A

Comparison of Anomaly Detection between Statistical Method and Undercomplete Autoencoder

Muhammad Ayaz Hussain, University of
Saif ur Rahman, University of
Ioannis Iossifidis, University of
Christian Klaus, University of

This paper is concerned about the use of unsupervised learning (semi-supervised) algorithms in power consumption data, as well as the comparison of the performance between statistical method and Undercomplete Autoencoder in classification. As power consumption varies over time, therefore the labeled data created by that, becomes obsolete over times. Therefore, supervised learning is not a viable method in the prediction of energy consumption over a long period of time.

1. INTRODUCTION

In modern days, the amount of energy consumption is becoming a more serious issue as several companies and industries are addressing this issue in order to contain their expenses as unexpected variations can incur additional operational costs to their facilities. These fluctuations in electricity consumption can arise from various factors such as excessive use of heavy equipment like electric heaters in winters, room coolers during the summers etc. In order to detect and classify unusual amount of power consumption, outlier detection algorithm is designed and implemented on the provided data and compared its classification efficiency against a semi-supervised autoencoder. Presently, it is very difficult to predict the energy consumption anomalies precisely, since there are many factors influencing the energy usage, such as weather condition [Wang et al. 2012], occupancy [Pan et al. 2007] and operation of appliances [Chenglei et al. 2015; Royapoor and Roskilly 2015]

2. RELATED WORKS

Previous studies in data-driven building energy consumption prediction have utilized several methods such as engineering methods [xiang Zhao and Magouls 2012], statistical method [Lei and Hu 2009], Artificial Neural Networks (ANN) [Ekici and Aksoy 2009; Kusiak et al. 2010], Support Vector Machines (SVM) [Li et al. 2009a], fuzzy logic and grey model techniques [Guo et al. 2011], decision trees [Tso and Yau 2007] etc.

Also there have been some studies which compared the effectiveness of different algorithms in energy consumption prediction by comparing the results of two or more algorithms on a similar dataset. For example, Li et al. [Li et al. 2009a] compared SVM and Back Propagation Neural Network (BPNN) to predict hourly cooling load in a building ; [Xuemei et al. 2009] compared LS-SVM and BPNN for building cooling load forecasting; [Liu and Chen 2013] compared Support Vector Regression (SVR) and ANN for lighting consumption for a building; [Penya et al. 2011; Fernndez et al. 2011] compared AR Model, ANN, autoregressive integrated moving average (ARIMA), and Bayesian Network on a university campus for short term load forecasting by incorporating energy information as well as metrological data; [Platon et al. 2015] compared ANN and Case based Reasoning (CBR) for the prediction of hourly electricity consumption of an institutional building; [Jain et al.] compared SVM and MLR on a residential building; [Hou et al. 2006] compared ARIMA and ANN (need to download paper); [Fan et al. 2014] compared MLR, ARIMA, SVM, RF, MLP, BT, MARS, and kNN on a high rise commercial building ; [Chou and Bui 2014] emphasized on building geometry in their dataset while comparing ANN, SVM, CART, CHAID, and GLR ; [Edwards et al. 2012] compared MLR, FFNN, SVM, LS-SVM, HME-FFNN, and FCM-FFNN;

[Li et al. 2009b; Li et al. 2010] compared SVM, BPNN, Radial Basis Function Neural Network (RBFNN), and General Regression Neural Network (GRNN) on multiple residential buildings; Dagnely et al. [Dagnely et al. 2015] compared Ordinary Least Squared (OLS) Regression and SVR for the prediction of hourly energy consumption by taking temporal predictors such as occupancy, working hours and days of the week as well as metrological predictors such as ambient temperature and irradiance for a building; [Massana et al. 2015] compared MLR, MLP, and SVR for Short-term load forecasting in a non-residential building. A hybrid neural net ARIMA have been used by Chou et al.[Chou and Bui 2014] to detect anomalies in the power consumption in an office space. In another paper, Capozolli et al. [Capozzoli et al. 2015] used Classification and Regression Tree (CART), K-Mean clustering, Artificial neural networks and basic ensembling method (ANN BEM) and Outlier Detection methods such as Generalized Extreme Studentized Deviation (GESD) and Peak detection method on a building cluster. Whereas, [Borges et al. 2013] compared SVM and Autoregressive (AR) Model on the energy consumption on a couple of university campuses for a load forecasting.

As far as the use of autoencoder for classification is concerned, those have been used in previous works for classification tasks such as [Fan et al. 2018] which used autoencoder based unsupervised anomaly detection method on a building energy data on a single facility which was in detail as several variables were considered.

Autoencoder is also been used for unsupervised classification of the traffic flow of data packet in Internet [Höchst et al. 2017]. Whereas, [Qi et al. 2017] used stacked autoencoders to diagnose and classify the faults in a rotating machinery.

3. METHODOLOGY

3.1. Data Preparation and Analysis

Initially, the raw data was provided by the company, which consisted of several entries. Those entries were obtained from the energy consumption data of 227 stores on 15 min basis for around 11 months as shown in the Figure 1. Analysis of Figure 1 shows that there is an obvious increase in power consumption during winters possibly due to the usage of electric heaters and an major drop in the power consumption around Christmas and New Year (just above January).

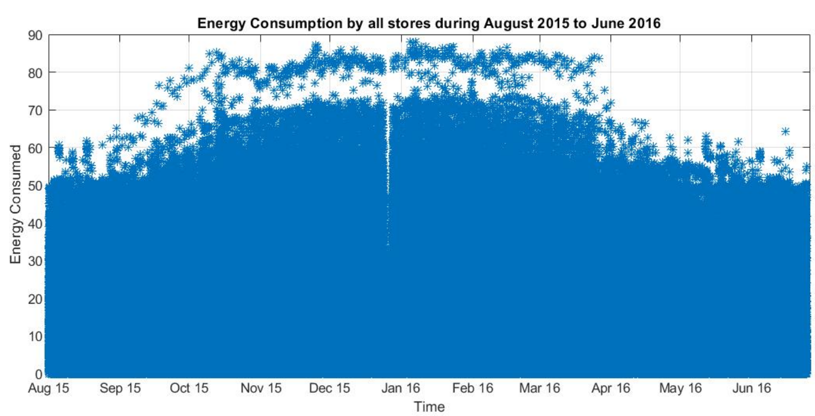


Fig. 1. Energy Consumption by all stores between August 2015 and June 2016

Afterwards, a random month is selected from the whole time period (in this case November 2015) for visualization and energy consumption of all stores is plotted as shown in Figure 2.

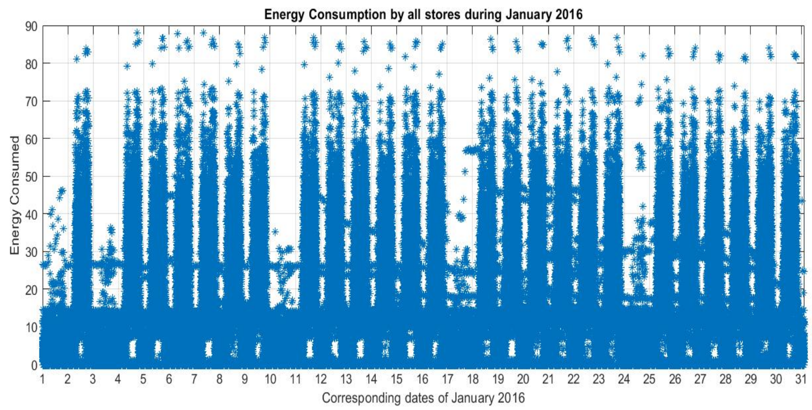


Fig. 2. Energy Consumption by all stores during January 2016

Still, it was difficult to pinpoint or visualize any single store from that plot. Therefore, a random store was chosen (in this case Store 105) for in depth analysis as shown in Figure 3.

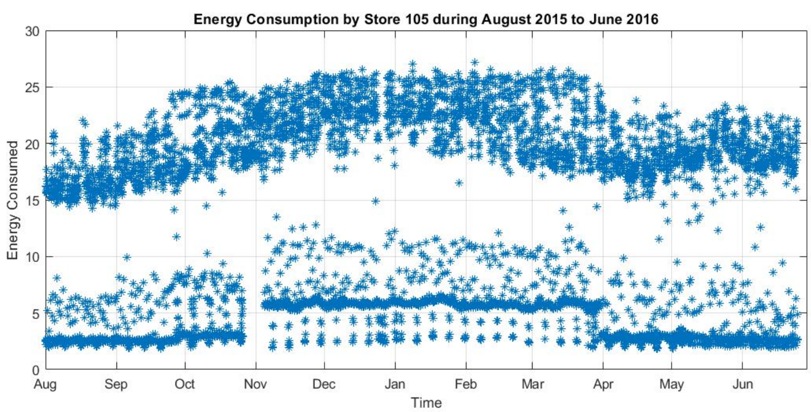


Fig. 3. Energy Consumption by Store 105 between August 2015 and June 2016

We can again see the noticeable increase in energy consumption during winter months from October till April as well as lower consumption during Christmas and New year holidays. Again, we used November which was randomly picked earlier to visualize the energy consumption by Store 105 as shown in Figure 4.

In order to select meaningful features to be implemented in Machine Learning or Statistical Modeling algorithms, it was necessary to analyze the given data throughly. After a comprehensive analysis, it was decided that the amount of energy consumption depends upon multiple factors, which are the store itself as all stores differs from

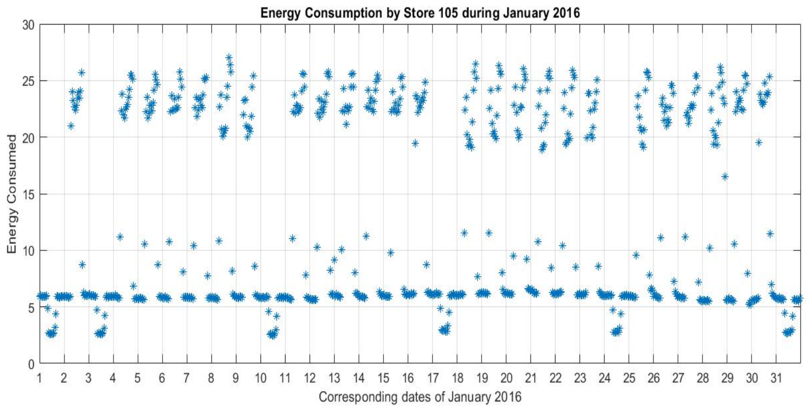


Fig. 4. Energy Consumption by Store 105 during January 2016

one another as each having its particular size, different number of energy consuming equipments and even have different geographical location which affect the amount of energy consumption. Aside from that, time of the day is another factor as during working hours energy consumption is higher. There is also a reduction in power consumption during the day even when the store is closed.

The data consisted of the energy readings from 227 different stores with 238 sensor values (as certain stores had multiple sensors) taken every 15 minute which spanned over a period of 11 months. The provided data was then throughly examined and it was subdivided into 2 main categories which were working days and non-working days as energy consumption showed different trends in both categories. Therefore, a Feature Vector was developed which consisted of respective month, hour of the day, store number, energy consumed, a tag for working day or non-working day. As the data was gathered from the sensors in the real world therefore, it was highly unbalanced outliers estimated to be around 2-5% of the total dataset.

3.2. Statistical Method for Outlier Detection

In order to detect the outliers, a feature vector was developed, which consisted of the hour of the day, store number, energy consumed, a tag for working and non-working day and on that data a modified version of Tukey's test [Mcgill et al. 1978] was performed, which involves the calculation of median of training data which is referred to $Q2$ and then again median of values lesser and greater than that of $Q2$ is calculated. Median of values lower than $Q2$ is refereed as $Q1$ (lower quartile) and those having greater than $Q2$ is referred as $Q3$ (upper quartile) as shown in Figure 8.

After determining $Q1$ and $Q3$, then Interquartile Range (IQR) is calculated by subtracting the value at $Q3$ by $Q1$.

After determining $Q1$ and $Q3$, then Interquartile Range (IQR) is calculated by subtracting the value at $Q3$ by $Q1$.

$$IQR = Q3 - Q1 \tag{1}$$

After calculating the IQR, we use it to calculate the Tukey's Fences or Inner Fences by the following equation,

$$InnerFenceLowerLimit = Q1 - \alpha(IQR) \tag{2}$$

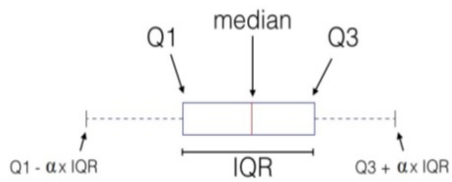


Fig. 5. Depiction of quartiles as used in the detection and removal of Outliers.

$$InnerFenceUpperLimit = Q1 + \alpha(IQR) \tag{3}$$

where,
α is the factor which varies the inner fence limits.

Any value lying beyond inner fence limits (upper and lower) can be considered as an outlier, and therefore, it is removed. A sample result of designed outlier detection algorithm is shown in Figure 8(a) and 9(a).

As the dataset used in this project was highly varied and there were several energy values that stayed the same for the whole period of time. In order to calculate the quartiles and medians, there must be at least 4 unique values in a dataset, therefore for those dataset having less unique values, some product of their respective average was used to determine Inner and Outer Fences.

3.3. Autoencoder

An autoencoder always consists of two parts, the encoder and the decoder, which can be defined as transitions ϕ and ψ such that:

$$\phi : \mathcal{X} \rightarrow \mathcal{F} \tag{4}$$

$$\psi : \mathcal{F} \rightarrow \mathcal{X} \tag{5}$$

$$\phi, \psi = \underset{\phi,\psi}{\arg\min}\ \|X - (\phi \circ \psi)X\|^2 \tag{6}$$

In a simple case, if there is one hidden layer, the encoder stage of an autoencoder takes the input $\mathbf{x} \in \mathbb{R}^d = \mathcal{X}$ and maps it to $\mathbf{z} \in \mathbb{R}^p = \mathcal{F}$.

$$\boldsymbol{z} = \boldsymbol{\sigma}(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}) \tag{7}$$

This image z is usually referred to as *code*, *latent variables* or *latent representation*. Here, σ is an element-wise activation function such as sigmoid function, hyperbolic tangent or a rectified linear unit. W is weight matrix and b is a bias vector. After that, the decoder stage of the autoencoder maps z to the reconstruction x′ of the same shape as x.

$$\boldsymbol{x'} = \boldsymbol{\sigma'}(\boldsymbol{W'}\boldsymbol{z} + \boldsymbol{b'}) \tag{8}$$

where, $\boldsymbol{\sigma'}$, $\boldsymbol{W'}$ and $\boldsymbol{b'}$ for the decoder may differ in general from the corresponding $\boldsymbol{\sigma}$, $\boldsymbol{W}$ and $\boldsymbol{b}$ for the encoder, depending on the design of the autoencoder.

Autoencoders are also trained to minimize reconstruction errors (such as squared errors):

$$\boldsymbol{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{x'}) = \|\boldsymbol{x} - \boldsymbol{\sigma'}(\boldsymbol{W'}(\boldsymbol{\sigma}(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b})) + \boldsymbol{b'})\|^2 \tag{9}$$

where, x is usually averaged over some input training set.

If the feature space $\mathcal{F}$ has lower dimensionality than the input space $\mathcal{X}$, then the feature vector $\phi(x)$ can be regarded as a compressed representation of the input x. If the hidden layers are larger than the input layer, an autoencoder can potentially learn the identity function and become useless. However, experimental results have shown that autoencoders might still learn useful features in these cases.

In case of Undercomplete autoencoders, $\mathcal{F}$ has lower dimensionality than the input space $\mathcal{X}$, then the feature vector $\phi(x)$ can be regarded as a compressed representation of the input x.

In our project, an undercomplete autoencoder was trained in a semi-supervised fashion on the values of normal energy consumption data. Afterwards, the trained model was used and evaluated on a pre-trained dataset.

4. EXPERIMENTS AND RESULTS

Outlier detection algorithm was developed and it performed adequately well in removing the outliers. Figure 6 shows the energy consumption data of each hour for all working days in January 2016, where as, figure 7 shows the same data as in figure 6 after the outlier removal algorithm implemented on it.

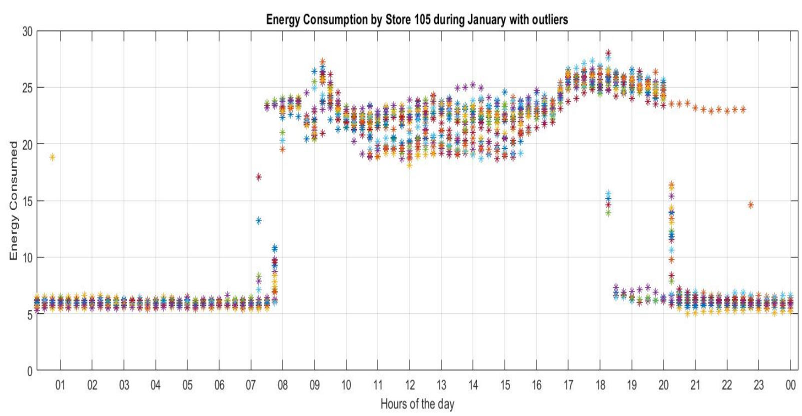


Fig. 6. Energy Consumption by Store 105 during working days of January 2016 with outliers

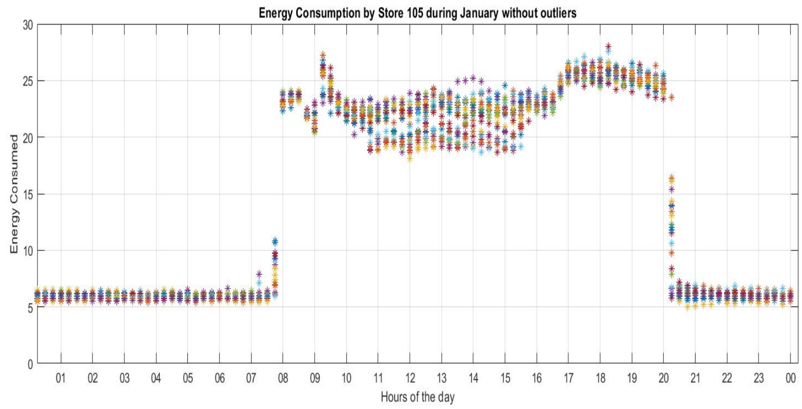
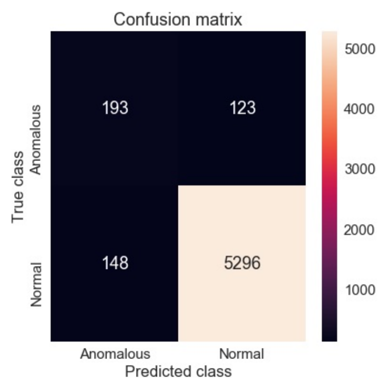
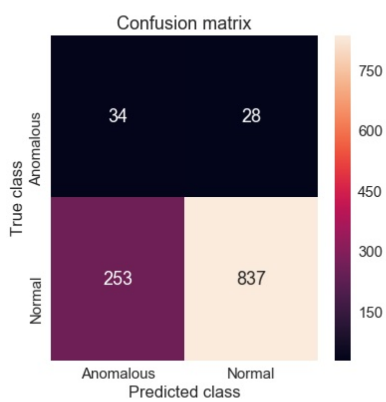


Fig. 7. Energy Consumption by Store 105 during working days of January 2016 without outliers

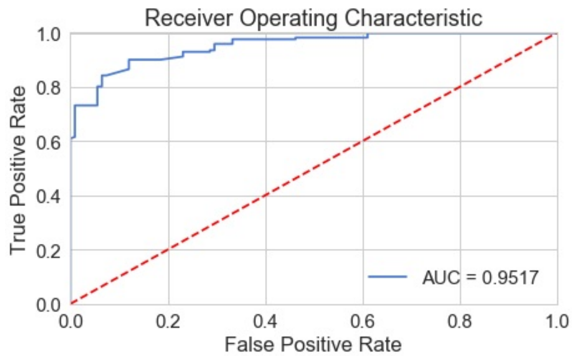


(a) ROC curve of Anomaly Detection Algorithm.

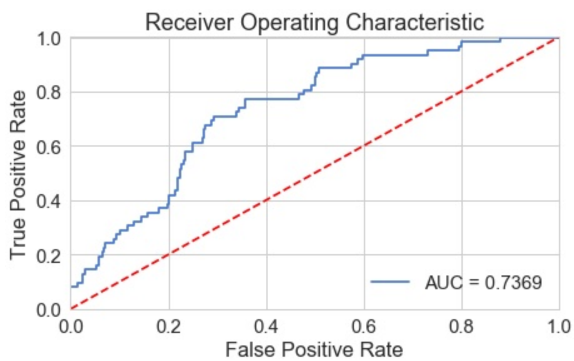


(b) Confusion Matrix of Autoencoder.

Fig. 8. Confusion Matrices



(a) ROC curve of Anomaly Detection Algorithm.



(b) ROC curve of Autoencoder.

Fig. 9. ROC curves

In Fig 8, we are comparing the accuracy of prediction of both algorithms with manually classified energy data of two stores using confusion matrices.

The same dataset which was used to make confusion matrices in Fig 8 is used to plot Receiver Operating Characteristics curves in Fig 9.

As we can see from the comparison of Fig 8 and Fig 9, we can interpret that the performance of Statistical method which is based on Tukey's test performed remarkably better with around 95% correlation with manual classification but the shortcoming of this method is that the dataset should be arranged in a proper manner as a single missing values in the dataset can make the classification ineffective. Whereas, classification using Autoencoder performed adequately with around 70-75% correlation with manually classified data but it required pre-classified data containing no anomalies.

5. DISCUSSION/OUTLOOK/CONCLUSION/FUTURE WORK

In this paper, the performance of a semi supervised Autoencoder and a statistical model for outlier detection was compared. The experiment resulted in the varying results from each algorithms as Autoencoder did not require data to be arranged precisely and were robust against mixed up dataset whereas, Statistical based outlier detection model performed more accurately in detecting outliers but it required careful arrangement of the data.

In future, it is planned to use more variables in the feature vector of energy consumption data such size of the facility, number of energy consuming equipments, occupancy of the building, and other sensor data such as outside temperature and solar radiation. As more data is embedded in the feature vector, there will be more focus on using totally unsupervised learning using Autoencoder to discover and learn new pattern in the given dataset and do the classification more precisely.

REFERENCES

Cruz E. Borges, Yoseba K. Penya, Ivn Fernndez, Juan Prieto, and Oscar Bretos. 2013. Assessing Tolerance-Based Robust Short-Term Load Forecasting in Buildings. *Energies* 6, 4 (2013), 2110–2129. DOI:http://dx.doi.org/10.3390/en6042110

Alfonso Capozzoli, Fiorella Lauro, and Imran Khan. 2015. Fault detection analysis using data mining techniques for a cluster of smart office buildings. *Expert Systems with Applications* 42, 9 (2015), 4324 – 4338. DOI:http://dx.doi.org/https://doi.org/10.1016/j.eswa.2015.01.010

H. Chenglei, L. Kangji, L. Guohai, and P. Lei. 2015. Forecasting building energy consumption based on hybrid PSO-ANN prediction model. In *2015 34th Chinese Control Conference (CCC)*. 8243–8247. DOI:http://dx.doi.org/10.1109/ChiCC.2015.7260948

Jui-Sheng Chou and Dac-Khuong Bui. 2014. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy and Buildings* 82 (2014), 437 – 446. DOI:http://dx.doi.org/https://doi.org/10.1016/j.enbuild.2014.07.036

Pierre Dagnely, Tom Ruette, Tom Tourwé, Elena Tsiporkova, and Clara Verhelst. 2015. Predicting Hourly Energy Consumption. Can Regression Modeling Improve on an Autoregressive Baseline?. In *Revised Selected Papers of the Third ECML PKDD Workshop on Data Analytics for Renewable Energy Integration - Volume 9518*. Springer-Verlag, Berlin, Heidelberg, 105–122. DOI:http://dx.doi.org/10.1007/978-3-319-27430-0_8

Richard E. Edwards, Joshua New, and Lynne E. Parker. 2012. Predicting future hourly residential electrical consumption: A machine learning case study. *Energy and Buildings* 49 (2012), 591 – 603. DOI:http://dx.doi.org/https://doi.org/10.1016/j.enbuild.2012.03.010

Betul Bektas Ekici and U. Teoman Aksoy. 2009. Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software* 40, 5 (2009), 356 – 362. DOI:http://dx.doi.org/https://doi.org/10.1016/j.advengsoft.2008.05.003

Cheng Fan, Fu Xiao, and Shengwei Wang. 2014. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy* 127 (2014), 1 – 10. DOI:http://dx.doi.org/https://doi.org/10.1016/j.apenergy.2014.04.016

Cheng Fan, Fu Xiao, Yang Zhao, and Jiayuan Wang. 2018. Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. *Applied Energy* 211 (2018), 1123 – 1135. DOI:http://dx.doi.org/https://doi.org/10.1016/j.apenergy.2017.12.005

I. Fernndez, C. E. Borges, and Y. K. Penya. 2011. Efficient building load forecasting. In *ETFA2011*. 1–8. DOI:http://dx.doi.org/10.1109/ETFA.2011.6059103

J.J. Guo, J.Y. Wu, and R.Z. Wang. 2011. A new approach to energy consumption prediction of domestic heat pump water heater based on grey system theory. *Energy and Buildings* 43, 6 (2011), 1273 – 1279. DOI:http://dx.doi.org/https://doi.org/10.1016/j.enbuild.2011.01.001

J. Höchst, L. Baumgrtner, M. Hollick, and B. Freisleben. 2017. Unsupervised Traffic Flow Classification Using a Neural Autoencoder. In *2017 IEEE 42nd Conference on Local Computer Networks (LCN)*. 523–526. DOI:http://dx.doi.org/10.1109/LCN.2017.57

Zhijian Hou, Zhiwei Lian, Ye Yao, and Xinjian Yuan. 2006. Cooling-load prediction by the combination of rough set theory and an artificial neural-network based on data-fusion technique. *Applied Energy* 83, 9 (2006), 1033 – 1046. DOI:http://dx.doi.org/https://doi.org/10.1016/j.apenergy.2005.08.006

R. K. Jain, T. Damoulas, and C. E. Kontokosta. *Towards Data-Driven Energy Consumption Forecasting of Multi-Family Residential Buildings: Feature Selection via The Lasso*. DOI:http://dx.doi.org/10.1061/9780784413616.208

Andrew Kusiak, Mingyang Li, and Zijun Zhang. 2010. A data-driven approach for steam load prediction in buildings. *Applied Energy* 87, 3 (2010), 925 – 933. DOI:http://dx.doi.org/https://doi.org/10.1016/j.apenergy.2009.09.004

F. Lei and P. Hu. 2009. A Baseline Model for Office Building Energy Consumption in Hot Summer and Cold Winter Region. In *2009 International Conference on Management and Service Science*. 1–4. DOI:http://dx.doi.org/10.1109/ICMSS.2009.5301031

Qiong Li, Qinglin Meng, Jiejin Cai, Hiroshi Yoshino, and Akashi Mochida. 2009a. Applying support vector machine to predict hourly cooling load in the building. *Applied Energy* 86, 10 (2009), 2249 – 2256. DOI:http://dx.doi.org/https://doi.org/10.1016/j.apenergy.2008.11.035

Qiong Li, Qinglin Meng, Jiejin Cai, Hiroshi Yoshino, and Akashi Mochida. 2009b. Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. *Energy Conversion and Management* 50, 1 (2009), 90 – 96. DOI:http://dx.doi.org/https://doi.org/10.1016/j.enconman.2008.08.033

Qiong Li, Peng Ren, and Qinglin Meng. 2010. Prediction model of annual energy consumption of residential buildings. In *2010 International Conference on Advances in Energy Engineering*. 223–226. DOI:http://dx.doi.org/10.1109/ICAEE.2010.5557576

D. Liu and Q. Chen. 2013. Prediction of building lighting energy consumption based on support vector regression. In *2013 9th Asian Control Conference (ASCC)*. 1–5. DOI:http://dx.doi.org/10.1109/ASCC.2013.6606376

Joaquim Massana, Carles Pous, Lloren Burgas, Joaquim Melendez, and Joan Colomer. 2015. Short-term load forecasting in a non-residential building contrasting models and attributes. *Energy and Buildings* 92 (2015), 322 – 330. DOI:http://dx.doi.org/https://doi.org/10.1016/j.enbuild.2015.02.007

Robert Mcgill, John W. Tukey, and Wayne A. Larsen. 1978. Variations of Box Plots. *The American Statistician* 32, 1 (1978), 12–16. DOI:http://dx.doi.org/10.1080/00031305.1978.10479236

Yiqun Pan, Zhizhong Huang, and Gang Wu. 2007. Calibrated building energy simulation and its application in a high-rise commercial building in Shanghai. *Energy and Buildings* 39, 6 (2007), 651 – 657. DOI:http://dx.doi.org/https://doi.org/10.1016/j.enbuild.2006.09.013

Y. K. Penya, C. E. Borges, D. Agote, and I. Fernndez. 2011. Short-term load forecasting in air-conditioned non-residential Buildings. In *2011 IEEE International Symposium on Industrial Electronics*. 1359–1364. DOI:http://dx.doi.org/10.1109/ISIE.2011.5984356

Radu Platon, Vahid Raissi Dehkordi, and Jacques Martel. 2015. Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis. *Energy and Buildings* 92 (2015), 10 – 18. DOI:http://dx.doi.org/https://doi.org/10.1016/j.enbuild.2015.01.047

Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu. 2017. Stacked Sparse Autoencoder-Based Deep Network for Fault Diagnosis of Rotating Machinery. *IEEE Access* 5 (2017), 15066–15079. DOI:http://dx.doi.org/10.1109/ACCESS.2017.2728010

Mohammad Royapoor and Tony Roskilly. 2015. Building model calibration using energy and environmental data. *Energy and Buildings* 94 (2015), 109 – 120. DOI:http://dx.doi.org/https://doi.org/10.1016/j.enbuild.2015.02.050

Geoffrey K.F. Tso and Kelvin K.W. Yau. 2007. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy* 32, 9 (2007), 1761 – 1768. DOI:http://dx.doi.org/https://doi.org/10.1016/j.energy.2006.11.010

Liping Wang, Paul Mathew, and Xiufeng Pang. 2012. Uncertainties in energy consumption introduced by building operations and weather for a medium-size office building. *Energy and Buildings* 53 (2012), 152 – 158. DOI:http://dx.doi.org/https://doi.org/10.1016/j.enbuild.2012.06.017

Hai xiang Zhao and Frdric Magouls. 2012. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews* 16, 6 (2012), 3586 – 3592. DOI:http://dx.doi.org/https://doi.org/10.1016/j.rser.2012.02.049

L. Xuemei, L. Jin-hu, D. Lixing, X. Gang, and L. Jibin. 2009. Building Cooling Load Forecasting Model Based on LS-SVM. In *2009 Asia-Pacific Conference on Information Processing*, Vol. 1. 55–58. DOI:http://dx.doi.org/10.1109/APCIP.2009.22