# Classification of Household Devices by Electricity Usage Profiles

Jason Lines[1], Anthony Bagnall[1], Patrick Caiger-Smith[2], and Simon Anderson[2]

[1] School of Computing Sciences
University of East Anglia
Norwich
UK
{j.lines,anthony.bagnall}@uea.ac.uk
http://www.uea.ac.uk/cmp
[2] Green Energy Options
Hardwick
Cambridge
UK
{paddy,simon}@greenenergyoptions.co.uk
http://www.greenenergyoptions.co.uk

**Abstract.** This paper investigates how to classify household items such as televisions, kettles and refrigerators based only on their electricity usage profile every 15 minutes over a fixed interval of time. We address this time series classification problem through deriving a set of features that characterise the pattern of usage and the amount of power used when a device is on. We evaluate a wide range of classifiers on both the raw data and the derived feature set using both a daily and weekly usage profile and demonstrate that whilst some devices can be identified with a high degree of accuracy, others are very hard to disambiguate with this granularity of data.

**Keywords:** time series classification, electricity device classification

## 1 Introduction

This paper investigates how to classify household items such as televisions, kettles and refrigerators based only on their electricity usage profile over a fixed interval of time. This research is part of a wider project investigating data mining electricity usage patterns generated by 'smart meters'. Smart meters record and transmit data on electricity usage of a whole home, a specific circuit or even an individual plug.

This project is supported by Cambridge based company Green Energy Options (GEO), who have developed a range smart metering devices. GEO have conducted a preliminary trial of the technology. Their monitoring devices were installed in 187 homes across East Anglia and the usage of individual devices and total household power consumption was recorded at 15 minute intervals for approximately a year in each household. The resulting data is described in Section 3.

One of the key components of the UK's strategy to reduce carbon emissions is the national plan to roll out smart metering devices to the 27 million homes in the country within the next decade. The cost of this program has been estimated to be 10 billion [8].

This cost is justified by the commonly cited statistic that smart meters reduce electricity usage by around 2.5% [2]. This source also states an example where a reduction of 20% is recorded. If this is accurate, smart meters offer a cost effective way of significantly reducing carbon emissions. However, this oft cited statistic has little basis in data and very little is known about the actual effect of smart meters and whether any observed initial reduction can be sustained in the medium or long term. Clearly, the act of collecting power usage data is in itself unlikely to modify long term behaviour; all smart meters are required to have an in-home display that describes usage. Very little is known about how people will react to smart meters and how best to use their output to encourage reduced consumption without a detrimental effect on a household's lifestyle. The success of a smart meter in altering consumer behaviour will thus be strongly influenced by:

1. what information can be extracted from the usage data;
2. how this information is presented to best inform the consumer; and
3. whether the consumer can be encouraged to interact with the device in order to act on this information.

GEO have included a range of features in their devices to help achieve these goals. Whilst our primary concern is how to extract knowledge from the data collected from smart meters, the nature of the models we form from the data are influenced by the second and third factors and thus ultimately help GEO provide the consumer with useful information. For example, one of the secondary uses of a smart meter could be to notify the consumer when a monitored device is malfunctioning or using more power than necessary. This offers the potential for saving the consumer money through reduced consumption and is a good way of demonstrating the utility of the device. A prerequisite for identifying faulty or inefficient behaviour is the classification of the type of device being monitored and a description of normal/efficient behaviour. Whilst it is possible to require the consumer to manually identify every device monitored, it is thought that this level of engagement will be hard to achieve. It is far more consumer friendly to be able to automatically identify a device through its usage profile. Hence we consider the time series classification problem of identifying device type through a daily or weekly demand profile. The main contribution of this paper is to define a new time series classification problem and to evaluate a range of strategies for best solving it.

To our knowledge this problem has not been addressed before. The rest of this paper is structured as follows. Section 2 provides background into time series classification and an overview of the strategies we have adopted for this problem.

Section 3 details the trial data used to form the classification problem and the preprocessing steps required. Section 4 presents the results of our experiments. Finally, we conclude with Section 5.

## 2 Time Series Classification

Suppose we have a set of $n$ time series, $T = \{\mathbf{t_1}, \ldots, \mathbf{t_n}\}$ where each series has $m$ ordered real valued observations $\mathbf{t_i} = \{t_{i1}, \ldots, t_{im}\}$ and a class label $c_i$ (note for simplicity we assume the series are of equal length, but this is not a requirement). Time series classification involves finding a function from the space of possible time series to the space of possible class labels. This differs from traditional classification problems in that the discriminating factors are assumed to be primarily embedded within the auto-correlation structure. All time series data mining relies to some degree on a measure of similarity between series. There are essentially three types of similarity: similarity in time (correlation based); similarity in shape (shape based) and similarity in change (autocorrelation based). A fuller discussion can be found in the literature [4, 7]. Section 3 details the trial data used to form the classification and the preprocessing steps required. There are a variety of approaches to time series classification, which can be summarised as follows:

**Ignore the time element.** If the series are of equal length and interval, it is possible to simply ignore the ordering and treat the problem as a traditional classification problem. This approach puts the onus on the classifier to model the interdependency between the attributes. It is potentially useful when attempting to classify based on correlation, but shape based similarity will be hard to detect and autocorrelation similarity impossible. One problem with this approach is that time series tend to have many features, hence some form of attribute selection or more usually feature creation is often employed.

**Specialised similarity measure.** Recent data mining research has focused on using specialised similarity measures such as dynamic time warping (DTW) [5] in conjunction with lazy classifiers [4] to capture both correlation based and shape based similarity. DTW is is a natural generalisation of using Euclidean distance based methods and is often seen as a means of compensating against slight phase shift rather than capturing phase independent similarity.

**Extract bespoke features.** The most common approach in the machine learning literature is to derive a set of summary features prior to classification (for example, see [10]). This can include time independent summary measures such as mean, variance, kurtosis and skewness and series characterisations such as slope and runs measures. Clearly, the nature of similarity captured is dependent on the features extracted.

**Transform into a different feature space.** An alternative approach seen in both the machine learning and data mining literature is to use transformations such as Spectral transforms, Autocorrelation function or Wavelets and classify in the transformed space. The aim of such as transformation is either dimensionality reduction to better approximate Euclidean distance [6] or to allow for the detection of shape based or change based similarity [3].

**Construct a model** The final commonly used approach is to construct a generative model of each series such as an autoregressive moving average (ARMA) model or hidden Markov model (HMM) and then to use the model parameters as features for classification [1, 9]. The generative model based schemes are best suited for detecting similarity in change and are hence the the least used approach, since most problems used in research are more suited to similarity in shape.

Clearly the approaches can be mixed. The main distinction is whether to preprocess the data to capture the different types of similarity or to embed the method within the classification algorithm, This is analogous to the difference between a filter and a wrapper approach to feature selection/creation. In this paper we concentrate on bespoke feature extraction.

## 3 The Data

The trial involved measuring the power consumption of 187 households for a variety of devices as identified by the participants. We extracted data on the ten most commonly identified devices: immersion heater; washing machine; fridge; freezer; fridge/freezer; oven/cooker; computer; television; and dishwasher. We created two classification problems: For the first set a case consisted of the daily measurements of the specified device (96 attributes), for the second set we used a week of readings (672 attributes). After data cleansing and validation, the daily set has 78,869 cases, the weekly set 9,215 cases. Figure 1 gives some examples of the resulting demand profiles.

This problem has several confounding factors that will make classification difficult. Firstly the fact that measurements are summed over 15 minutes makes it harder to detect devices that peak over a short period. For example, a kettle will consume a large amount of power whilst on, but will only be on for a two or three minutes; when summed over 15 minutes it will be harder to distinguish from a device such as a dish washer or washing machine which consume lower power but will be on for the whole period. Secondly, there will be a seasonal variation in the use of devices such as immersion heaters. Thirdly, we would also expect it to be hard to distinguish between similar devices such as a fridge and a fridge/freezer and finally, we would expect considerable variation between different devices of the same class.

Since our objective is to be able to identify a device for a new user with no labelled usage history we need to design our experiment to avoid a potential bias into our experiments. It will surely be easier to identify a device for a

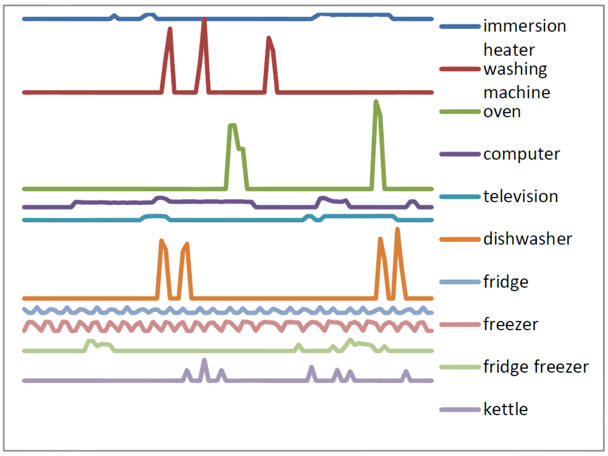


Fig. 1: Examples of daily profiles for the ten devices considered.

single household than across all households, thus if measurements from a single household appear in both the training and testing sets, our accuracy estimates are liable to be over optimistic for unseen households. Hence we design all cross validation experiments so that the test and train splits are always composed of different households.

An examination of Figure 1 highlights the nature of the similarity measures this problem will require and hence the transformations we consider. Firstly, when a device is on and the level of power used are clearly important. Hence our first approach is to simply use the raw data. However, it seems unlikely that a correlation based approach will be sufficient, given the variability of usage pattern within each class. Our second approach is to derive a set of features describing the distribution of power used when a device is on and the distribution of length of time on. Table 1 lists these features and presents the summary statistics averaged over all devices of each class of device. The mean values in Table 1 shows that there are clusters of power consumption over a 15 minute interval, which can be characterised as low (computer, freezer, fridge, fridge, TV and kettle) and high (dishwasher, immersion heater, oven and washing machine). There is also a wide variation between classes in the duration of usage; the average time on for computers is approximately 7 hours, for cookers 42 minutes and for kettles 16 minutes (skewed because the minimum on-time is 15 minutes). This indicates that these statistics may be useful in constructing classifiers.

## 4 Results

Table 2 gives the accuracy results for a ten fold cross validation (where no single household appears in both the training and testing fold) using five different classifiers on the daily and weekly data sets. For the daily data, we observe

Table 1: Summary statistics for the daily data set. The proportion of the data in each class is given in brackets. Each data is average over all cases of the given class. So, for example, the minimum power usage in any one 15 minute period for a computer is 26.35 Wh when we average across all computers, assuming any power is being used at all.

| | Computer | Dish Washer | Freezer | Fridge | Fridge/ Freezer | Immersion Heater | Kettle | Oven/ Cooker | Television | Washing Machine |
|---|---|---|---|---|---|---|---|---|---|---|
| Summary statistics for power usage when a device is in use (Wh) | | | | | | | | | | |
| min | 26.35 | 276.01 | 17.61 | 12.34 | 19.07 | 151.07 | 61.44 | 252.05 | 29.65 | 274.96 |
| max | 39.79 | 457.92 | 37.34 | 27.07 | 45.42 | 245.49 | 113.98 | 423.94 | 45.24 | 375.88 |
| mean | 33.13 | 365.13 | 27.14 | 20.10 | 28.91 | 201.80 | 84.94 | 328.79 | 39.36 | 324.14 |
| std dev | 6.42 | 102.69 | 8.11 | 4.55 | 7.28 | 75.65 | 27.10 | 114.63 | 11.67 | 81.13 |
| skewness | 0.07 | 0.03 | -0.22 | -0.55 | 0.45 | 0.11 | 0.17 | 0.26 | -1.08 | 0.03 |
| kurtosis | 2.62 | -1.44 | 1.35 | 0.64 | 5.05 | 0.53 | -0.94 | -0.73 | 3.42 | -1.04 |
| Summary stats of device usage tendancies | | | | | | | | | | |
| percentage of time on | 0.40 | 0.04 | 0.47 | 0.39 | 0.45 | 0.20 | 0.05 | 0.06 | 0.25 | 0.03 |
| first time on | 33.73 | 51.57 | 1.69 | 1.83 | 5.87 | 28.88 | 32.34 | 61.15 | 45.09 | 45.78 |
| Summary stats of the number of time steps a device is on for | | | | | | | | | | |
| nos runs | 3.03 | 2.13 | 23.18 | 17.91 | 14.27 | 6.55 | 4.55 | 1.94 | 2.71 | 1.70 |
| run min | 28.00 | 1.61 | 5.99 | 2.30 | 5.86 | 3.58 | 1.02 | 2.45 | 8.04 | 1.32 |
| run max | 33.68 | 2.17 | 9.60 | 5.90 | 13.17 | 7.99 | 1.27 | 3.36 | 16.80 | 1.59 |
| run mean | 30.54 | 1.88 | 6.97 | 3.55 | 8.73 | 5.43 | 1.07 | 2.84 | 11.82 | 1.44 |

Table 2: Classification accuracy (and standard deviation) for a ten fold cross validation on the daily and weekly data sets.

| **Daily data set** | | | |
|---|---|---|---|
| | **Naive Bayes** | **C4.5** | **Random Forest** |
| Raw Data | 38.40% (4.69) | 56.60% (4.17) | 61.34% (4.67) |
| Derived Features | 44.01% (4.74) | 58.89% (5.01) | 59.04% (4.11) |
| | **SMO (SVM)** | **NN ($k = 21$)** | **NN ($k = 51$)** |
| Raw Data | 43.52% (5.38) | 56.72% (4.85) | 53.82% (5.36) |
| Derived Features | 59.46% (4.24) | 60.86% (5.82) | 60.95% (6.3) |
| **Weekly data set** | | | |
| | **Naive Bayes** | **C4.5** | **Random Forest** |
| Raw Data | 41.43% (3.16) | 46.42% (4.46) | 55.81% (6.27) |
| Derived Features | 44.80% (6.44) | 62.90% (4.62) | 64.81% (5.5) |
| | **SMO (SVM)** | **NN ($k = 21$)** | **NN ($k = 51$)** |
| Raw Data | 48.79% (6.81) | 31.50% (6.89) | 26.48% (6.51) |
| Derived Features | 54.40% (5.73) | 63.25% (6.44) | 63.17% (6.28) |

that using the derived features improves the performance of all the classifiers except random forest, but that this improvement is small and the best overall performance is with random forest on the raw data. This suggests that the inbuilt ensemble mechanism of the random forest classifier is at least as good at capturing the inherent similarity as our derived features. Further experiments (not reported here) showed that there was also no improvement with dynamic time warping and FFT derived feature sets. However, for multi class problems such as this accuracy tends not to tell the whole story. Tables 3 and 4 show the confusion matrices for the random forest classifier on the daily raw and derived data sets. These tables demonstrate that the confusion for the raw data seems to be more widely distributed between all the classes, whereas on the derived feature set random forest is more systematic in it's mistakes.

Table 3: Confusion matrix for Random Forest on the daily raw data

| | **a** | **b** | **c** | **d** | **e** | **f** | **g** | **h** | **i** | **j** |
|---|---|---|---|---|---|---|---|---|---|---|
| a = computer | 4471 | 36 | 520 | 126 | 39 | 280 | 69 | 273 | 3649 | 156 |
| b = dishwasher | 174 | 5021 | 5 | 5 | 2 | 121 | 16 | 928 | 28 | 320 |
| c = freezer | 596 | 3 | 4060 | 2016 | 318 | 12 | 56 | 33 | 173 | 12 |
| d = fridge | 108 | 9 | 2403 | 3801 | 385 | 4 | 3 | 9 | 35 | 7 |
| e = fridgeFreezer | 93 | 2 | 1177 | 755 | 49 | 2 | 0 | 13 | 122 | 0 |
| f = immersionHeater | 631 | 263 | 188 | 98 | 18 | 520 | 583 | 409 | 543 | 146 |
| g = kettle | 28 | 26 | 36 | 7 | 0 | 72 | 8000 | 668 | 40 | 145 |
| h = ovenCooker | 564 | 600 | 268 | 80 | 29 | 139 | 1066 | 8166 | 632 | 305 |
| i = television | 4547 | 81 | 206 | 167 | 167 | 310 | 82 | 504 | 9392 | 284 |
| j = washingMachine | 66 | 358 | 11 | 13 | 0 | 30 | 323 | 547 | 118 | 4897 |

Table 4: Confusion matrix for Random Forest on the daily derived features

| | **a** | **b** | **c** | **d** | **e** | **f** | **g** | **h** | **i** | **j** |
|---|---|---|---|---|---|---|---|---|---|---|
| a = computer | 2912 | 0 | 388 | 20 | 76 | 332 | 173 | 90 | 2825 | 2 |
| b = dishwasher | 0 | 3785 | 1 | 0 | 0 | 52 | 0 | 1414 | 6 | 932 |
| c = freezer | 306 | 8 | 3704 | 2106 | 369 | 12 | 154 | 6 | 65 | 1 |
| d = fridge | 66 | 0 | 1952 | 3717 | 486 | 1 | 2 | 0 | 222 | 0 |
| e = fridgeFreezer | 272 | 0 | 447 | 907 | 166 | 5 | 126 | 0 | 96 | 0 |
| f = immersionHeater | 734 | 169 | 10 | 3 | 16 | 814 | 379 | 481 | 211 | 103 |
| g = kettle | 78 | 0 | 205 | 5 | 13 | 145 | 7954 | 72 | 97 | 32 |
| h = ovenCooker | 52 | 1100 | 5 | 0 | 0 | 486 | 71 | 6158 | 37 | 1802 |
| i = television | 2950 | 12 | 41 | 125 | 20 | 168 | 159 | 103 | 7789 | 36 |
| j = washingMachine | 0 | 1217 | 0 | 0 | 0 | 30 | 103 | 1990 | 40 | 2230 |

The results for the weekly data set are more clear cut, in that the classifiers trained on the derived features clearly outperform those trained on the raw data. The random forest confusion matrices given in Table 5 and 6 further demonstrate the improved performance.

The largest source of confusion is the expected problem of distinguishing between fridge, freezer and fridge freezer. Computer and television are also often confused. Table 7 shows the accuracy on the daily data set when we merge the classes fridge, freezer and fridge freezer (cold group) and computer and television

Table 5: Confusion matrix for Random Forest on the weekly raw data

| | **a** | **b** | **c** | **d** | **e** | **f** | **g** | **h** | **i** | **j** |
|---|---|---|---|---|---|---|---|---|---|---|
| a = computer | 360 | 13 | 85 | 22 | 2 | 31 | 5 | 24 | 437 | 26 |
| b = dishwasher | 14 | 389 | 0 | 6 | 0 | 25 | 25 | 243 | 27 | 160 |
| c = freezer | 49 | 1 | 409 | 177 | 34 | 1 | 10 | 4 | 32 | 2 |
| d = fridge | 39 | 4 | 245 | 352 | 17 | 1 | 3 | 3 | 11 | 0 |
| e = fridgeFreezer | 9 | 1 | 98 | 78 | 6 | 0 | 0 | 1 | 16 | 0 |
| f = immersionHeater | 48 | 61 | 21 | 21 | 2 | 35 | 76 | 39 | 63 | 32 |
| g = kettle | 4 | 27 | 1 | 0 | 0 | 14 | 733 | 69 | 20 | 113 |
| h = ovenCooker | 48 | 94 | 40 | 3 | 2 | 8 | 56 | 1169 | 87 | 79 |
| i = television | 365 | 56 | 41 | 35 | 15 | 23 | 19 | 58 | 985 | 42 |
| j = washingMachine | 7 | 124 | 0 | 0 | 0 | 7 | 99 | 146 | 26 | 705 |

Table 6: Confusion matrix for Random Forest on the weekly derived features

| | **a** | **b** | **c** | **d** | **e** | **f** | **g** | **h** | **i** | **j** |
|---|---|---|---|---|---|---|---|---|---|---|
| a = computer | 382 | 0 | 45 | 0 | 3 | 32 | 18 | 1 | 358 | 1 |
| b = dishwasher | 0 | 520 | 2 | 0 | 0 | 2 | 0 | 125 | 3 | 220 |
| c = freezer | 23 | 1 | 375 | 184 | 62 | 2 | 24 | 1 | 3 | 0 |
| d = fridge | 2 | 0 | 201 | 358 | 69 | 1 | 0 | 0 | 14 | 0 |
| e = fridgeFreezer | 24 | 0 | 72 | 70 | 22 | 7 | 0 | 0 | 1 | 0 |
| f = immersionHeater | 86 | 18 | 1 | 1 | 5 | 71 | 53 | 48 | 34 | 5 |
| g = kettle | 8 | 0 | 24 | 1 | 2 | 17 | 886 | 3 | 4 | 5 |
| h = ovenCooker | 9 | 61 | 1 | 0 | 4 | 69 | 1 | 1233 | 14 | 150 |
| i = television | 271 | 25 | 4 | 14 | 1 | 32 | 2 | 17 | 952 | 4 |
| j = washingMachine | 2 | 207 | 0 | 0 | 0 | 0 | 12 | 178 | 4 | 668 |

into screen group. Unsurprisingly, the accuracy is much higher. For the daily data, we again observe that the derived features improve accuracy across all reported classifiers except random forest, which again recorded the best accuracy using the raw data. Tables 8 and 9 show the confusion matrices for the random forest classifier on the daily raw and derived data sets. These tables demonstrate that confusion has been significantly reduced when compared to using the full set of classes, and the difference between the confusion of raw and derived features has also been reduced.

The results of the weekly data are again much more clear cut, with derived features clearly outperforming raw data. Tables 10 and 11 show the confusion matrices for the random forest classifier on the raw and derived feature data. They demonstrate a pattern similar to the first round of experiments, where the confusion for the raw data appears to be more widely distributed than the derived features.

Table 7: Classification accuracy using cold and screen groups

| **Weekly data set** | | | | |
|---|---|---|---|---|
| | **C4.5** | **Random Forest** | **SMO (SVM)** | **NN ($k = 21$)** |
| Raw Data | 77.09% (3.92) | 81.27% (4.62) | 56.47% (7.18) | 74.08% (5.20) |
| Derived Features | 78.25% (4.18) | 77.38% (3.69) | 74.90% (4.19) | 78.56% (4.20) |
| **Weekly data set** | | | | |
| | **C4.5** | **Random Forest** | **SMO (SVM)** | **NN ($k = 21$)** |
| Raw Data | 65.36% (4.52) | 72.96% (5.56) | 60.53% (9.10) | 39.89% (6.40) |
| Derived Features | 78.19% (4.11) | 80.32% (4.69) | 69.02% (3.57) | 78.03% (4.95) |

Table 8: Confusion matrix for Random Forest on the daily raw data with cold and screen groups

| | **a** | **b** | **c** | **d** | **e** | **f** | **g** |
|---|---|---|---|---|---|---|---|
| a = tvGroup | 22614 | 88 | 1174 | 443 | 127 | 554 | 359 |
| b = dishwasher | 244 | 5029 | 19 | 100 | 14 | 888 | 326 |
| c = coldGroup | 1033 | 8 | 15096 | 8 | 59 | 36 | 16 |
| d = immersionHeater | 1347 | 215 | 309 | 425 | 594 | 355 | 154 |
| e = kettle | 105 | 23 | 44 | 73 | 7980 | 661 | 136 |
| f = ovenCooker | 1443 | 605 | 262 | 116 | 988 | 8117 | 318 |
| g = washingMachine | 257 | 339 | 34 | 42 | 323 | 535 | 4833 |

Table 9: Confusion matrix for Random Forest on the daily features with cold and screen groups

| | **a** | **b** | **c** | **d** | **e** | **f** | **g** |
|---|---|---|---|---|---|---|---|
| a = tvGroup | 16746 | 10 | 625 | 370 | 237 | 192 | 41 |
| b = dishwasher | 6 | 3781 | 1 | 48 | 0 | 1442 | 912 |
| c = coldGroup | 1125 | 8 | 13900 | 13 | 145 | 4 | 1 |
| d = immersionHeater | 961 | 170 | 50 | 705 | 365 | 560 | 109 |
| e = kettle | 232 | 0 | 215 | 136 | 7913 | 73 | 32 |
| f = ovenCooker | 70 | 1137 | 2 | 509 | 69 | 6148 | 1776 |
| g = washingMachine | 50 | 1195 | 1 | 28 | 93 | 2018 | 2225 |

Table 10: Confusion matrix for Random Forest on the weekly raw data with screen and cold groups

| | **a** | **b** | **c** | **d** | **e** | **f** | **g** |
|---|---|---|---|---|---|---|---|
| a = screenGroup | 22614 | 88 | 1174 | 443 | 127 | 554 | 359 |
| b = dishwasher | 244 | 5029 | 19 | 100 | 14 | 888 | 326 |
| c = coldGroup | 1033 | 8 | 15096 | 8 | 59 | 36 | 16 |
| d = immersionHeater | 1347 | 215 | 309 | 425 | 594 | 355 | 154 |
| e = kettle | 105 | 23 | 44 | 73 | 7980 | 661 | 136 |
| f = ovenCooker | 1443 | 605 | 262 | 116 | 988 | 8117 | 318 |
| g = washingMachine | 257 | 339 | 34 | 42 | 323 | 535 | 4833 |

Table 11: Confusion matrix for Random Forest on the weekly features with screen and cold groups

| | **a** | **b** | **c** | **d** | **e** | **f** | **g** |
|---|---|---|---|---|---|---|---|
| a = screenGroup | 16746 | 10 | 625 | 370 | 237 | 192 | 41 |
| b = dishwasher | 6 | 3781 | 1 | 48 | 0 | 1442 | 912 |
| c = coldGroup | 1125 | 8 | 13900 | 13 | 145 | 4 | 1 |
| d = immersionHeater | 961 | 170 | 50 | 705 | 365 | 560 | 109 |
| e = kettle | 232 | 0 | 215 | 136 | 7913 | 73 | 32 |
| f = ovenCooker | 70 | 1137 | 2 | 509 | 69 | 6148 | 1776 |
| g = washingMachine | 50 | 1195 | 1 | 28 | 93 | 2018 | 2225 |

## 5 Conclusions and Future Work

In this paper we have proposed the time series classification problem of classifying household goods based solely on the electricity usage of the device as measured by a GEO smart meter. The ability to automatically detect the type of a device gives insights into the breakdown of the household usage pattern and offers the potential for providing useful feedback to the consumer, both in terms of minimizing their usage and in fault detection. We have assessed alternative classifiers and transformation for this problem and conclude that with a weekly profile we can accurately discriminate between classes of device by deriving a set of descriptive features and using a random forest or nearest neighbour classifier.

Data mining of smart meter data is going to be crucial in order to get the best value out of the massive investment required for the national role out program. This problem represents just one potential secondary use of the data. We may be able to achieve improved classification performance through consideration of more complex transformations and ensemble classifiers.

## References

1. A. J. Bagnall and G. J. Janacek, *Clustering time series from arma models with clipped data*, 10th International Conference on Knowledge Discovery in Data and Data Mining (ACM SIGKDD 2004), 2004, pp. 49–58.
2. S. Darby, *Making it obvious: designing feedback into energy consumption*, pp. 685–696, 2001.
3. G. J. Janacek, A. J. Bagnall, and M. Powell, *A likelihood ratio distance measure for the similarity between the fourier transform of time series*, Proceedings of the Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2005, 2005.
4. E. Keogh and S. Kasetty, *On the need for time series data mining benchmarks: A survey and empirical demonstration*, Data Mining and Knowledge Discovery **7** (2003), no. 4, 349–371.
5. E. Keogh and M. Pazzani, *Scaling up dynamic time warping for datamining applications*, Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '00, 2000, pp. 285–289.
6. F. Mörchen, *Time series feature extraction for data mining using DWT and DFT*, Tech. Report 3, Departement of Mathematics and Computer Science Philipps-University Marburg, 2003.
7. F. Mörchen, I. Mierswa, and A. Ultsch, *Understandable models of music collections based on exhaustive feature generation with temporal statistics*, 12th International Conference on Knowledge Discovery in Data and Data Mining (ACM SIGKDD 2006), 2006, pp. 882–891.
8. Department of Energy and Climate Change, *Smart metering implementation programme*, White paper, Department of Energy and Climate Change, July 2010.
9. P. Smyth, *Clustering sequences with hidden markov models*, Advances in Neural Information Processing Systems, vol. 9, 1997.
10. D. Toshniwal, *Feature extraction from time series data*, International Journal of Computational Methods in Sciences and Engineering **8** (2009), no. 1, 99–110.