<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *International Conference on Data Mining*.

Citation for the original published paper:

Nordahl, C., Persson, M., Grahn, H. (2017)

Detection of Residents' Abnormal Behaviour by Analysing Energy Consumption of Individual Households.

In: *Proceedings of the 17th IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 729-738).

<https://doi.org/10.1109/ICDMW.2017.101>

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:bth-15565>

Detection of Residents' Abnormal Behaviour by Analysing Energy Consumption of Individual Households

Christian Nordahl, Marie Persson and Håkan Grahn

Department of Computer Science and Engineering

Blekinge Institute of Technology, 371 79, Karlskrona, Sweden

Email: {christian.nordahl, marie.persson, hakan.grahn}@bth.se

Abstract—As average life expectancy continuously rises, assisting the elderly population with living independently is of great importance. Detecting abnormal behaviour of the elderly living at home is one way to assist the eldercare systems with the increase of the elderly population. In this study, we perform an initial investigation to identify abnormal behaviour of household residents using energy consumption data. We conduct an experiment in two parts, the first to identify a suitable prediction algorithm to model energy consumption behaviour, and the second to detect abnormal behaviour. This approach allows for an initial step for the elderly care that has a low cost, is easily deployable, and is non-intrusive.

I. INTRODUCTION

With an ever improving healthcare system, average life expectancy is continuously rising. By year 2020, projections state that approximately 20% of the world population will be age of 60 or older [1]. A higher life expectancy provides challenges to ageing adults, e.g. limitations in vision, physical activities, and cognitive decline. These limitations also become challenges for the healthcare and eldercare systems, both in terms of scale and economy. Assisting the elderly population with living independently at home is one way to approach these concerns.

As smart meters are being deployed world wide by electrical providers, there is an opportunity to provide a low cost approach to remote monitoring of residents that is non-intrusive. Smart meters are energy consumption meters that allows for high resolution energy consumption measurements, and the data can be gathered remotely. It has been shown that the use of electrical appliances directly relate to how the residents perform their activities of daily living (ADLs) [2]. We believe that this can also be shown in how the residents adhere to their normal behaviour of energy consumption.

We have collected energy consumption data from 17 different households in Sweden, provided by a Swedish electricity provider. The collected data is measured with an interval of 1 minute. We will investigate the possibility to detect abnormal behaviour of the resident by analysing the household energy consumption.

In this paper we make the following contributions:

- We model the normal energy consumption behaviour of 17 households by applying the principles of Very Short Term Load Forecasting (VSTLF).
- We investigate the possibility of detecting abnormal behaviour utilising the prediction models modelling the normal energy consumption behaviour.

To summarize, we investigate the possibility to use household energy consumption to detect abnormal behaviour of the residents. We believe this could be a possible first step for eldercare systems to monitor individuals remotely with low cost and no intrusion. As the next step following this study, we intend to collect data from a representative test group in collaboration with our local municipality and eldercare.

II. BACKGROUND

In this section, we first define two scenarios of abnormal behaviour that are of interest. We then describe the basics of regression based anomaly detection, followed up by energy consumption forecasting. We conclude this section with a review of previous work related to ours.

A. Abnormal Behaviour

We have identified two scenarios that are of significance to detect in this type of context; S1: increase of energy consumption during late evenings and nights, and S2: lack of energy consumption during mornings.

Scenario S1 is an indication of a phenomenon called *sundowning*, which often is associated with Alzheimer's

disease and in other forms of dementia [3]. When the sun is settling down, the subjects affected with sundowning gets more confused, restless, and agitated etc.

Scenario S2 represents the resident staying in bed longer than usual. This could be an indication of the resident being sick and thereby staying longer in bed.

B. Regression Based Anomaly Detection

The first step of applying regression to detect anomalies in time series data is to fit a model to the data. The model, modelling the normal behaviour, is then used to predict values. Predicted values are compared against the observed by calculating the residual. The residual is then used either directly by its magnitude, or with statistical tests to determine if it is an anomaly or not [4]

The anomalies found using the method above are point anomalies, i.e. single data points identified as abnormal. In our future studies we will investigate the use of contextual anomalies.

C. Forecasting Energy Consumption

Forecasting energy consumption has been approached in many different ways, including Support Vector Regression (SVR) [5], Linear Regression (LR) [6], and exponential smoothing [7]. The most popular being Neural Networks (NN), which Hippert et al. [8] have conducted a thorough survey on. In this context, when identifying abnormal behaviour of elderly people, we generally want to identify if something abnormal has happened as soon as possible. Therefore, we focus on approaches from Very Short Term Load Forecasting (VSTLF), a research area aimed at forecasting energy consumption minutes to hours ahead.

In VSTLF, the main type of features are previous energy consumptions during the same time on earlier days. These are chosen to capture the daily and weekly trends. It has been shown that fully capturing these trends is important for accurate prediction models [9], [10].

The average use of VSTLF is for electricity providers to estimate how much energy is needed in an area. Studies have mainly been focusing on energy consumption data from entire cities, regions, and countries. Energy consumption data on such a high level shows clear consumption trends. In this study, we apply the principles of VSTLF on individual households to model normal energy consumption behaviour. The energy consumption from individual households are more prone to differ from day to day. For example, simply arriving home a couple of hours late from work will alter the energy consumption behaviour for the majority of an evening.

D. Related Work

Detecting abnormal energy consumption behaviour has gained more interest with the addition of smart meters. Smart meters allow both residents to be more aware of their own energy consumption and electricity providers to have a better estimate on how much energy is needed in an area. Zhang et al. [11] analyses energy consumption data on a household level to identify days when the residents have gone on vacation. They compare the accuracy of regression, entropy, and clustering based anomaly detection, with the regression based providing the best results. Chou and Telaga [12] investigate the possibility of real time anomaly detection for smart meter energy consumption in an office building. Auto-Regressive Integrated Moving Average (ARIMA) and a hybrid NN and ARIMA model were used to model the normal energy consumption behaviour. The authors used the two-sigma rule to determine if the actual energy consumption was abnormal. Meaning, if the actual energy consumption was more than 2 standard deviations away from the predicted energy consumption, it was labeled as abnormal.

Smart homes allow for a vast amount of different sources to collect data from. Research in the Ambient Assisted Living (AAL) area investigate the use of ambient sensors and actuators to determine if a resident is behaving out of the ordinary. Streaming data from the sensors and actuators can be used directly to determine the residents normal behaviour and if any deviation has occurred [13]. Another possibility is to take the sensor data and map it into activities using Activity Recognition (AR). Identified activities are then used to determine how well they perform activities of daily living (ADLs). Sprint et al. [14] takes activities identified by AR and places them in time windows. These windows are compared to previous weeks' time windows to determine if a significant change has happened. The main drawback in these approaches is that smart homes are not that common yet. Installing sensors could also be seen as intrusive by the resident, and deploying this in a large scale takes a lot of effort.

Alcála et al. [15] focus on using energy consumption to determine if the residents behaviour is deviating from the norm. The authors use Non-Intrusive Monitor Loading (NILM), which takes a household's aggregated energy consumption and disaggregates it into individual appliances. They divide a day into time windows, and defines the probability of the appliances being used in these time windows using the Dempster-Shaffer theory.

To detect abnormal activities, they compare the current day's time window to the same time window on previous days and calculate the probability of the residents deviating from the normal behaviour. However, disaggregating energy consumption into individual appliances is a complex task. The authors made the assumption that the disaggregation has already been performed, using data that was already disaggregated.

Chalmers et al. [16] investigate the use of smart meters and smart plugs to detect abnormal behaviour. They propose a system that analyses smart meter data, in conjunction with health data, to determine if a resident is deviating from their normal behaviour. The authors also conduct a case study of an elderly resident with a smart meter and smart plugs. They show differences in energy consumption on different mornings, arguing that abnormal behaviour could be identified using smart meter data. In line with the work of Chalmers et al., we have collected data and have started an initial study to investigate the possibility to detect abnormal behaviour using energy consumption data.

III. DATA ANALYSIS

We use energy consumption data measured at 1 minute intervals from 17 different households in Sweden. Initially, we received data from 20 households, but 3 of them were discarded due to missing values. The data collection time spans from the 7th of February 2017 to the 1st of April 2017 and was provided by a Swedish electricity provider. In total for the 17 households, there are 1,311,720 measurements with 664 missing values (0.05%) remedied by applying linear interpolation.

Each household's average energy consumption and standard deviation is presented in Table I. The households average energy consumption range from 0.0186 kWh to 0.0588 kWh per minute with standard deviations from 0.0076 kWh to 0.0322 kWh.

Two weeks of energy consumption for four households are shown in Figure 1, where the differences between the households' energy consumption are easier to identify. We can observe that the households are distinctive, both in terms of the *amount* of consumed energy and *how* the energy is consumed. For example, house D has a dense usage pattern where the energy consumption is focused around 0.03 kWh per minute. On the other hand, house K consistently has a low consumption during mornings and high peaks during late afternoons and evenings.

In general, it is not easy to identify a daily or weekly pattern for the households due to frequent variations in

TABLE I
AVERAGE ENERGY CONSUMPTION PER MINUTE AND STANDARD DEVIATION FOR EACH HOUSEHOLD.

House	Mean (kWh)	STD (kWh)
A	0.031135	0.016534
B	0.018647	0.017853
C	0.035664	0.021904
D	0.030324	0.012230
E	0.036505	0.016897
F	0.056365	0.020543
G	0.019731	0.010121
H	0.051781	0.030274
I	0.058827	0.026903
J	0.037286	0.032240
K	0.029254	0.027309
L	0.027475	0.019119
M	0.038118	0.022269
N	0.036136	0.025799
O	0.038120	0.028139
P	0.043946	0.018218
Q	0.009349	0.007618

the data. This makes the identification of an energy consumption trend even harder. Therefore, we create three additional data sets where we decrease the frequency and aggregate the energy consumption. The additional data sets have the intervals 10, 30, and 60 minutes.

Figure 2 shows the energy consumption of house K in the four data sets over two weeks (February 13-27, 2017). Comparing the different data sets, we can observe that with a frequency decrease it is easier to identify a daily pattern of energy consumption. With a larger time window between measurements we allow for more flexibility of the household residents. This yields a larger margin of error for how the residents use their appliances. In Figure 2 we see that peaks and slumps of consumption occur at roughly similar times each day. For example, February 22 and 23 are nearly identical, only differing in the height of the peak during the evening.

As described in Section II-C, identification of daily and weekly trends is important for accurate energy consumption prediction. This section shows that there are energy consumption trends present in the data, albeit not that clear for all households.

IV. APPROACH

In this section we present our approach to detect abnormal behaviour. First, we decide on what features

Fig. 1. Energy consumption of four different households from February 13-27 2017.

Fig. 2. Energy consumption of household K in all four data sets from February 13-27 2017.

we will include in our different feature sets. We then choose three prediction algorithms to model normal behaviour. Finally, we describe our approach to detect anomalies.

A. Feature Selection

As described in Section III, we identified trends of energy consumption for our different households. Giving a good indication that using previous energy consumptions to predict future energy consumptions is a reasonable

approach. This is in line with the literature of VSTLF where previous energy consumptions are extensively used as features [9], [17].

To decide which features to choose, we calculate the autocorrelation coefficient (r_k) for all four data sets. The r_k coefficient gives a value of how much the energy consumption at time t correlates to the energy consumption k lags away. A lag in this context stands for the number of time periods between the measurements.

Fig. 3. Average autocorrelation of all households for each data set up to four weeks back in time. Each vertical line represents a 24 hour time span.

For example, in the one minute data set one lag away would be one minute, two lags two minutes and so on. The autocorrelation is calculated as follows:

$$r_k = \frac{\sum_{t=k+1}^n (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=k+1}^n (X_t - \bar{X})^2} \quad (1)$$

where X_t is the energy consumption at time t , $\bar{X}$ is the average energy consumption, and X_{t-k} is the energy consumption at k lags away from t . Values range from -1 to 1, where a value closer to either -1 or 1 shows a strong negative or a positive correlation respectively. Values closer to 0 indicate less or no correlation between the measurements.

In Figure 3 we plot the average household autocorrelation for all data sets. We can observe that with lower measurement frequencies, we have a clearer correlation between measurements. Comparing 10, 30, and 60 minutes data sets to the 1 minute data set, the peaks in the graph are more distinctive. Giving an indication that it is easier to discover a consumption trend of individual households with a lower measurement frequency.

The strongest correlation in all data sets is the energy consumption one lag before the current. This drops off quickly and starts to enter a daily pattern of correlation. Each data set has a recurring trend of correlation at exact days before the current energy consumption. For example, in the 60 minute data set we have peaks of correlation at lag 24, 48, 72, and so on. These peaks of correlation indicate that using the energy consumption

measurements at the same time from days before are suitable features.

To show the difference in correlation between the households, we plot the autocorrelation for two different households in Figure 4. We can observe a clear difference in how the energy is consumed between them. Household L follows the daily trend of energy consumption we have emphasized on. On the other hand, household B does not follow the daily trend but instead has a generally higher correlation for the whole week before the current energy consumption.

Based on the results from calculating the autocorrelation coefficient (r_k) we design six different feature sets. The first two feature sets, F_a and F_b , both contain energy consumption measurements right before the load being predicted. F_a consists of 5 measurements right before the predicted at times $t, \dots, t-4$, and F_b contains 10 measurements at times $t, \dots, t-9$. Both feature sets consist of the measurements that have the highest correlation to the energy consumption being predicted. They are chosen as baseline feature sets, that only takes into account the energy consumption close to the predicted energy consumption. No daily or weekly trends are incorporated into these two feature sets.

In feature sets F_c and F_d , we incorporate weekly behaviour by adding energy consumption from the week previous to the prediction day as additional features. F_c is a super set of F_a , with the addition of energy consumption at times $t+1, \dots, t-4$ from the same day

Fig. 4. Autocorrelation of two households. Each vertical line represents a 24 hour time span.

one week before. Note that the time $t+1$ was added from the previous week, as it has the highest correlation from that day to the value we want to predict. F_d has the same addition of features as F_c , only with more measurements, i.e. the energy consumption at times $t+1, \dots, t-9$ from the previous week.

For feature set F_e , we aim to solely focus on daily behaviour in order to compare the effects of daily to weekly trends. We therefore take energy consumption from the lags right before the energy consumption that we want to predict and from the six consecutive days before. As we have more days to collect features from in this feature set, we only choose three measurements from each day. The features chosen were energy consumption at times $t, t-1$ on the prediction day, and at times $t+1, t, t-1$ on the six consecutive days before.

Finally, for F_f we continue with the weekly behaviour as we did in feature sets F_c and F_d . F_f is a super set of F_c with added energy consumption of two weeks before at times $t+1, \dots, t-4$. All feature sets are summarized in Table II.

B. Prediction Algorithms

The first step to detecting abnormal behaviour is being able to model normal behaviour. Initially, we investigated how SVR, NNs, and LR perform in modelling the normal behaviour of all the 17 households. All three algorithms have shown good results for prediction in the VSTLF literature [5], [17]. However, initial experiments showed that NNs did not perform well with our data sets, therefore we discarded NNs from this study.

1) *Support Vector Regression*: SVR is an extension of the Support Vector Machine (SVM) that allows for prediction of continuous values [18]. SVMs have been used to address a wide variety of classification problems, including image analysis [19] and text categorisation

TABLE II
INFORMATION ABOUT THE FEATURE SETS EXTRACTED FROM OUR DATA SETS. X_t STANDS FOR THE ENERGY CONSUMPTION ON THE PREDICTION DAY AT TIME t AND XD_t^n FOR THE ENERGY CONSUMPTION n DAYS BEFORE THE PREDICTION DAY AT TIME t .

Feature set	Number of features	Variables
F_a	5	$X_t, \dots, X_{t-4}$
F_b	10	$X_t, \dots, X_{t-9}$
F_c	11	$F_a, XD_{t+1}^7, \dots, XD_{t-4}^7$
F_d	21	$F_b, XD_{t+1}^7, \dots, XD_{t-9}^7$
F_e	20	X_t, X_{t-1} and $XD_{t+1}^n, XD_t^n, XD_{t-1}^n$ for $1 \leq n \leq 6$
F_f	17	$F_c, XD_{t+1}^{14}, \dots, XD_{t-4}^{14}$

[20]. The SVM classifies instances by finding a hyperplane that separates two classes with the largest margin to any point within the training sets. A small subset of the data defines the decision boundary of the SVM, called support vectors. To create non-linear boundaries, the data can be projected into higher dimensions by using kernel functions. Both SVR and SVM rely on a small subset of the data to create their decision boundaries, making them tend to be resistant to overfitting the training data.

Initial tests showed that the linear and the radial basis kernel performed similarly well, which is why we choose both for our experiments. The linear kernel is configured with $C = 250$ and $\epsilon = 0.01$, and the radial basis function kernel with $C = 500$, $\epsilon = 0.005$, and $\gamma = 0.15$. All parameters were chosen empirically.

2) *Linear Regression*: LR is a statistical method used for forecasting. LR approximates the relationship between the independent variables and the dependent variable with a straight line, assuming there is a linear relationship between the two. The line that fits the data

the best is chosen by minimizing the sum of squared errors, using ordinary least squares method.

C. Anomaly Detection

We use the prediction algorithms presented in the previous subsection to detect anomalies. We calculate the residual between the actual energy consumption and the prediction models predicted energy consumption. The residual is then divided by the actual energy consumption, obtaining the relative distance between the two energy consumptions.

The relative distance, R , between the actual and predicted energy consumption is calculated as follows:

$$R = \frac{|F_t - A_t|}{A_t} \quad (2)$$

where F_t is the predicted energy consumption and A_t is the actual energy consumption at time t . If $R > \alpha$, where α is a user defined parameter acting as a threshold, the energy consumption is considered abnormal.

V. EXPERIMENTS

We divide our experiments into two parts; *A*: modelling normal behaviour, and *B*: identifying abnormal behaviour. In the first part, we determine the accuracy of the prediction algorithms when modelling the normal behaviour of the 17 households. In the second part, we investigate how different combinations of prediction algorithms, feature sets, and data sets detect abnormal behaviour. We use the implementations of each prediction algorithm from the Scikit-learn module [21].

A. Modelling Normal Behaviour

We perform a 5-fold cross validation to determine the performance of the prediction algorithms when modelling the normal behaviour. We use Mean Absolute Percentage Error (MAPE) to evaluate the prediction models which gives the accuracy of the predictions in percentage. It is defined as follows:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \times 100 \quad (3)$$

where A_t is the actual value at time t , F_t is the predicted value at time t , and n is the number of measurements.

B. Identifying Abnormal Behaviour

We randomly select three households from our data set to insert abnormal behaviour, households D, I, and K. The two scenarios defined in Section II-A are inserted in the final week of data separately, creating a data set for each scenario. Scenario S1, *sundowning*, we insert

by taking the peak of consumption during a randomly selected evening, in this case the 29th of March, and repeat it for 4 additional hours. The additional 4 hours are the ones considered to be the abnormal behaviour. For households D and I, we take the energy consumption between 19:00 and 21:00 and repeat it until 01:00 the next day. For household K we choose the consumption between 16:00 and 18:00 and repeat it until 22:00.

Likewise, we insert scenario S2 the same way on the same day. We take the energy consumption of household D between 02:00 and 04:00 and repeat it until 08:00. For households I and K we choose the consumption between 01:00 and 03:00 and repeat it until 07:00.

We train the prediction algorithms on all the data leading up to the final week, i.e. the data from the 7th of February to, and including, the 25th of April. The final weeks energy consumption is predicted and the method described in Section IV-C is used to determine if the energy consumption is normal or not. We then evaluate the anomaly detection by; 1) the detection rate (often referred as recall), which is the ratio between correctly classified anomalies and the number of anomalies, and 2) the false alarm rate, the ratio between normal data points classified as anomalous and the number of normal data points.

VI. RESULTS AND ANALYSIS

A. Cross Validation Results

The results from cross validating the 1, 10, 30, and 60 minute data sets are presented in Tables III, IV, V, and VI respectively.

Comparing the four data sets, we observe that the one minute data set is the most accurate in modelling the normal behaviour. The MAPE scores are more stable for all prediction algorithms compared to the other data sets, varying between 16.87-18.17% for LR, 29.06-31.52% for SVR with linear kernel, and 20.76-23.10% for SVR with the RBF kernel. The one minute data set contains more energy consumption observations compared to the other data sets, i.e. we have more observations to train the prediction algorithms with, making the result in line with expectations.

However, the second most accurate data set is the 60 minute data set, where we have the least amount of observations. In the 60 minute data set SVR with linear kernel has its best accuracy, with a MAPE score between 21.30-22.75%. LR has its second best prediction accuracy with a MAPE between 23.86-24.86%. The SVR with RBF kernel does however vary a lot in this data set,

TABLE III
MAPE SCORES FROM 5 FOLD CROSS VALIDATION FOR THE THREE
PREDICTION ALGORITHMS ON THE 1 MINUTE DATA SET.

Feature set	Linear Regression	SVR (Linear)	SVR (RBF)
F_a	17.19	29.66	21.17
F_b	16.87	29.06	20.76
F_c	17.73	30.56	22.38
F_d	17.36	29.84	21.64
F_e	17.93	31.52	23.10
F_f	18.17	30.17	22.77

TABLE IV
MAPE SCORES FROM 5 FOLD CROSS VALIDATION FOR THE THREE
PREDICTION ALGORITHMS ON THE 10 MINUTE DATA SET.

Feature set	Linear Regression	SVR (Linear)	SVR (RBF)
F_a	38.08	34.10	31.48
F_b	35.56	32.73	28.88
F_c	39.47	35.57	34.24
F_d	37.10	34.14	37.72
F_e	39.88	35.89	40.58
F_f	43.50	39.02	51.32

TABLE V
MAPE SCORES FROM 5 FOLD CROSS VALIDATION FOR THE THREE
PREDICTION ALGORITHMS ON THE 30 MINUTE DATA SET.

Feature set	Linear Regression	SVR (Linear)	SVR (RBF)
F_a	30.02	27.26	26.07
F_b	29.88	27.12	34.59
F_c	30.86	27.85	42.08
F_d	30.80	27.84	48.39
F_e	31.61	28.48	50.87
F_f	34.29	32.24	57.31

TABLE VI
MAPE SCORES FROM 5 FOLD CROSS VALIDATION FOR THE THREE
PREDICTION ALGORITHMS ON THE 60 MINUTE DATA SET.

Feature set	Linear Regression	SVR (Linear)	SVR (RBF)
F_a	24.26	21.75	28.81
F_b	23.86	21.30	44.20
F_c	24.35	22.05	42.33
F_d	24.23	21.79	37.36
F_e	24.86	22.74	40.51
F_f	24.17	22.25	39.82

with a MAPE between 28.81-42.33%, which is also its second best accuracy in the data sets.

The different feature sets do affect the accuracy of

modelling the normal behaviour. Overall, feature set F_b proves to be the most accurate for all data sets and prediction models, except for SVR with RBF kernel on the 30 and 60 minute data sets. Feature sets F_e and F_f seem to be alternating which is producing the worst results, depending on the data sets and the prediction algorithms. For example, in the one minute data set F_f provides the lowest accuracy for LR, while F_e is the lowest for SVR with linear kernel in the same data set.

We observe that SVR with RBF kernel has a varying prediction accuracy on our data. The accuracy on the one minute data set is among the best of all different configurations. However in the other 3 data sets, the accuracy is not consistent throughout the different feature sets, e.g. in the 30 minute data set where it has a MAPE of 26% on feature set F_a and a MAPE of 57.31% on feature set F_f . Both LR and SVR with linear kernel have consistent results on the 1 and 60 minute data sets, varying only 1-2%. In the 10 and 30 minute data sets, there is a larger variance in accuracy for both algorithms.

B. Abnormal Behaviour

Based on the results presented in the previous section, we exclude the 10 and 30 minute data sets and the SVR with RBF kernel. The accuracy is not enough to build an accurate model of the normal behaviour. The α parameter was set to 0.22 for the 1 minute data set and 0.27 for the 60 minute data set. The parameter values were chosen based on the MAPE scores from the previous section. The results for detecting the abnormal behaviour in the 1 minute data set are presented in Table VII and the 60 minute data set in Table VIII.

In the one minute data set, both LR and SVR have a poor detection rate for scenario S1. LR detecting 8-9% and SVR detecting 10-12% of the abnormal instances. However, between the two prediction algorithms we see a larger difference in the amount of normal instances classified as anomalies. LR has a false alarm rate of 24.62-26.12% while SVR has between 33.31-40.48%. For scenario S2 in the same data set, both algorithms have a higher detection rate. LR varying between 29.72-34.58% and SVR between 44.72-51.11% depending on the feature set.

In the 60 minute data set, the SVR remains at a similar false alarm rate for both scenarios, differentiating 0-4% between the different feature sets. However, we can observe a clear increase of false alarm rates for LR, varying between 40.29-48.47% for scenario S1 and 40.29-48.26% for scenario S2. The detection rate for scenario S1 is increased substantially for both algorithms

TABLE VII

DETECTION RESULTS FOR SVR LINEAR KERNEL AND LINEAR REGRESSION ON THE 1 MINUTE DATA SET. DR STANDS FOR DETECTION RATE, THE RATE OF IDENTIFIED ANOMALIES, AND FR STANDS FOR FALSE ALARM RATE, THE RATE OF NORMAL INSTANCES CLASSIFIED AS ABNORMAL.

Scenario	Feature set	Linear Regression		SVR (Linear)	
		DR (%)	FR (%)	DR (%)	FR (%)
S1	F_a	8.33	26.12	11.39	40.48
	F_b	8.33	24.62	12.22	39.93
	F_c	8.47	25.56	11.25	39.43
	F_d	8.47	24.13	11.11	38.24
	F_e	8.75	25.42	11.67	38.16
	F_f	9.44	25.92	10.00	33.31
S2	F_a	34.44	26.04	51.11	40.26
	F_b	31.39	24.54	48.33	39.70
	F_c	34.03	25.47	49.31	39.25
	F_d	29.72	24.04	47.08	38.05
	F_e	33.19	25.32	48.75	38.01
	F_f	34.58	25.77	44.72	33.08

TABLE VIII

DETECTION RESULTS FOR SVR LINEAR KERNEL AND LINEAR REGRESSION ON THE 60 MINUTE DATA SET. DR STANDS FOR DETECTION RATE, THE RATE OF IDENTIFIED ANOMALIES, AND FR STANDS FOR FALSE ALARM RATE, THE RATE OF NORMAL INSTANCES CLASSIFIED AS ABNORMAL.

Scenario	Feature set	Linear Regression		SVR (Linear)	
		DR (%)	FR (%)	DR (%)	FR (%)
S1	F_a	58.33	48.47	66.67	38.04
	F_b	58.33	46.01	66.67	35.38
	F_c	41.67	46.83	50.00	37.83
	F_d	41.67	44.99	41.67	34.97
	F_e	41.67	40.70	41.67	35.79
	F_f	41.67	40.29	41.67	34.56
S2	F_a	66.67	48.26	16.67	38.24
	F_b	58.33	45.81	33.33	35.38
	F_c	66.67	46.63	33.33	37.83
	F_d	58.33	44.38	41.67	35.17
	F_e	66.67	40.49	25.00	35.79
	F_f	50.00	40.29	41.67	33.54

in the 60 minute data set. For scenario S2, LR also has an increased detection rate while SVR decreases.

Generally in both data sets, feature set F_f provides the least amount of false alarms for both algorithms. The feature set achieving the highest detection rate varies more between the different configurations, but overall F_a achieves the highest detection rate.

VII. DISCUSSIONS

Results from the first part of the experiments indicate that applying VSTLF on individual household's energy consumption to model normal behaviour is a promising approach. We achieve a prediction accuracy that is satisfactory and allow for detection of abnormal energy consumption behaviour. For further studies, we assume that setting parameters for the prediction algorithms on a household level, or clusters of households, would further improve the prediction accuracy.

The second part of the experiment demonstrate the possibility to detect anomalous points of energy consumption. We assume that putting the point anomalies into context would allow us to draw more conclusions about the residents behaviour, especially with regards to health concerns. For example, a resident arriving at home one hour later than normal but follows the normal behaviour for the remainder of the evening. This may be flagged as abnormal behaviour, even though it is normal.

Initially, we expected the feature sets F_e and F_f to generate higher prediction accuracy and detection rate compared to the others. These feature sets capture more of the weekly and daily energy consumption behaviours. Instead, F_b , the feature set whose features are the energy consumptions right before the predicted, achieved the best prediction accuracy when modelling the normal behaviour. While F_f did provide the lowest false alarm rate for both data sets, the prediction accuracy when modelling the normal behaviour was generally the worst.

One possible reason to why F_e and F_f did not perform as expected is that both F_e and F_f have less data compared to F_b , in particular. F_e extracts features from the six previous days, causing it to have six days less of data instances compared to F_a and F_b . Likewise, F_f has features up to two weeks before. Having only seven weeks of data, this could be a reason why the performances of F_e and F_f are not in line with expectations.

The 1 minute and 60 minute data sets allowed for the best prediction accuracy for the prediction algorithms. LR had the best prediction accuracy of the entire experiment on the 1 minute data sets, and SVR had its best MAPE score on the 60 minute data set. We expected that the 60 minute data set would allow for better detection rates, which the results generally support. However, the false alarm rates for LR increases in the 60 minute data set compared to the 1 minute data set. The SVR did remain at similar levels for both data sets in terms of false alarm rate, but the detection rate was worse when detecting scenario S2 compared to the 1 minute data set.

VIII. CONCLUSIONS AND FUTURE WORK

The aim of this paper is to conduct an initial study on using energy consumption data gathered from smart meters to detect abnormal behaviour of residents. We apply the concepts of VSTLF to build a model of normal behaviour and predict future energy consumption. The results from the cross validation indicate that it is a promising approach to model the normal behaviour of the households.

Two scenarios indicating abnormal behaviour were defined and inserted into the data, and used for evaluating the prediction algorithms' ability to detect abnormal behaviour. Preliminary results indicate that it is possible to detect abnormal behaviour, but requires further study.

We have identified a number of directions for future work. First and foremost, we will collect data from a pilot study of elderlies in collaboration with our local municipality and eldercare. We will also investigate the addition of water consumption and weather data to further improve accuracy of modelling normal behaviour. We also plan to investigate the use of contextual anomalies instead of point anomalies to be able to draw more conclusions about the residents behaviour.

ACKNOWLEDGEMENTS

We would like to thank IIOX and EON for their participation in this study and for providing the energy consumption data. This work is part of the research project "Scalable resource-efficient systems for big data analytics" funded by the Knowledge Foundation (grant: 20140032) in Sweden.

REFERENCES

- [1] U. Nations, "World population ageing: 1950-2050," *New York: Department of Economic and Social Affairs*, 2002.
- [2] X. Zhang, T. Kato, and T. Matsuyama, "Learning a context-aware personal model of appliance usage patterns in smart home," in *Innovative Smart Grid Technologies-Asia (ISGT Asia)*, 2014 IEEE, pp. 73–78, IEEE, 2014.
- [3] L. K. Evans, "Sundown syndrome in institutionalized elderly," *Journal of the American Geriatrics Society*, vol. 35, no. 2, pp. 101–108, 1987.
- [4] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.
- [5] A. Setiawan, I. Koprinska, and V. G. Agelidis, "Very short-term electricity load demand forecasting using support vector regression," in *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, pp. 2888–2894, IEEE, 2009.
- [6] T. Hong, M. Gui, M. E. Baran, and H. L. Willis, "Modeling and forecasting hourly electric load by multiple linear regression with interactions," in *Power and Energy Society General Meeting, 2010 IEEE*, pp. 1–8, IEEE, 2010.
- [7] J. W. Taylor, "An evaluation of methods for very short-term load forecasting using minute-by-minute british data," *International Journal of Forecasting*, vol. 24, no. 4, pp. 645–658, 2008.
- [8] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on power systems*, vol. 16, no. 1, pp. 44–55, 2001.
- [9] I. Koprinska, M. Rana, and V. G. Agelidis, "Correlation and instance based feature selection for electricity load forecasting," *Knowledge-Based Systems*, vol. 82, pp. 29–40, 2015.
- [10] I. Drezga and S. Rahman, "Input variable selection for ann-based short-term load forecasting," *IEEE Transactions on Power Systems*, vol. 13, no. 4, pp. 1238–1244, 1998.
- [11] Y. Zhang, W. Chen, and J. Black, "Anomaly detection in premise energy consumption data," in *Power and energy society general meeting, 2011 IEEE*, pp. 1–8, IEEE, 2011.
- [12] J.-S. Chou and A. S. Telaga, "Real-time detection of anomalous power consumption," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 400–411, 2014.
- [13] T. Rodner and L. Litz, "Data-driven generation of rule-based behavior models for an ambient assisted living system," in *Consumer Electronics, Berlin (ICCE-Berlin), 2013. ICCEBerlin 2013. IEEE Third International Conference on*, pp. 35–38, IEEE, 2013.
- [14] G. Sprint, D. Cook, R. Fritz, and M. Schmitter-Edgecombe, "Detecting health and behavior change by analyzing smart home sensor data," in *Smart Computing (SMARTCOMP), 2016 IEEE International Conference on*, pp. 1–3, IEEE, 2016.
- [15] J. M. Alcalá, J. Ureña, Á. Hernández, and D. Gualda, "Assessing human activity in elderly people using non-intrusive load monitoring," *Sensors*, vol. 17, no. 2, p. 351, 2017.
- [16] C. Chalmers, W. Hurst, M. Mackay, and P. Fergus, "Smart health monitoring using the advance metering infrastructure," in *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on*, pp. 2297–2302, IEEE, 2015.
- [17] W. Charytoniuk and M.-S. Chen, "Very short-term load forecasting using artificial neural networks," *IEEE transactions on Power Systems*, vol. 15, no. 1, pp. 263–268, 2000.
- [18] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in neural information processing systems*, pp. 155–161, 1997.
- [19] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE transactions on Neural Networks*, vol. 10, no. 5, pp. 1055–1064, 1999.
- [20] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," *Machine learning: ECML-98*, pp. 137–142, 1998.
- [21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.