Comparison of Anomaly Detection between Statistical Method and Undercomplete Autoencoder

Muhammad Ayaz Hussain
dept. name of organization (of Aff.)
name of organization (of Aff.)
Bochum, Germany
email address

Muhammad Saif ur Rahman
dept. name of organization (of Aff.)
name of organization (of Aff.)
Bochum, Germany
email address

Ioannis Iossifidis
dept. name of organization (of Aff.)
name of organization (of Aff.)
Bochum, Germany
email address

Christian Klaus
dept. name of organization (of Aff.)
name of organization (of Aff.)
Bochum, Germany
email address

***Abstract*—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.**

I. 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 such variations or anomalies, anomaly detector algorithm is designed and implemented on the provided data. 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 [1], occupancy [2] and operation of appliances [3], [4].

II. Related Works

Previous studies in data-driven building energy consumption prediction have utilized several methods such as Engineering methods [5], statistical method [6], Artificial Neural Networks (ANN) [7], [8], support vector machines (SVM) [9], fuzzy logic and grey model techniques [10], decision trees [11] 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. [9] compared SVM and Back Propagation Neural Network (BPNN); Borges et al. [12] compared SVM and Autoregressive (AR) Model; Xuemei et al. [13] compared LS-SVM and BPNN; Liu and Chen [14] compared Support Vector Regression (SVR) and ANN; Penya et al. [15] compared AR Model, ANN,autoregressive integrated moving average(ARIMA), and Bayesian Network; Platon et al. [16] compared ANN and Case based Reasoning (CBR); Jain et al. [17] compared SVM and MLR; (this paper linked from somewhere else) Hou et al. [18] compared ARIMA and ANN (need to download paper); Fan et al. [19] compared MLR, ARIMA, SVM, RF, MLP, BT, MARS, and kNN; Chou and Bui [20] compared ANN, SVM, CART, CHAID, and GLR; Edwards et al. [21] compared MLR, FFNN, SVM, LS-SVM, HME-FFNN, and FCM-FFNN; Li et al. [22], [23] compared SVM, BPNN, Radial Basis Function Neural Network(RBFNN), and General Regression Neural Network (GRNN); Dagnely et al. [24] [22] compared Ordinary Least Squared (OLS) Regression and SVR; Massana et al. [25] compared MLR, MLP, and SVR; and Fernandez et al. [26] compared AR, polynomial model, ANN, and SVM.

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. [27] 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 [28]. Whereas, Qi et al. [29] used stacked autoencoders to diagnose and classify the faults in a rotating machinery.

III. Methodology

A. Data Exploration

In order to select meaningful features to be implement Machine Learning or Statistical Modeling, 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

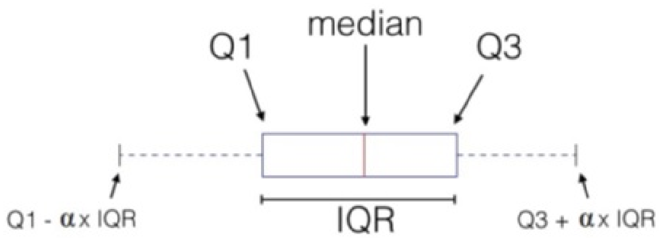


Fig. 1: Depiction of quartiles as used in the detection and removal of Outliers.

stores differs from 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 stores with 238 sensor values (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.

B. Statistical Method for Anomaly 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 Tuckey's test 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 Tuckey's Fences or Inner Fences by the following equation,

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

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

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 3.

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.

C. Autoencoders for Anomaly Detection

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 $\mathbf{z}$ 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 $\mathbf{z}$ to the reconstruction $\mathbf{x'}$ of the same shape as $\mathbf{x}$.

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

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 minimise reconstruction errors (such as squared errors):

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

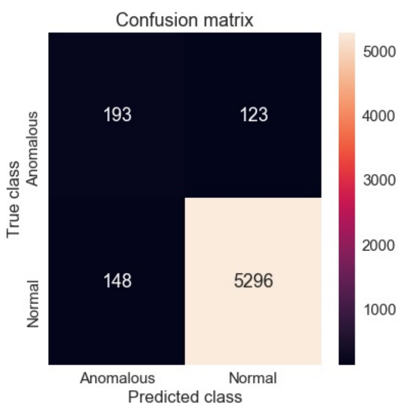
where, $\boldsymbol{x}$ is usually averaged over some input training set.

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

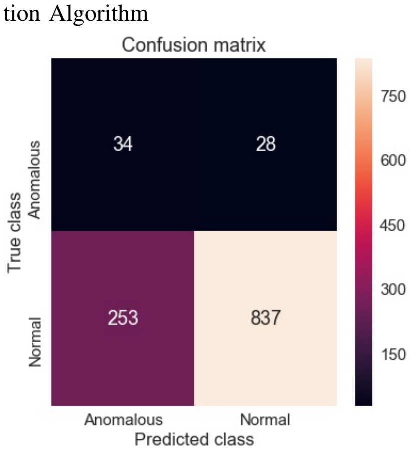
In our project, the Autoencoder was trained in a semi-supervised fashion on the values of normal energy consumption data. Afterwards, the trained model was used and evaluated on a pre-trained dataset.

D. Performance Comparison between Anomaly Detection and Autoencoder

In Fig 2, we are comparing the accuracy of prediction of both algorithms with manually classified energy data of two stores using confusion matrices.



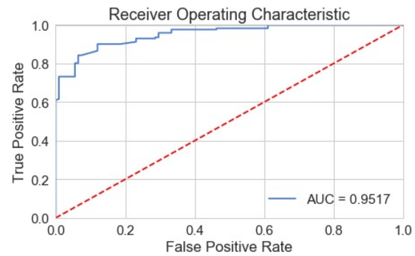
(a) Confusion Matrix of Anomaly Detection Algorithm



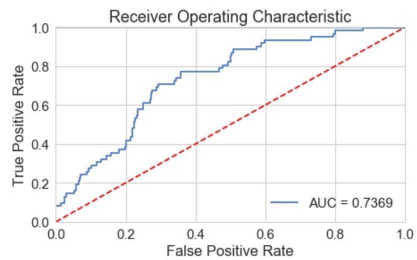
(b) Confusion Matrix of Autoencoder.

Fig. 2: Confusion Matrices

The same dataset which was used to make confusion matrices in Fig 2 is used to plot Receiver Operating Characteristics curves in Fig 3.



(a) ROC curve of Anomaly Detection Algorithm.



(b) ROC curve of Autoencoder.

Fig. 3: ROC curves

As we can see from the comparison of Fig 2 and Fig 3, we can interpret that the performance of Statistical method which is based on Tuckey's test performed remarkably better with around 95% similarity 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.

REFERENCES

[1] Liping Wang, Paul Mathew, and Xiufeng Pang. Uncertainties in energy consumption introduced by building operations and weather for a medium-size office building. *Energy and Buildings*, 53:152-158, 2012.

[2] X. Feng, D. Yan, and T. Hong, "Simulation of occupancy in buildings, energy and buildings," pp. 348-359, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/ S0378778806002404

[3] H. Chenglei, L. Kangji, L. Guohai, and P. Lei, "Forecasting building energy consumption based on hybrid pso-ann prediction model," in *2015 34th Chinese Control Conference (CCC)*, July 2015, pp. 8243-8247.

[4] M. Royapoor and T. Roskilly, "Building model calibration using energy and environmental data," *Energy and Buildings*, vol. 94, pp. 109 - 120, 2015. [Online]. Available: http://www. sciencedirect.com/science/article/pii/S0378778815001553

[5] H. xiang Zhao and F. Magouls, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586 - 3592, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/ S1364032112001438

[6] F. Lei and P. Hu, "A baseline model for office building energy consumption in hot summer and cold winter region," in 2009 *International Conference on Management and Service Science*, Sept 2009, pp. 1-4.

[7] B. Bekta Ekici and U. Aksoy, "Prediction of building energy consumption by using artificial neural networks," vol. 40, pp. 356-362, 05 2009.

[8] A. Kusiak, M. Li, and Z. Zhang, "A data-driven approach for steam load prediction in buildings," *Applied Energy*, vol. 87, no. 3, pp. 925 - 933, 2010. [Online]. Available: http://www. sciencedirect.com/science/article/pii/S0306261909003808

[9] Q. Li, Q. Meng, J. Cai, H. Yoshino, and A. Mochida, "Applying support vector machine to predict hourly cooling load in the building," *Applied Energy*, vol. 86, no. 10, pp. 2249 - 2256, 2009. [Online]. Available: http://www.sciencedirect.com/ science/article/pii/ S0306261908003176

[10] J. Guo, J. Wu, and R. Wang, "A new approach to energy consumption prediction of domestic heat pump water heater based on grey system theory," *Energy and Buildings*, vol. 43, no. 6, pp. 1273 - 1279, 2011. [Online]. Available: http://www. sciencedirect.com/science/article/pii/S0378778811000107

[11] G. K. Tso and K. K. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy*, vol. 32, no. 9, pp. 1761 - 1768, 2007. [Online]. Available: http://www.sciencedirect.com/ science/article/pii/S0360544206003288

[12] C. E. Borges, Y. K. Penya, I. Fernndez, J. Prieto, and O. Bretos, "Assessing tolerance-based robust short-term load forecasting in buildings," *Energies*, vol. 6, no. 4, pp. 2110-2129, 2013. [Online]. Available: http://www.mdpi.com/1996-1073/6/4/2110

[13] L. Xuemei, L. Jin-hu, D. Lixing, X. Gang, and L. Jibin, "Building cooling load forecasting model based on ls-svm," in *2009 AsiaPacific Conference on Information Processing*, vol. 1, July 2009, pp. 55-58.

[14] D. Liu and Q. Chen, "Prediction of building lighting energy consumption based on support vector regression," in *2013 9th Asian Control Conference (ASCC)*, June 2013, pp. 1-5.

[15] Y. K. Penya, C. E. Borges, D. Agote, and I. Fernndez, "Shortterm load forecasting in air-conditioned non-residential buildings," in *2011 IEEE International Symposium on Industrial Electronics*, June 2011, pp. 1359-1364.

[16] R. Platon, V. R. Dehkordi, and J. Martel, "Hourly prediction of a buildings electricity consumption using case-based reasoning, artificial neural networks and principal component analysis," *Energy and Buildings*, vol. 92, pp. 10 - 18, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/ S0378778815000651

[17] 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.* [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/ 9780784413616.208

[18] Z. Hou, Z. Lian, Y. Yao, and X. Yuan, "Cooling-load prediction by the combination of rough set theory and an artificial neuralnetwork based on data-fusion technique," *Applied Energy*, vol. 83, no. 9, pp. 1033 - 1046, 2006. [Online]. Available: http:// www.sciencedirect.com/science/article/pii/S0306261905001315

[19] C. Fan, F. Xiao, and S. Wang, "Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques," *Applied Energy*, vol. 127, pp. 1 - 10, 2014. [Online]. Available: http://www. sciencedirect.com/science/article/pii/S0306261914003596

[20] J.-S. Chou and D.-K. Bui, "Modeling heating and cooling loads by artificial intelligence for energy-efficient building design," *Energy and Buildings*, vol. 82, pp. 437 - 446, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/ S037877881400574X

[21] R. E. Edwards, J. New, and L. E. Parker, "Predicting future hourly residential electrical consumption: A machine learning case study," *Energy and Buildings*, vol. 49, pp. 591 - 603, 2012. [Online]. Available: http://www.sciencedirect.com/ science/article/pii/S0378778812001582

[22] Q. Li, Q. Meng, J. Cai, H. Yoshino, and A. Mochida, "Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks," *Energy Conversion and Management*, vol. 50, no. 1, pp. 90 - 96, 2009. [Online]. Available: http://www.sciencedirect.com/ science/article/pii/S0196890408003154

[23] Q. Li, P. Ren, and Q. Meng, "Prediction model of annual energy consumption of residential buildings," in *2010 International Conference on Advances in Energy Engineering*, June 2010, pp. 223-226.

[24] P. Dagnely, T. Ruette, T. Tourw, E. Tsiporkova, and C. Verhelst, "Predicting hourly energy consumption. can regression modeling improve on an autoregressive baseline?" pp. 105-122, 09 2015

[25] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, "Short-term load forecasting in a non-residential building contrasting models and attributes," *Energy and Buildings*, vol. 92, pp. 322 - 330, 2015. [Online]. Available: http://www. sciencedirect.com/science/article/pii/S0378778815001024

[26] I. Fernndez, C. E. Borges, and Y. K. Penya, "Efficient building load forecasting," in *ETFA2011*, Sept 2011, pp. 1-8

[27] Cheng Fan, Fu Xiao, Yang Zhao, and Jiayuan Wang. Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. *Applied Energy*, 211:1123 - 1135, 2018.

[28] J. Höchst, L. Baumgärtner, M. Hollick, and B. Freisleben. Unsupervised traffic flow classification using a neural autoencoder. In *2017 IEEE 42nd Conference on Local Computer Networks (LCN)*, pages 523-526, Oct 2017.

[29] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu. Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery. *IEEE Access*, 5:15066-15079, 2017