appliance usage forms cycles of weekly periods.

In what follows, we propose a novel algorithm that, by taking the aforementioned dependencies and cyclic behaviour into account, efficiently predicts future activities.

3.2 The Prediction Algorithm

As mentioned earlier, the foundations of our prediction algorithm rely on the EGH method. However, as EGH is not designed for detecting human activities, we tailor the model to fit our settings by exploiting the periodic features of the human everyday routine. In addition, we further advance EGH by using an adaptive prediction threshold value in order to improve the efficiency of prediction. In particular, if the probability of the occurrence of an event (i.e. usage of an appliance) exceeds this threshold, the algorithm predicts that the event will occur. Given this, we first detail the training phase of our approach in Section 3.2.1, where we use a set of training data to build up a dependency model for the correlations between the usage of appliances. We then continue with the description of our human routine model in Section 3.2.2. Based on these models, we then construct a mixture model of the significant episodes (i.e. sets of possible inter–dependency rules, see Section 3.2.1 for more details) in order to calculate the probability of activities' occurrence in Section 3.2.3. Finally, Section 3.2.4 focuses on the prediction model in detail.

3.2.1 The Inter–Dependency Model

To build the inter–dependency model, we rely on the EGH approach described by Srivatsa *et al.* [18] as follows. Suppose that the inter–dependency between the usage of different appliances follow some probabilistic patterns. For example, a typical pattern can be the following: it is likely that the dryer is also used after the usage of the washing machine, or the use of the microwave follows the use of the oven within two days. In our model we denote these patterns as *episodes* (i.e. a sequence of appliance activities).

To evaluate the likelihood of an episode, we calculate its probability of occurrence, given the history of appliance usage. To do so, we use the EGH approach, which assigns a discrete hidden Markov model (HMM) to the corresponding episode. In particular, suppose that episode α consists of N activities $(a_{l_1,t_1}, a_{l_2,t_2}, \dots, a_{l_N,t_N})$ such that $t_1 \leq t_2 \cdots \leq t_N$. EGH assigns a HMM $H_\alpha = (S, \Delta_\alpha, \eta_\alpha)$ to α such that $S = \{1, \dots, 2N\}$, denotes the state space, $\Delta_\alpha = (a_{l_1,t_1}, a_{l_2,t_2}, \dots, a_{l_N,t_N})$ denotes the activities, and η_α is the noise parameter. The latter is set equal to $\frac{T - Nf_\alpha}{T}$, where T is the total number of activities in the training dataset and f_α is the number of times the episode α occurres in the training dataset, if it is less than $\frac{M}{M+1}$ and to 1 otherwise. The intuition behind the use of Λ_α is that it represents a Markov model of a sequences activities that contains the corresponding episode (see [18] for more details).

Now, we calculate its corresponding frequency f_α of each possible episode α within the dataset (i.e. the number of times the episode occurs in a non–overlapping way). Hereafter, we only consider those episodes that have a frequency f_α higher than $\frac{T}{N(\alpha)M}$, where M is the number of activity types, and $N(\alpha)$ is the number of activities within α. We denote these as *significant episodes*. The reason we focus on these episodes (and thus, ignore the rest) is that the others are unlikely to occur given the training data set. These significant episodes can be regarded as rules that model the inter–dependency between the occurrence of different activities.

3.2.2 The Human Routine Model

By building up the set of significant episodes, we can then predict the occurrence of activities within the next time step by analysing whether they can be a part of a significant episode. However, as the number of significant episodes can be an exponential in size of the training dataset, EGH is inefficient computation–wise. In addition, EGH might overestimate the occurrence of activities, due to redundant episodes. In particular, due to the cyclic nature of human routines, a sequence of activities that consists of two non–overlapping, but identical, episodes can also be regarded as a significant episode. This might lead to inaccurate estimation of the probability of an activity's occurrence.

To address these challenges, we reduce the set of potential significant episodes by exploiting the cyclic features of human everyday routine. In particular, we assume that human behaviour in home energy usage follows a weekly cycle. Thus, if the goal is to predict whether a target activity type l occurs on the specific day d, we only consider activities that happen at most one week earlier than an occurrence of l on the same day d in the past. More formally, let K denote the number of occurrences of the target activity type l on the specific day d of the week in the activity usage history h_{t-1}. Thus, for each label l and the prediction day of the week d, from the original training dataset D, we extract a training set $D_{l,d} = \{X_i\}_{i=1}^K$, where $X_i = < x_{t_i-7}, \dots, x_{t_i-1} >$ is the weekly preceding window of activities from x that immediately preceded the i^{th} occurrence of l in x, and t_i is the time that the target activity type l occurred at the i^{th} in the activity sequence.

Given this reduced training dataset $D_{l,d}$, we then use the EGH approach to identify the significant episodes. The intuition behind this technique can be described as follows. We assume that the activities are typically influenced only by activities within a week time (i.e. older activities do not have affect on them), it is more efficient to only consider these past activities. By doing so, we can reduce the computational costs and also improve the quality of prediction (as we will demonstrate later in Section 4).

3.2.3 The Mixture Model

Given the episode reduction using the human routine model, we now turn to the discussion of how to use these episodes to predict the future activities. To do so, we first analyse the joint influence of these episodes on the probability of a single activity's future occurrence. Suppose that for a given training data set $D_{l,d} = \{X_i\}_{i=1}^K$, we have calculated a set of significant episodes, denoted as $F^s = \{\alpha_1, \dots, \alpha_J\}$, and each HMM H_{α_j} of episode α_j. Now, to predict activity a in the next time step t, we use these episodes in order to calculate the probability of occurrence. To do so, we calculate the probability that a is a part of a significant episode. However, as an episode typically has a certain positive probability of indicating the occurrence of a, we have to take into account all of them. To model the effect of this joint influence, we compute a mixture model Λ_l (i.e. a combination of probabilistic processes) of the significant episodes' HMMs. This mixture model then can be used to predict the future occurrences of the target activity a. In what follows, we first build the aforementioned mixture model and then demonstrate how to predict future activities.

Now, the likelihood function of the training dataset D under a mixture model Λ_l can be written as follows:

$$P[D|\Lambda_l] = \prod_{i=1}^{K} P[X_i|\Lambda_l] = \prod_{i=1}^{K} \left(\sum_{j=1}^{J} \theta_j P[X_i|H_{\alpha_j}] \right) \quad (1)$$

where θ_j, $j = 1..J$ are the mixture coefficients of Λ_l (with $\theta_j \in [0, 1]$ for all j, and $\sum_{j=1}^{J} \theta_j = 1$). Recall that each HMM H_{α_j} is fully characterised by the significant episode α_j and its noise parameter η_{α_j}. Given this, the likelihood of the activity sequence X_i, given the HMMs $\{H_{\alpha_j}\}_{j=1}^{J}$, is computed by approximating the likelihood along the corresponding most likely state sequence:

$$P[X_i|\Lambda_{\alpha_j}] = \left(\frac{\eta_{\alpha_j}}{M}\right)^{|X_i|} \left(\frac{1-\eta_{\alpha_j}}{M}\right)^{|\alpha_j| f_{\alpha_j}(X_i)} \quad (2)$$

where $|X_i|$ denotes the length of sequence, X_i, $f_{\alpha_j}(X_i)$ denotes the non-overlapped occurrences-based frequency of α_j in the sequence X_i, and $|\alpha_j|$ denotes the size of the episode α_j.

We use the Expectation Maximisation (EM) algorithm to estimate the set of mixture coefficients of the mixture model Λ_l. In particular, the algorithm is initialised with the current guess for the mixture coefficients, denoted by $\Theta^g = \{\theta_1^g, \ldots, \theta_J^g\}$. These mixture coefficient values is initially set to be uniform, that is, $\theta_j^g = \frac{1}{J}$ for every $j \in J$. We then use these current guesses to update the mixture coefficients as follows. Let $\Theta^{new} = \{\theta_1^{new}, \ldots, \theta_J^{new}\}$ denote the new values of these coefficients. Given this, we have:

$$\theta_q^{new} = \frac{1}{K} \sum_{i=1}^{K} P[q|X_i, \Theta^g] \quad (3)$$

where $q = 1..J$. Let $P[q|X_i, \Theta^g]$ denote the posterior probability for the q^{th} mixture component, with respect to the window $X_i \in D_l$, which can be computed using Bayes' Rule:

$$P[l|X_i, \Theta^g] = \frac{\theta_l^g P[X_i|H_{\alpha_l}]}{\sum_{j=1}^{J} \theta_j^g P[X_i|H_{\alpha_a}]} \quad (4)$$

The new set of mixture coefficients Θ^{new} is then used as the current set of guesses (i.e. Θ^g) of the mixture coefficients. The process is repeated until the coefficients converge.

3.2.4 The Prediction Model

Given the mixture model $\Lambda_l = \{(\alpha_j, \theta_j)\}_{j=1,\ldots,J}$, we now turn to the prediction phase of our approach. Let t denote the current time. For the set of target activity labels $l \in$, we want to predict their occurrences in the next day, $t + 1$. As we are mainly interested in occurrences of recent activities of the users, therefore, we construct a 7—length window of activities from the weekly period $[t - 7, t]$. The recent list of activities can be written as $X = [a_{t-7+1}, \ldots, a_t]$. We then estimate the likelihood of this recent activity sequence, given the mixture model, Λ_l, that is obtained from the training phase. The algorithm determines the occurrence of the target activity at time step $(t + 1)$ based on the value of the threshold. In particular, if the probability of the window under the mixture model is greater or equal than the prediction threshold, the algorithm predicts that the target activity will occur at the next time step $(t + 1)$. Otherwise, if the probability of the window under the mixture model is less than the threshold value, the algorithm predicts that the target activity will not occur at the next time step $(t + 1)$. This raises the difficult task of how to define the reasonable threshold value that would work best for each dataset in the scenario.

Finally, we compute the value of the prediction threshold as follows. We first compute the likelihood of each preceding window of the target activity, X_i where $i = 1, 2, \ldots, K$, under the mixture model Λ_l. Then, the final threshold value we use for prediction is the minimum value of all the likelihood values of preceding windows, which can be written as below:

$$\varsigma = min\{P[X_i|\Lambda_l]\} \quad (5)$$

where $i = 1, 2, \ldots, K$. Note that since none of the existing prediction algorithms are designed specifically for predicting the usage of electrical appliances in the home, our agent based method is the first that addresses this particular problem. In what follows, we demonstrate the by using our approach, an agent can achieve efficient performance in accurately predicting the usage of electrical consumption activities.

4. EMPIRICAL EVALUATION

Given the prediction model, we now turn to demonstrate how our algorithm outperforms the existing prediction algorithms in predicting the next day usage of electrical appliances in the home. To do so, we first introduce a set of benchmark algorithms against which we compare our method (Section 4.1). We also detail two real–world datasets that we use in our experiments in Section 4.2. Finally, we show our results in Section 4.3.

4.1 Benchmark Algorithms

As mentioned in Section 1, related work has typically focused on single user behaviour prediction and dependency model prediction for non–human data. Given this, to demonstrate that these algorithms are not designed for our settings, we choose a number of state–of–the–art methods from these domain to benchmark against. In particular, we compare our method against the following approaches:

- Pitman–Yor Process (PYP): This algorithm is designed for predicting the presence at locations of a single user [10]. In particular, it regards a set of binary observations (i.e. whether the user is at a certain place at a particular time), modelled by beta distributions. The parameters of these distributions are conditioned on a *day type* category for that particular day. These day types are latent states that enable the clustering of behaviours, introducing dependencies between each separate observation of the day (allowing prediction). They are generated by a Dirichlet distribution with unknown component coefficient parameters. All of these parameters are inferred from the training data during learning. The prediction is then done by finding the probability of the day type given the day of the week (see [10, 19] for more details).

- PCIM: The piece–wise constant conditional intensity model (PICM) is a state–of–the–art approach in predicting multiple–source web data where data from different sources might depend on each other. In particular, it uses a set of piece–wise constant dependency functions to capture the correlation between labels (i.e. data from different sources). It uses these functions to create a decision learning tree to describe the inter–dependency model. Based on this model, it then estimates the probability of event occurrence in the future by using forward and importance sampling (for more details, see [12]).

- EGH: We also compare our algorithm against the original EGH method to demonstrate that, by adding the extensions described in Section 3, we make the approach suitable for our domain, and thus, advance its performance.

In addition, for the sake of simplicity, we refer to our algorithm as EGH-H (i.e. EGH for human routine prediction).

4.2 Real–World Datasets

In this section, we describe the REDD dataset [14] and our own collected data. These datasets, in fact, are collected from real–world applications and are used in our experiments to evaluate our algorithm and the benchmark approaches.

4.2.1 The REDD dataset

The REDD data set includes six different houses. These houses have been monitored for approximately 35 days with sub-meters installed on multiple relevant electrical home appliances. The data in the REDD set is the power consumption for the specific devices every 3 seconds. We converted the raw data of power consumption into a list of cyclic on-off events as follows:

- We set a threshold of power consumption (typically 55W) to determine the periods that the appliances turned on. We store all these segments of durations that the appliances turned on.
- We set a *gap allowance parameter* for two consecutive segments. If the gap between these two consecutive segments is greater than the gap allowance, we connect these two segments together, and considered as one segment.
- We select a *noise removal parameter* to filter the noise of the data. All the segments that are less than the noise parameter are removed.

The *gap allowance parameters* and *noise removal parameter* are adjusted to adapt to the behaviour of the appliances. For example, the period of *dishwasher* cycle is typically over 30 minutes, thus we set the noise parameters for the full cycle of using dishwasher is up to 30 minutes. In addition, the power consumption of the dishwasher is controlled by the built-in sensor of temperature in the dishwasher. Given this, the power consumption consumed for the dishwasher is fluctuated. However, the gap between these two consecutive periods that the power consumption over 55W is less than 10 minutes. Thus, we set the gap allowance parameter for the dishwasher is 10 minutes. Then, we observed that there were 3 houses which do not have enough information to judge the performance of the prediction. Hence, we only carry out our tests on data from 3 houses. For those houses, we use the first 20 days as a training data set, and the remaining 15 days as a testing set.

4.2.2 Data Collected from FigureEnergy

In addition to the REDD dataset, we also use another dataset collected from homeowners in the UK. In particular, this included 13 participating homes. Each user (i.e. homeowner) was given a smart meter, which integrated into the user's home and transferred data into the application's server over the internet. Users then could observe their aggregated energy consumption from their web browser using FigureEnergy, a web–based application designed for appliance usage labelling (see [4] for more details).

Type	Start time	End time	Energy usage	Baseline	UserID
oven	2011-08-31 18:58:27	2011-08-31 19:42:47	1.562	0.069	32
kettle	2011-09-01 07:21:05	2011-09-01 07:26:17	0.094	0.007	32
shower	2011-09-01 08:12:45	2011-09-01 08:21:08	0.102	0.0144	32
tv	2011-09-01 17:54:16	2011-09-01 19:18:20	0.3902	0.151	32
stove	2011-09-01 19:18:20	2011-09-01 19:41:16	0.585	0.0396	32

Table 1: An example of data collected from FigureEnergy.

This application allows users to identify and label individual activities as follows. By clicking on the graph with their mouse and dragging, users can select a segment and fill information about the activities that they spent. There is a preset list of labels that users can choose from for their activities. The labels will be on the aggregate energy consumption to show users the results that they have annotated (see Figure 1). An example of the collected data can be seen in Table 1.

In the next section, we describe the experimental settings, and empirically evaluate the performance of our methods to other state–of–the-art on REDD, and FE datasets.

4.3 Experimental Results

The experimental settings are described as follows. Since we only consider binary prediction, we set a value of threshold at 0.5 to determine occurrence of the labels. That is, for a specific label at a given time, if the predicted probability is greater or equal than 0.5, the label is considered to have a high probability of occurring. Otherwise, if the predicted probability is less than 0.5, the label is considered not to occur. We first use the REDD dataset to evaluate the performance of the algorithms (see Section 4.3.1) and then continue with the data collected from FigureEnergy (Section 4.3.2). We also compare the average running time of the algorithms in Section 4.3.3.

4.3.1 Performance on REDD Data

Here, we run our algorithms to predict all the labels of the REDD dataset. We then compute the *F–score* of each algorithm to measure their accuracy of prediction. The results can be seen in Table 2. In overall, our method outperforms other state–of–the-art by up to 40%. In particular, it is better than PYP, EGH, and PCIM by approximately 73%, 40%, and 75% on average, respectively. Note that since home 1

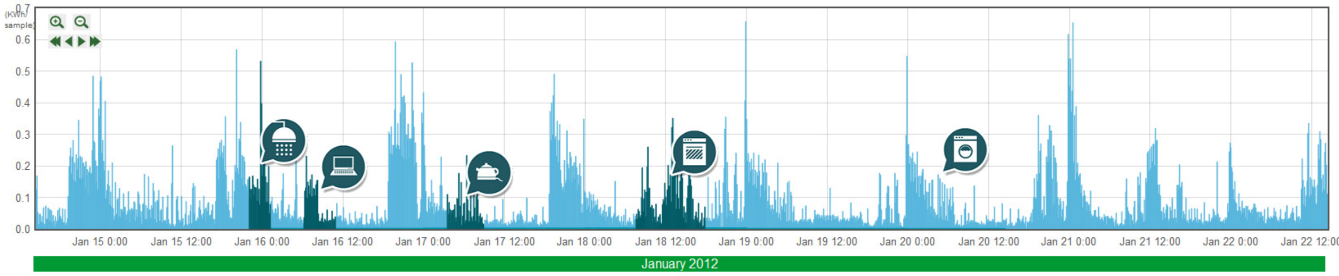

Figure 1: An example of using annotation in the FE application.

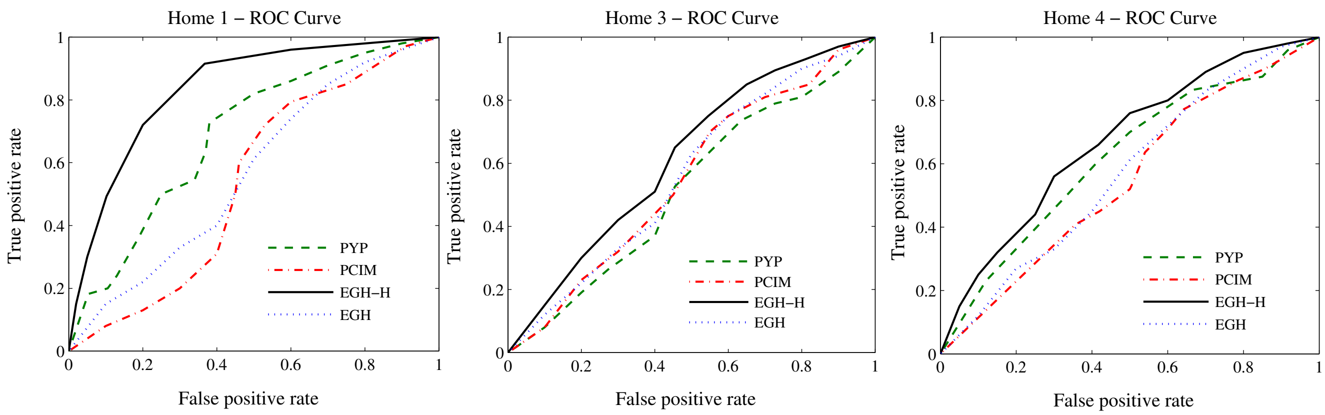

Figure 2: ROC curve of the algorithms run on three homes from REDD.

Home ID	No. Appliances	PYP	PCIM	EGH	EGH–H
1	8	0.770	0.330	0.743	**0.844**
3	3	0.190	0.308	0.571	**0.772**
4	3	0.4	0.714	0.368	**0.75**

Table 2: Overall F–score results three homes from REDD.

has the most detailed data, all the algorithms typically provide their best performance on this home. An exception is the PCIM method, which performs by far the worst. The reason here is that due to the large size of available data, the PCIM overfits the inter–dependency model (since it does not take into account the cyclic feature of human routine). Given this, it fails to correctly detect the occurrence of activities.

To better demonstrate this, we depict the receiver operating characteristic (ROC) curve of the algorithms for each home in Figure 2. From this figure, we can see that our algorithm dominates all the others. In particular, the area under the curve (AUC) of EGH–H in home 1 is 0.84, while the AUC value for PYP, EGH, and PCIM is 0.68, 0.56, and 0.53, respectively. We can also observe that since data from homes 3 and 4 is less detailed, all the algorithms provides worse performance, compared to themselves in home 1. However, our algorithm still dominates the benchmark approaches.

We continue with evaluating the performance on appliances that are likely to be strongly correlated with each other. In particular, in our initial observations, we learnt that the kitchen's appliances are most likely to be used together by consumers. Therefore, we focus on a set of kitchen appliances, such as oven, dishwasher, microwave, kitchen outlet 1, or kitchen outlet 2, to perform our tests. We generate all possible combinations of these appliances for the test data, then matched each combination to the houses that have these labels of activities. We then compute the F–score of the algorithms for each test data. The results are shown in Table 3. On average, the F–score obtained by our method is approximately 0.84, while the F–score of PYP is 0.76, PCIM is 0.50, and EGH is approximately 0.73. Hence, we can conclude that our method has better prediction of 13% compared to PYP, 15% compared to EGH, and 90% compared to PCIMs in predicting appliance dependencies.

4.3.2 *Performance on Data from FigureEnergy*

In this section, we test the performance on three selected homes from the FE dataset. In particular, the other homes did not provide sufficient data. Thus, we could not be able to set up a proper training dataset for those homes. Similar to the previous section, we also consider the overall performance of the algorithms. Note that within the FE data set, the labels of energy usage activities were mainly annotated by consumers. Thus, the uncertainty of the labels is high and this uncertainty in labels could cause the learning structure of dependencies to behave incorrectly, and hence worsen the prediction performance[1]. Therefore we selected labels that occurred sufficiently in both training and testing datasets. The F–score of the algorithms for each house is depicted in Table 4. We can observe that, due to the uncertainty of the

[1]This is an aspect which we will further investigate as part of our future work

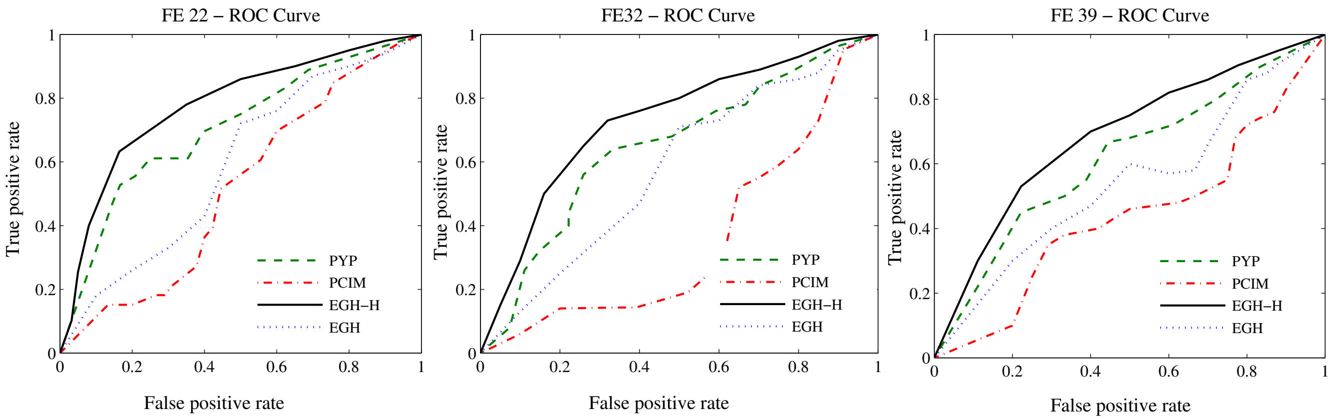

Figure 3: ROC curves of the algorithms for three homes from FigureEnergy.

Test Data	PYP	PCIM	EGH	EGH–H
{Oven, dishwasher}	0.4444	0.486	0.536	0.667
{Oven, microwave}	0.860	0.698	0.776	0.88
{Oven, kitchen outlets 1}	0.813	0.697	0.750	0.850
{Oven, microwave, Dishwasher}	0.762	0.560	0.732	0.842
{Oven, Dishwasher, microwave, kitchen outlets1}	0.849	0.308	0.800	0.883
{Oven, Dishwasher, microwave, kitchen outlets1, kitchen outlets2, washing machine}	0.838	0.279	0.792	0.903

Table 3: F-Score performance of predicting appliance dependencies in REDD

Home ID	No. Appliances	PYP	PCIM	EGH	EGH–H
FE22	6	0.550	0.40	0.6	**0.613**
FE32	4	0.488	0.444	0.619	**0.667**
FE39	4	0.444	0.491	0.609	**0.623**

Table 4: Overall F–score results on three homes from FigureEnergy.

homeowners' manual labelling process, the performance of the algorithms are much lower, compared to the case of the REDD dataset. However, EGH–H still provides the highest accuracy in predicting future activities.

For more detailed analysis, we also plot the ROC curve of the algorithms for these homes in Figure 3. From this figure, we can observe that PYP provides the second best performance (after EGH–H). A possible reason is the following: In the FigureEnergy dataset, the collected labels typically

Test Data	PYP	PCIM	EGH	EGH–H
{tv, kettle}	0.736	0.434	0.732	0.75
{toaster, microwave}	0.514	0.444	0.486	0.561
{computer, tv}	0.705	0.692	0.643	0.727

Table 5: F-Score results of predicting appliance dependencies in FigureEnergy.

show independence from each other. Thus, a naive extension of the single human behaviour prediction such as PYP is expected to work well in these settings.

To justify the previous argument, we also analysed the dataset for appliances with high possible inter–dependency. The list of these is depicted in Table 5. In addition, we also show the performance of the algorithms in predicting these dependencies. Here, our algorithm also outperforms the others. However, the improvement is not that high, compared to the case of REDD data. In particular, EGH–H outperforms PYP, PCIM and EGH by 4%, 34%, and 10%, respectively.

4.3.3 *Average Running Time of the Algorithms*

Having evaluated the prediction accuracy of the algorithms, we now turn to evaluate the running time of each algorithm. In particular, we run the algorithms on an Intel(R) Xeon(R) computer (64–bit operating system) with 2.67 GHz and 12GB. The results measured in seconds are depicted in Table 6. We can observe that on average, our algorithm is 1504.78, 119.3, and 151.19 times faster than PYP, PCIM, and EGH on average. In addition, we can also see that even in the case of 8 labels, the running time of the benchmark algorithms becomes extremely high (13, 6, and 20 minutes for PYP, PCIM, and EGH). In contrast, the running time of EGH–H still remains under 2 seconds. Given this, our algorithm needs significantly less computation time, while providing the highest accuracy, compared to the state–of–the–art benchmarks. This implies that our algorithm could be used for interactive feedback, where the agent suggests homeowners different home energy consumption plans in real–time, as it can use our algorithm to quickly predict the next–day usage, based on the real–time feedback of users.

5. CONCLUSIONS AND FUTURE WORK

Home ID	No. of appliances	PYP	PCIM	EGH	EGH–H
Home 1	8	779.95	371.3	1221.72	1.406
Home 3	3	240.64	8.8	1.05	0.067
Home 4	3	328.81	12.84	2.09	0.288
FE 22	6	494.97	43.769	2.6	0.97
FE 32	4	289.76	19.9	1.26	0.16
FE 39	4	255.28	18.897	0.886	0.172

Table 6: Running time on homes from REDD and FigureEnergy.

We investigated the problem of predicting the usage of electrical appliances in the home. To solve this problem, we proposed EGH–H, the first algorithm that addresses human behaviour prediction within the energy management domain by extending the EGH algorithm. In particular, our algorithm combines an inter–dependency model between the usage of appliances with the exploitation of the cyclic features of homeowners' everyday routine. We also demonstrated through extensive evaluations, using real–world data taken from the REDD and FigureEnergy datasets, that our algorithm outperforms state–of–the–art methods by up to 40% in prediction accuracy. As a result, our work could potentially form an efficient solution to real–world home energy management systems, where usage predictions are needed to optimally schedule the electrical consumption of the home.

Note that during the experiments on data from the FigureEnergy, all the algorithms (including ours) provides low performance. This is due to the uncertainty within the labelling process of homeowners. Since our current model does not take into account this source of uncertainty, it is not trivial that how our approach can be extended to such settings. Given this, we aim to further study prediction with noisy or uncertain labels as future work. In addition, we intend to improve the quality of prediction by allowing interactive feedback from users, where the agent can use these feedbacks to learn and refine its prediction in real–time, as the running time of our algorithm makes it suitable for such scenarios.

6. REFERENCES

[1] P. A. Aguilera, A. Fernández, R. Fernández, R. Rumí, and A. Salmerón. Bayesian networks in environmental modelling. *Environmental Modelling & Software*, 26(12):1376–1388, Dec. 2011.

[2] H. Bapierre, G. Groh, and S. Theiner. A variable order markov model approach for mobility prediction. *Pervasive Computing*, 2011.

[3] C. Cho, Y. Wu, S. Yen, Y. Zheng, and A. L. P. Chen. On-line rule matching for event prediction. *The VLDB Journal*, 20(3):303–334, July 2010.

[4] E. Costanza, S. D. Ramchurn, and N. R. Jennings. Understanding domestic energy consumption through interactive visualisation: a field study. *International Conference on Ubiquitous Computing*, 2012.

[5] Department of Energy & Climate Change. Smarter Grids : The Opportunity. (December), 2009.

[6] Department of Energy & Climate Change. UK National Strategy for Climate and Energy : Transition to a Low Carbon Society. Technical report, 2009.

[7] Department of Energy & Climate Change. Energy Consumption in the United Kingdom. Technical report, 2010.

[8] V. Didelez. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):245–264, Jan. 2008.

[9] N. Eagle and A. S. Pentland. Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7), 2009.

[10] H. Gao, J. Tang, and H. Liu. Exploring social-historical ties on location-based social networks. *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.

[11] M. C. González, C. A. Hidalgo, and A. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–82, June 2008.

[12] A. Gunawardana, C. Meek, and P. Xu. A Model for Temporal Dependencies in Event Streams. *NIPS*, 2011.

[13] D. Heckerman, D. M. Chickering, C. Meek, R. Rounthwaite, and C. Kadie. Dependency Networks for Inference, Collaborative Filtering, and Data Visualization. *Journal of Machine Learning Research*, 1:49–75, 2001.

[14] J. Z. Kolter and M. J. Johnson. REDD : A Public Data Set for Energy Disaggregation Research. In *In proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability*, number 1, pages 1–6, 2011.

[15] J. McInerney, A. Rogers, and N. R. Jennings. Improving location prediction services for new users with probabilistic latent semantic analysis. *4th International Workshop on Location-Based Social Networks*, 2012.

[16] S. D. Ramchurn, P. Vytelingum, A. Rogers, and N. Jennings. Agent-Based Control for Decentralised Demand Side Management in the Smart Grid. In *The Tenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, pages 5 – 12, 2011.

[17] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. T. Campbell. A spatio-temporal prediction framework for pervasive systems. *STAMI Workshop at IJCAI*, 6696:152–169, 2011.

[18] L. Srivatsan, P. Sastry, and K. Unnikrishnan. Discovering frequent episodes and learning hidden markov models: A formal connection. *IEEE Transactions on Knowledge and Data Engineering*, 17(11), 2012.

[19] S. Tominaga, M. Shimosaka, R. Fukui, and T. Sator. A variable order markov model approach for mobility prediction. *Pervasive Computing*, 7319, 2012.