Anomaly detection in electricity consumption data of buildings using predictive models

Jesse Eisses

6352189

Bachelor thesis for Artificial Intelligence

18 ECs

University of Amsterdam

Faculty of Science

Science Park 904

1098 XH Amsterdam

Supervisor

dhr. dr. M.W. van Someren

m.w.vansomeren@uva.nl

June 27, 2014

Abstract

In the electricity consumption data of buildings it is hard to see where the electricity is being used for. This makes it hard to detect a malfuncioning piece of equipment. This paper uses several machine learning techniques to automatically detect anomalies in the electricity consumption data. As an example, the data of 5 university buildings is analyzed. Features are constructed for the outdoor climate and for representing time. Several machine learning techniques are compared for creating models, including kNN, SVM and ANN. It turns out that an ANN has the best performance in this setting. A seperate model is built for each building. The models are then used for anomaly detection using a threshold method. The anomalies are compared between buildings, this gives more insight in the nature of the anomly.

Contents

1   Introduction

A buildings electricity usage can be hard to keep track of. In big buildings, such as universities and government institutions, hundreds of people make use of the same electricity source. They charge their phones, use their laptops and ride the elevators on varying schedules. The building also has systems running that control the indoor climate, power the fridges and refresh the air. If one of these systems fails or is misconfigured (if it has an *anomaly*) this can have a negative impact on the energy consumption. For example, an air conditioning could be misconfigured to cool air during the winter.

Energy efficiency is an important topic these days, so it is desirable to fix anomalies as soon as possible. If anomalies could be detected automatically and be reported to the building manager, the energy waste of an anomaly is minimized. Unfortunately this is not an easy task, as the buildings often do not have detailed information about the consumption of the individual parts and people inside.

This report tries to solve this problem by *creating predictive models* for the *automated anomaly detection* in the *electricity consumption data* of the buildings. The energy consumption data that will be analyzed consists of measurements of the total electricity consumption of the building on a constant time interval. Because the project has a limited scope other forms of energy (most notably gas) will not be discussed. The buildings that will be analyzed are university buildings from the UvA[1] and HvA[2] that made their data available for research, they will be used as an example throughout the paper but the methods and theory do not depend on them.

In the next chapter the available data will be discussed. An overview of the electricity data will be presented and analyzed. The weather data and data about the buildings will also be highlighted. In chapter 4 the data will be processed into more descriptive features. The features will be compared and their correlation with the electricity consumption are calculated.

Chapter 5 will introduce multiple algorithms for creating models that fit the data: kNN, SVM and ANN models will be created. Some effort is put in finding the best parameters for the algorithms. A separate model will be created for each building. Because anomalies can occur on different time scales, models will be created for hourly and daily intervals.

In chapter 6 the best performing model is used to find anomalies in the energy consumption. If the real value of a sample differs too much from the predicted value it will be classified as an anomaly. The best threshold for the classification is related to the standard deviation of the error and will be determined empirically.

Lastly, in chapter 7, anomalies from different buildings will be compared to find similarities. The comparison gives more insight in the nature of the anomalies. For example, if anomalies occur in multiple buildings at the same time, it is possible the anomaly is in the training data instead of the building electricity consumption.

---

[1]University of Amsterdam
[2]Hogeschool van Amsterdam

2 Theoretical background

This paper is not the first to create predictive models or to use them for anomaly detection. Some of the previous work and relevant theories will be discussed here.

2.1 Previous work

The topic of this paper is anomaly detection in energy consumption data of buildings. Several projects handled the same problem in previous years using the same building data. Jasper [16] tried to use unsupervised learning on a detailed dataset, which did not yield good results. Another group [6] created models for the gas data using linear regression; they also made a working model for the amount of people in the building. In addition, 2 other groups [2] and [14] used a different dataset to predict energy consumption and used the results for anomaly detection. This paper can be seen a continuation on their efforts. The previous papers provide a foundation to work from but leave enough room for expansion and improvement.

2.2 Theory

The data that is being analyzed is a time series; for which several forecasting techniques have already been explored. Until recently, time series analysis has been dominated by linear statistical models. As shown by Bontempi *et al* [1], machine learning models have been yielding good results. For example, Lora *et al* [12] showed that a simple k Nearest Neighbors method can already be successful. Nogales *et al* [13] later show that Dynamic regression also performs well on similar data. Several papers suggest that artificial neural networks are the best method for prediciting such data [7, 5, 8, 10].

In long term time series (such as the data in this paper) it is necessary to describe the cyclic characteristic of the data. This has been discussed by Gao *et al* [7], who suggests several features to describe these cycles. Drezga [4] and Gonzales [8] have also researched useful features for machine learning on time series.

An important part of this project is finding descriptive features for the buildings. This topic has not had much attention, but Ekici and Aksoy [5] analyze certain aspects of buildings in an ANN setting, which is certainly applicable in the use case of this project.

Once a model has been constructed, it is straight forward to perform the anomaly detection. A recent visual approach to anomaly detection is done by Janetzko *et al* [11].

The above mentioned papers all take a supervised learning approach, and most of them have positive results. A similar methodology will be used in this paper. In contrast several projects tried an unsupervised approach using clustering, with varying success [16, 18, 17].

3 Data

To predict the electricity consumption input variables are required that are related to the consumption. The variables need to be extracted from several data sources. There are 2 kinds of data that are used:

1. Building data

2. Local weather data

Building data is often kept private by the owners, so permission had be asked to analyze for inclusion in this project. Fortunately, most weather data is publicly available on the web. This chapter discusses the source and properties of these datasets.

3.1 Building data

There are several buildings that made their energy consumption data available for this research, they are listed in table 1. The data of each building consists of the electricity consumption at a constant time interval. This time interval differs per building and is depicted in the *granularity* column. The lowest common frequency in the data is 1 hour, which means there are 24 measurements each day. In the context of anomaly detection this should be sufficient, as anomalies that last shorter than 1 hour have little influence on the total electricity consumption. Looking at lower resolutions also adds noise to the data which makes prediction more complicated. For that reason this paper will focus on the hourly consumption data.

The other data sets are converted to hourly granularity by averaging the consumption over each hour.

3.2 Weather data

The previous papers [2, 16, 14, 6] that researched anomaly detection on this data, showed that certain weather features are good estimators for the electricity consumption. The weather data is collected

Feature	Unit
Temperature	$0.1\ C$
Wind speed	$0.1\ m/s$
Global radiation	J/cm^2
Rain	$0.1\ mm$
Sun duration	0.1 hr

Table 2: Weather data from the KNMI that is relevant to the electricity consumption.

from the KNMI[3], that keeps a database of local weather measurements in the Netherlands. The weather station closest to the buildings in table 1 is Schiphol. The data is measured on an hourly granularity (which matches the building data) and has a lot of variables. In total there are 21 variables in the dataset. A selection of 5 variables that are likely relevant to the consumption are listed in table 2.

3.3 Basic analysis

To understand the data above it helps to visualize it. The electricity consumption is the most important as this is the variable to be predicted. The data can differ a lot per building. This is illustrated in figure 1 where the consumptions of 2 buildings are plotted. There is a clear periodicity in the data of each building, but the relation between the buildings is not clear. Figure 1(b) shows the consumption for 2 different years for the *TTH* building. It shows there is similarity between years; seasonal and monthly cycles are visible. This is important because it means a well formed model should be able to predict the consumption with high accuracy. Because the relation between the consumption of different buildings is not clear (which makes it complicated to model) a separate model will be created for each building.

---

[3]Koninklijk Nederlands Meteorologisch Instituut http://www.knmi.nl

Code	Building	Period	Granularity
761 - KMH	Koetsier Montaigenhuis	2008-2014	5 min
763 - KSH	Kohnstammhuis	2011-2014	15 min
764 - TTH	Theo Thijssenhuis	2011-2014	15 min
882 - WBW	Kroonstate	2008-2014	5 min
425 - Singel	Universiteitsbibliotheek	2012-2014	5 min
505 - G647	Science Park	2013-2014	1 hr

Table 1: Buildings used in this research that made their energy consumption data available.
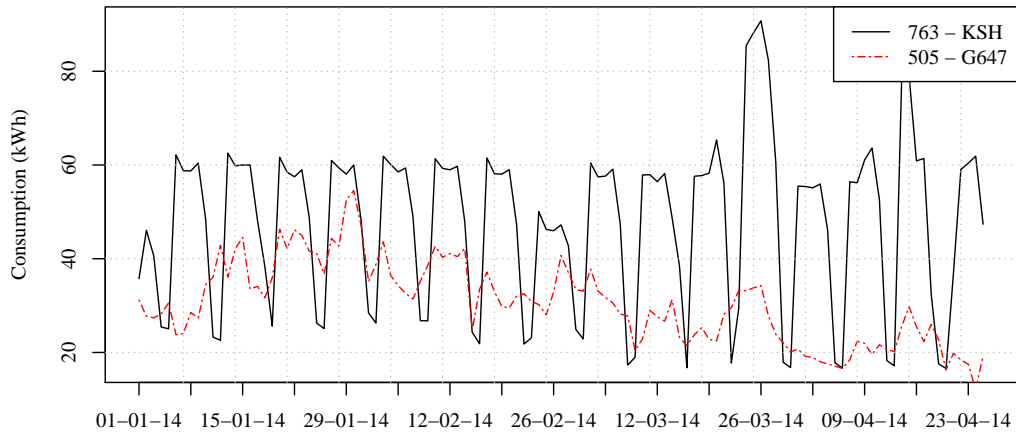
(a) Electricity consumption of KSH and SP in 2014, on a daily scale.
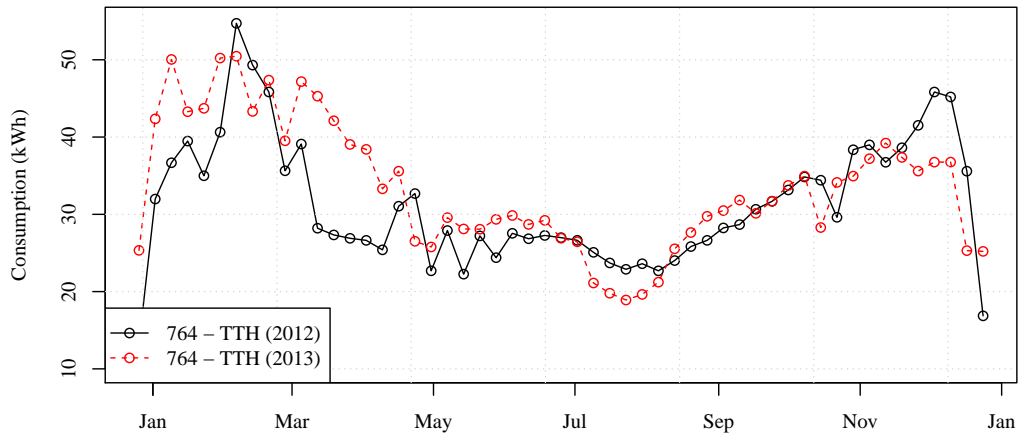
(b) Electricity consumption of TTH for the years 2012 and 2013 on a weekly scale.

Figure 1: Graphs of a small section of the electricity consumption.

4   Features

The data collected in the previous chapter will now be preprocessed and analyzed. Afterwards, the most useful features can be used to train predictive models. Not all data will be useful as an input feature. If a feature is not related to the electricity consumption it is better to leave it out, because it can cause overfitting. This chapter will introduce a set of features that is descriptive enough for model creation.

4.1   Weather features

Weather features were created straight from the KNMI data collected in the previous chapter. Only features relative to electricity consumption were considered. Table 3 shows the selected weather features.

Feature	Value
Temperature	Celcius
Global radiation	J/cm_2
Wind speed	m/s

Table 3: Weather features

The *rain* and *sun duration* variables are not included, as it turned out they did not contribute to the model accuracy. This can be explained because the sun duration variable is largely embedded in the radiation feature, while the rain variable might not be related to electricity consumption at all.

4.2   Holidays and weekends

During holidays and weekends the energy consumption is different from regular usage. Because this can cause problems for the machine learning algorithms, the data is split in 3 parts:

1. Weekday (Monday to Friday)

2. Weekend (Saturday and Sunday)

3. Vacation (any official holiday)

Feature	Value
Hour	0-23
Weekday	1-5
Cos hour	$\cos\left(Hour \cdot 2\pi/24\right)$
Sin hour	$\sin\left(Hour \cdot 2\pi/24\right)$
Cos weekday	$\cos\left(Weekday \cdot 2\pi/5\right)$
Sin weekday	$\sin\left(Weekday \cdot 2\pi/5\right)$

Table 4: Time features

This means a model must be trained for each of the datasets. Because the weekend and vacation data are stable and have a low consumption footprint, they are not interesting for anomaly detection, as an anomaly will occur anywhere there is a positive consumption. This report will only concentrate on the weekday dataset, but the same methods will apply for the other datasets as well.

4.3   Time features

Intuitively time is the most predictive feature for the electricity consumption. For example, if a sample has a time feature "Monday 6am", the electricity consumption will be low because the building is still closed. There are several measures of time that can be used as features. Firstly, the *time of day* gives the hour of the day between 0:00 and 23:00. Secondly, the *weekday* gives the day of the week (Monday - Friday).

Most learning algorithms work best with numeric features. The weekday and time of day can be transformed to numeric by using their index, 0-23 and 0-6 respectively. Further, some learning algorithms (e.g. neural networks) work better with continuous variables. To transform the time of day and weekday to a continuous scale, seasonal indices as suggested by [7] and [4] can be used. This yields the time features in table 4.

4.4   Time series features

A time series is an ordered dataset that is sampled on equal time intervals. The data used in

Feature	Value
p_1	Consumption at $t-1$.
$\dots$	$\dots$
p_n	Consumption at $t-n$.

Table 5: Time series features

Feature	Correlation
Temp	0.1664771
Rain	-0.0139693
Rain amount	-0.0192769
Snow	-0.0044864
Cloudiness	0.0104912
Radiation	0.4212074
weekday	-0.3711609
weekdaySin	0.0555005
weekdayCos	-0.3132163
hourSin	0.3056213
hourCos	-0.1348746

Table 6: Features and their raw correlation with electricity consumption.

this project is such a time series. Predicting data in a time series is called forecasting. As shown by [10, 12, 1], forecasting can be successful on electricity consumption data. Forecasting tries to predict the value at time t using the n previous observations, as shown in equation 1.

$$y_t = s(y_{t-1}, y_{t-2}, \dots, y_{t-n}) \qquad (1)$$

Where y_t is the energy consumption at time t, and s is a predictive function. It is possible to incorporate forecasting theory in our current setting by introducing the features $p_1, \dots, p_n$ from table 5.

Time series features are efficient in one-step-ahead forecasting. However, there are some drawbacks which makes them less useful for anomaly detection. A model depending on the previous observations has difficulty adjusting to change. The models that were trained with the features from table 5 had a higher accuracy, but they failed at anomaly detection. The reason is that every time the consumption changes, the prediction lags behind 1 or 2 hours. These lags are often detected as anomalies. The models trained without time series features had a lower accuracy, but did not suffer from this lag and performed better at anomaly detection.

Time series features are an interesting concept, but as they fail in the current setting they will not be used further in this paper.

4.5 Correlation

Now the raw data is converted to features ready for machine learning, it is good to check how much the features are related tin what measureo the electricity consumption. There are statistical methods that calculate such relations. The correlation measures the strength of the linear relation between to variables as a number between -1 and 1, where -1 and 1 represent a strong relation and 0 represents no relation. To find the importance of each feature the correlation with the electricity consumption will be calculated and is shown in table 6.

Note that the correlation only captures a simple relation between variables, a variable with 0 correlation can still be useful in a model that uses more complex relations. However it is a fact that the features with high correlation are strongly related to the electricity consumption.

5 Models

A model predicts the value for a sample with a specific formula, that takes the sample's features as inputs. In other words, it defines a relation between the features and the value to be predicted. The challenge in creating a good model is to find a formula that is similar to the formula that drives the underlying data. For example, if there is a linear relation between the features and the values, the formula would be of the form:

$$f(x) = ax + b \tag{2}$$

To fit this model to the data, the function must be tuned to find the optimal values for a and b. This is done by a learning algorithm.

When the relation between features and values is less obvious, for example non-linear, more advanced formulas have to be used. Over the last few years dozens of models and learning algorithms have been constructed for many purposes. In this chapter some of these will be implemented and evaluated for the electricity data.

5.1 Training and test set

For model creation and evaluation the data must be split in multiple parts. A *train set* is used to train the models, while a separate *test set* is used to evaluate the models. Both the training and the test will use the hourly data set. In the following experiments the following data will be used:

	Train	**Test**
Start date	2008-01-01	2014-01-01
End date	2012-12-31	2014-04-28
Samples	24384	1826
Building	761-KMH	761-KMH

5.2 k Nearest Neighbors

In order to create a baseline for other learning algorithms, a basic model will be constructed using k-Nearest-Neighbors [3]. kNN has several aspects that make it a suitable as a baseline (as explained by Bontempi [1]) which are:

- It is easy to implement and to interpret the results.

- It does not make assumptions on the underlying function that drives the data, and has no assumptions about the nature of the noise.

- It can quickly be applied to large and mutating datasets, as no training is needed.

In kNN the value of a sample is derived by finding the k most similar samples in the past, and combining their values. The value for k was determined by comparing multiple models, and was set at 3.

5.3 Neural networks

An artificial neural network (ANN) is a more advanced machine learning method. The concept of an ANN is inspired by how neurons in the human brain work. The network consists of several layers of *neurons*, the neurons in adjacent layers are connected by weights. The inputs and output of the model always form the first and last layer in the network, between them is a variable amount of hidden layers. The neural network used in this paper is illustrated in figure 3.

A neural network has the following parameters:

- h number of hidden layers

- $size$ the number of nodes in each hidden layer

- $decay$ rate of decay for unused weights, used to prevent over fitting

The parameters for an ANN are important for its performance and require some additional work. Much research has been done in ANN parameter selection, as explained by Drezga *et al.* in [4]. Previous research has shown [7, 5, 8] that 1 hidden layer is enough for most use cases.
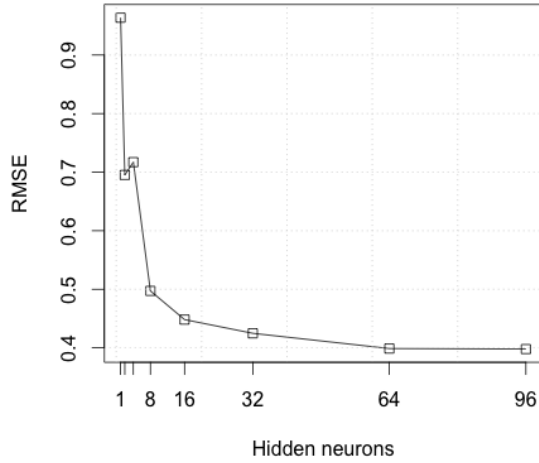
Figure 2: Error measures the the $size$ parameter

The decay parameter was not important in this setting, as no over fitting occurred. It was held constant at $5 \cdot 10^{-4}$.

The $size$ parameter is the most important. A neural network with size 1 (it has 1 neuron in the hidden layer) is a linear model. The more neurons are added the more complex the relation between variables becomes. The value was estimated empirically by training multiple models and comparing their error, the results are in figure 2. In the graph it is visible that after 32 neurons the profit diminishes.

5.4 SVM

Another popular machine learning algorithm is the support vector machine (SVM). There has been research by Tay [15] that shows an SVM performing better than a neural network in a similar setting. Tay's research relied mostly on time series features so it is interesting to check if an SVM is still a good option.

By design a SVM works as a binary linear classifier, which means it can only classify samples in one category or another. It does this by separating the training samples with a linear boundary in the best way possible, where the best boundary has the biggest margin between the samples. However, the SVM can be used when the samples are not linear separable (e.g. the electricity consumption data) by implementing a Kernel function. In this paper the same Gaussian kernel is used as described by Tay, as it has good performance in the general use case.

The Gaussian kernel works by allowing training samples to override the boundary by a certain error. The goal of the SVM it then to minimize this error for all samples, while keeping the margin between samples and the boundary as big as possible. There is payoff to be made between the maximal error and minimal boundary which is defined by the parameter C: the importance of the error. Figure 4 shows the influence of C on the training error. Larger values of C require more training to fit the model.
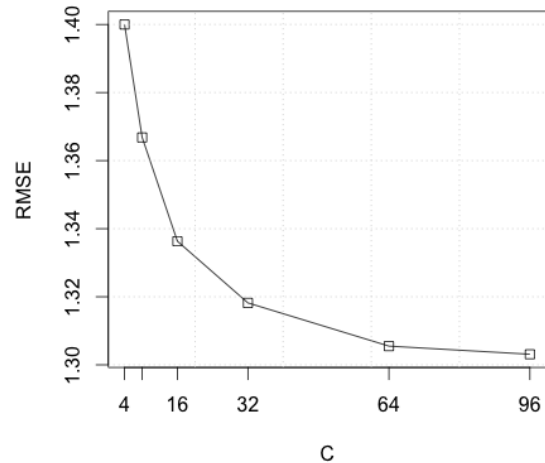

Figure 4: Parameter tuning for SVM

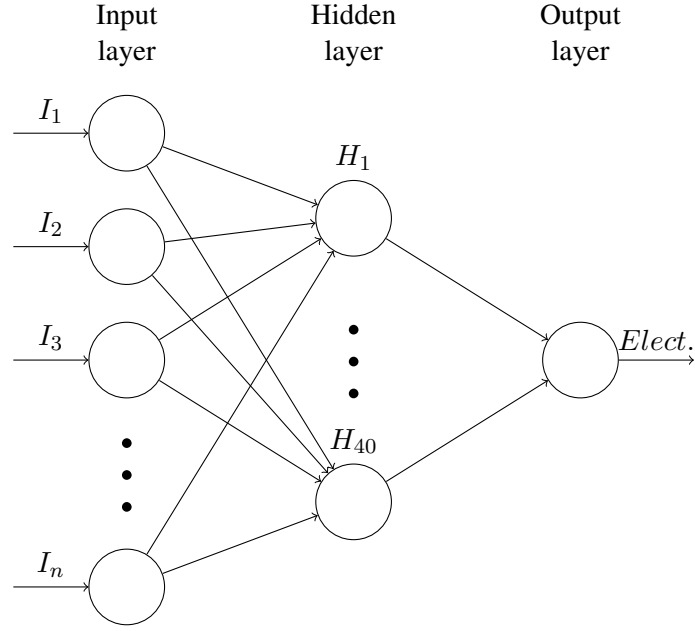The optimal value for C is around 40, to the same as the ANN size in figure 2 (coincidentally).

Figure 3: Layout of the neural network used in this paper

5.5 Comparison

Table 7 shows the results for the 3 models we created. The error measures depict the same outcome but for the data in this project MASE is the most accurate[9]. The error is measured by using a large train set (2008-2013) and a smaller test set (2014). It is clear the neural network is the best model.

Model	MAPE	MASE
kNN	69%	1.38
ANN	83%	0.91
SVM	69%	1.14

Table 7: Different models and their error measures

The errors are for the KMH building on a test set of the first 4 months of 2014. Some of the other buildings have more accurate models (for the TTH building the ANN has a MAPE of 90%).

6 Anomaly detection

In the previous chapter a model has been made for each building that accurately predicts the electricity consumption. This model can be now be used the find anomalies in the data. The idea is to let the model predict the consumption on time t. If the difference between the prediction and the actual value is above a certain threshold we classify it as an anomaly. The error is defined as:

$$E = |p - r| \tag{3}$$

Where p is the predicted value and r is the real value.

There are multiple ways to test the anomaly detection. Of course it can be applied to live data, but this is hard to acquire and to validate. An alternative is to apply it to the training data and try to find historical anomalies.

A sample will be classified an outlier if its error is above a certain threshold. This threshold can be determined by experimentation. Intuitively something is an outlier if its error is higher than the other errors. The *three-sigma-rule* is a method that implements this intuition. It states that something is an outlier if the error is larger than 3 times the standard deviation. This is a simple way of outlier detection, but it can work well if the models from the previous chapter are accurate enough. The threshold will be determined over the train set, not just the prediction part, to make it more robust.

6.1 An example: KSH

The outlier detection can be tested by applying it to part of the test data. In the following example, data from the KSH building is used. For this building 2 models are made with the method from chapter 5: a daily model and an hourly model. These models were trained on data from the period *2013-06-01* to *2014-04-26*. The anomaly detection will be applied to from *2014-01-01* to *2014-04-26*.

6.1.1 Detection

Firstly, the threshold must be calculated by finding the standard deviation of the error on the training data. The models are used to predict the train data, these are the black lines in figure 5(a). The error function 3 can now be applied, which yields the standard deviation:

$$\sigma = std(E) \tag{4}$$

Secondly, the models are used to predict the data from the anomaly detection set, and the error is calculated with 3. Anomalies are found by filtering this data on $> 3\sigma$. The results are shown in figure 5. The vertical lines are times where an anomaly is found; the line is drawn between the prediction and the real value, so the length of the line is equal to the size of the error.

Figure 5(a) shows the daily analysis for the anomaly detection set. The hourly anomalies are less stable than the daily anomalies: there are more unexpected outliers and fluctuations. This is because the daily data is in fact a smoothing of the hourly data. This smoothing aspect of the daily data can be used to eliminate noise. The daily scale is used to find anomalies on long time periods, these anomalies are then analyzed on an hourly scale.

6.1.2 Analysis

Two types of anomalies can been seen in the results:

1. Isolated anomalies (*peaks*), such as the point at *12/03* in figure 5(a).

2. Sequence anomalies (*groups*), such as the 3 points around *25/03* in figure 5(a).

The $peak$ anomalies are common in the hourly data but are hard to explain. The $goup$ anomalies are common on both time scales and they can be explained most of the time. If the hourly data contains a $group$ anomaly this often results in an anomaly on the daily scale as well.

date	H	Avg. $|E|$	explanation
14-04-2014	12	35	Unknown
14-04-2014	14	33	Unknown
14-04-2014	19 - 23	48	Building is open for an introduction evening.
15-04-2014	0	47	Possibly still the introduction evening
15-04-2014	8 - 10	48	Open day for new students; could cause higher energy consumption
15-04-2014	13 - 16	39	Open day for new students
16-04-2014	0	31	Unknown
18-04-2014	9 - 16	40	Building closed for "Goede vrijdag" (dutch holiday)
21-04-2014	9 - 17	42	Building closed for Eastern
21-04-2014	21 - 22	33	Building closed for Eastern

Table 8: The anomalies from 5(c) listed with explanation and error value. *Group* anomalies are concatenated and their error value is averaged to make the view more compact.

As an example the anomalies from figure 5(c) are listed in table 8. Note that most of the anomalies have a logical explanation. In fact, most of the anomalies in table 8 should have been accounted for in the training data. For example, the 2 holidays (Eastern en "Goede vrijdag") should have been filtered from the training data by the *holiday* feature.

The 3 unidentified anomalies are all *peak* anomalies. They are interesting because they could point to real anomalies resulting in energy inefficiency. In this case, the error value for these anomalies is lower than the others, which could mean that the anomaly threshold is too low. There are however many significant *peak* anomalies in the data that can not be explained, such as the first anomaly in figure 5(b).
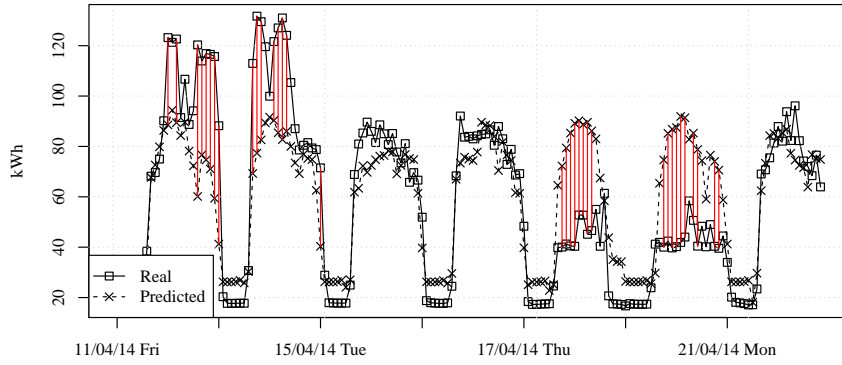
A more detailed anomaly explanation or the detection for more buildings does not fit within the scope of this project. However, the next chapter will try to combine anomalies from different buildings for further analysis.
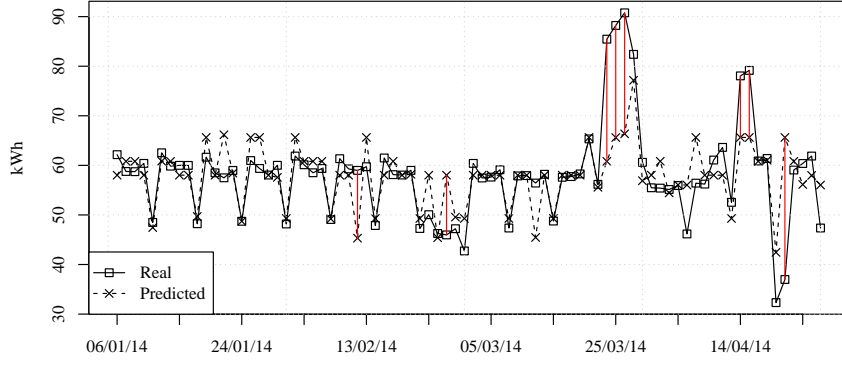
(a) Daily scale
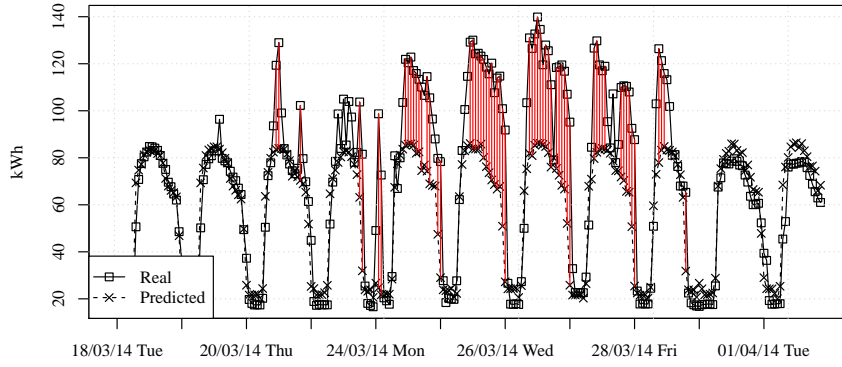
(b) Section of the first plot on an hourly scale

(c) Section of the first plot on an hourly scale

Figure 5: Fragments of the anomaly detection for building 763-KSH. The vertical lines point to detected anomalies, their length is the distance between the predicted and real value. **Note:** due to an unforeseen offset, the labels depict the *end* of the day.

7 Building comparison

In the previous chapter anomalies were detected with a model for each building. While these models are separate, they could have some aspects in common. There are multiple ways in which the models can be combined or compared, but this is subject of its own. This paper will only cover the simplest comparison: comparing the anomalies themselves. This can lead to insight about their nature and validity.

In table 5(c) the *KSH* building shows an anomaly during Eastern, because this holiday was not in the training data. As this is a mistake in the model building rather than a building anomaly, it should be detect as such. Because this anomaly is in the training data it should also occur in the other buildings. This should become visible when comparing the anomalies from different buildings. Figure 6 shows the result for this comparison.

As the graph shows, the buildings have multiple anomalies in common. The 2 anomalies in figure 6(b) are Eastern and "Goede vrijdag"; Dutch holidays. It is now possible to filter these out of the anomaly list, as they are most likely the result of a wrong model prediction, instead of a building anomaly.

Some of the building have more anomalies than the others. The main reason is that these buildings (TTH, KSH) only have 1 or 2 years of training data available, while the other buildings have up to 7 years of training data, which makes the predictions more stable.
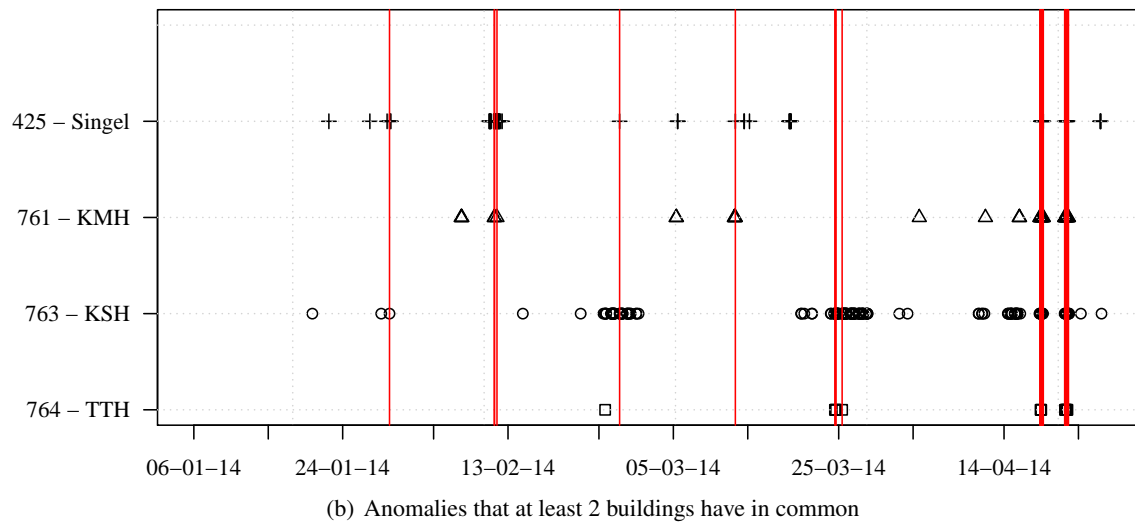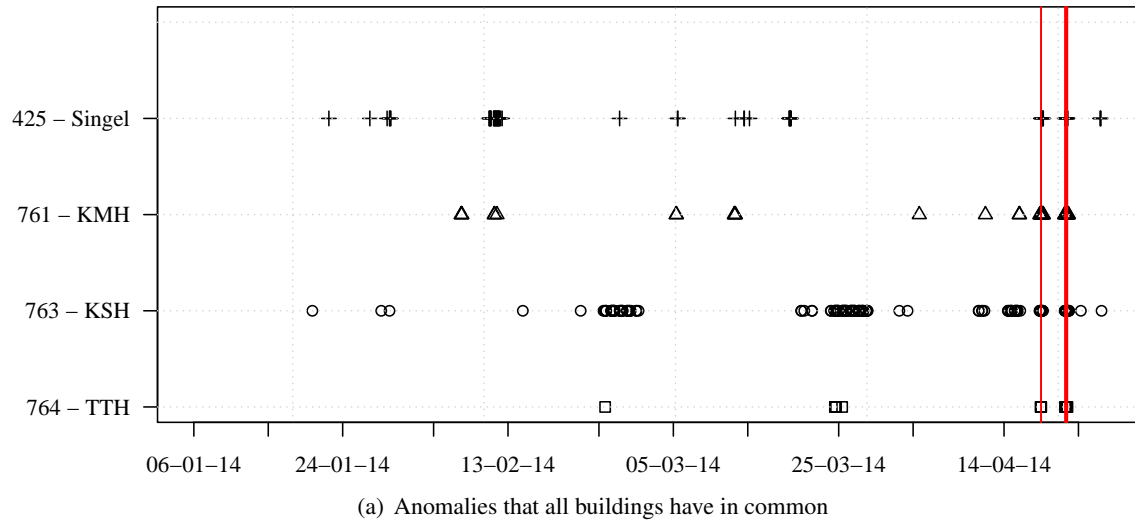
(a) Anomalies that all buildings have in common

(b) Anomalies that at least 2 buildings have in common

Figure 6: Anomalies from different buildings in 2014. Plots are on an hourly timescale.

8   Conclusion

In this paper, 3 methods were used to create a model for the prediction of electricity consumption data: k nearest neighbors, Support Vector Machine and artificial neural network. For each model the optimal parameters were found. The ANN turned out the best with an error of 14%.

Anomaly detection was done with the ANN model. For each building several anomalies were found and displayed. Some of the major anomalies were explained. The remaining unexplained anomalies contain the *real* anomalies, were electricity is being wasted. Unfortunately it is not possible to detect which of these anomalies they are. A list of suspicious anomalies (which are less likely to be noise) were collected and sent to the building managers.

The anomalies from different buildings were then compared to find commonalities. Anomalies that were common in all buildings were always explainable by a major event that caused unusual energy consumption.

8.1   Discussion

This paper took a complete approach to anomaly detection. It improved on previous research by further exploring model creation and anomalies. But the research is not finished yet, there are many ways to continue from here. It is also not certain the method in this paper is the best approach. A more advanced time series algorithm might also have succeeded at anomaly detection. However, I believe this project had good results for the problem at hand. There are multiple ways this research can be extended, which will be covered in the next paragraph.

8.2   Future work

The final results of this paper, which are the identified anomalies, are still crude. Not much can be said about them, except that they represent an un-usual amount of electricity consumption. To make sense of them the anomalies could be further analyzed, for example by looking for monthly or yearly recurring anomalies in a building.

Another way to get deeper insight in the anomalies is to look at a smaller time scale. By analyzing consumption on a 15 minute frequency, the structure of the anomaly could provide extra information.

Another way to extend this research is to generalize the model. In the current setting, a model is created for each building. This requires a lot of effort: every building must have enough data to train the model and the training can take a long time. Also maintaining the models is troublesome when each model must be updated with new data. There is a lot to be gained from creating a general building model, which would require more building parameters (e.g. isolation factor, size, build year).

References

[1] Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Aël Le Borgne. Machine learning strategies for time series forecasting. In *Business Intelligence*, pages 62–77. Springer, 2013.

[2] Kasper Bouwes, Tom Koenen, Rafiek Mohamedjoesoef, and Vincent Velthuis. Machine learning en klimaatbheer. Paper Leren en Beslissen, 2014.

[3] Belur V Dasarathy. Nearest neighbor (NN) norms: NN pattern classification techniques. 1991.

[4] IIII Drezga and S Rahman. Input variable selection for ann-based short-term load forecasting. *Power Systems, IEEE Transactions on*, 13(4):1238–1244, 1998.

[5] Betul Bektas Ekici and U Teoman Aksoy. Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software*, 40(5):356–362, 2009.

[6] Ysbrand Galama, Martijn Loos, Jonas Lodewegen, and Erwin van den Berg. Leren en beslissen: Luchtkwaliteitsbeheer. 2nd year project, 2013.

[7] Feng Gao, Xiaohong Guan, Xi-Ren Cao, and Alex Papalexopoulos. Forecasting power market clearing price and quantity using a neural network method. In *Power Engineering Society Summer Meeting, 2000. IEEE*, volume 4, pages 2183–2188. IEEE, 2000.

[8] Pedro A González and Jesús M Zamarreño. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings*, 37(6):595–601, 2005.

[9] Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.

[10] Syed M Islam, Saleh M Al-Alawi, and Khaled A Ellithy. Forecasting monthly electric load and energy for a fast growing utility using an artificial neural network. *Electric Power Systems Research*, 34(1):1–9, 1995.

[11] Halldór Janetzko, Florian Stoffel, Sebastian Mittelstädt, and Daniel A Keim. Anomaly detection for visual analytics of power consumption data. *Computers & Graphics*, 38:27–37, 2014.

[12] Alicia Troncoso Lora, Jesús Manuel Riquelme Santos, José Cristóbal Riquelme, Antonio Gómez Expósito, and José Luís Martínez Ramos. Time-series prediction: Application to the short-term electric energy demand. In *Current Topics in Artificial Intelligence*, pages 577–586. Springer, 2004.

[13] Francisco Javier Nogales, Javier Contreras, Antonio J Conejo, and Rosario Espínola. Forecasting next-day electricity prices by time series models. *Power Systems, IEEE Transactions on*, 17(2):342–348, 2002.

[14] Marco Schaap, Iason de Bondt, and Tom van den Bogaart. Anomalie detectie: Gasverbruik hva. 2nd year project, 2014.

[15] Francis EH Tay and Lijuan Cao. Application of support vector machines in financial time series forecasting. *Omega*, 29(4):309–317, 2001.

[16] Jasper Van Enk. Finding anomalies in building management data using clustering algorithms. BSc scriptie, 2013.

[17] Jarke J Van Wijk and Edward R Van Selow. Cluster and calendar based visualization of

time series data. In *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on*, pages 4–9. IEEE, 1999.

[18] T Warren Liao. Clustering of time series data a survey. *Pattern Recognition*, 38(11):1857–1874, 2005.