**DATA GENERATION – ELECTRICAL APPLIANCES CONSUMPTION DATA**

**Abstract**

Electrical appliance consumption data is very useful in planning and forecasting to aid efficient use and allocation of energy resource. Unfortunately, real world electrical appliance datasets are very scarce, expensive and difficult to collect due to the time and amount of resources needed to acquire the data.

This paper proposes to use a deep generative model such as Generative Adversarial Networks (GANs) and Variational Auto Encoders (VAEs) to generate new realistic synthetic electricity consumption data for different regions in the USA which can then be used with other statistical models for forecasting and analysis.

**Introduction**

Energy consumption is growing year on year worldwide with residential household electrical appliances making up about 34% of the total electricity consumption in the USA and is projected to rise to 39% by 2030 [1].

Therefore, to meet up with the growing demand of electricity, power companies need to build models to better forecast and plan the energy use; to achieve this they need an estimation of energy consumption at a household level. Energy consumption forecasting problem is a time series regression task, it consists of predicting the energy consumption for the future by using a customer‘s historical electricity usage data to produce accurate forecasts for planning and scheduling optimization.

Often in practical situations, the available historical data is not sufficient to capture the uncertainties in energy consumption; a major setback in modeling the electrical consumption is the limitation of robust real-time data where a large amount of historical/forecasted generation and consumption data is required. [2]

An approach to solving the above issue of limited historical data is to generate scenarios, modelling possible trajectories of different household electrical appliance. These scenarios can then be incorporated into the forecasting models for efficient planning and optimization.

This paper aims to use a Deep Generative Model for generating synthetic historical electricity data for residential households.The remainder of the paper is organized as follows:

* Section 2 includes a review on various Generative models and application
* Section 3 describes the proposed methodology
* Section 4 demonstrates the results.
* Section 5 concludes the study and discusses some future work.

**Review Literature**

Deep learning technology has been widely used to generate data in recent years mainly with application to images.

Two of the most commonly used Deep Generative models are Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN).

Generative Adversarial Networks (GANs) [3] has been successfully applied in many ways for data generation such as in images [14] and natural language [15].

VAEs have already shown promise in generating many kinds of complicated data, including handwritten digits [4, 5], faces [4, 6, 4], house numbers [8,6], CIFAR images [9], physical models of scenes [4], segmentation [7], and predicting the future from static images [11].

Conditional Generative Adversarial Networks (CGAN)[12], provided a way for GAN to generate samples conditioned on class labels[13]using a Generative Adversarial Networks (GAN), specifically a Deep Convolutional Generative Adversarial Networks based model to generate domestic solar production and electricity consumption scenarios, they also used a conditional Deep Convolutional Generative Adversarial Networks to generate synthetic data conditioned on site specific conditions.

Data disaggregation is another way Generative model are applied in generating electricity consumption, also known as energy disaggregation/non-intrusive load monitoring (NILM). This provides residents with an accurate view and understanding of their energy consumption and can potentially help in reducing the peak energy consumption and facilitating efficient usage and conservation of energy. It disaggregates the aggregated energy signal into individual appliance signals. A Variational recurrent neural network was used to learn the abstraction of the aggregated energy consumption over latent variables at training time and then generates all the individual appliance signals jointly by sampling from the latent variables at test time [17].

**Methodology**

**Dataset**

The dataset used was obtained from Pecan street dataset (DATAPORT). The data consists of energy consumption readings at 1-minute, 5-minutes and 1-hour interval. The data covers the period of 2013 to 2019, across seven states in the US and from 1534 houses.

In this project the 1-hour interval reading was obtained from houses in California, Colorado and Texas for the year 2015 which amounted to 1467 households. The model was trained using the Texas and Colorado data, while California Data used as the Validation set.

The data consists of reading for 68 appliances, but only the top 15 with the most energy consumption were selected.

**Data preprocessing**

**Data cleaning:** The dataset contained missing values, which were filled by grouping the categories by State, Month, and Time of the day and using the average to replace the missing values.

**Data Normalization:** To accelerate the training processes and to obtain good results, the dataset was normalized to a range between 0 and 1. These provided bounds for our data range and allows us to generate new solar and load data based on DC system sizes and peak consumption levels, achieving scale invariance.

**Generative Model**

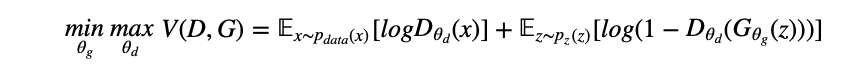
**Variational Autoencoders (VAE** ): VAE [4] is a regularized version of the traditional autoencoder (AE). It consists of two parts: an inference network that maps an input to a posterior distribution of latent codes and a generative network that tries to reconstruct the original input conditioned on the latent encoding. By imposing a prior distribution on the latent codes, the model enforces the distribution over the latent variable to be smooth. This enables proper sampling from the model via sampling from latent to input space.

The full objective of the VAE is written as:

**

**Generative Adversarial Network (GAN):** Generative Adversarial Nets (GAN)[4] is an adversarial process that involves two opposing networks, where the components (the generator and the discriminator) are neural net. The first net generates data, and the second net tries to tell the difference between the real data and the fake data generated by the first net. The second net will output a scalar [0,1]which represents a probability of real data.

Formally, the objective is written as:

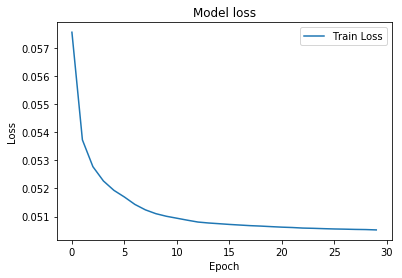
****

**Results**

**4.1. Model Training**

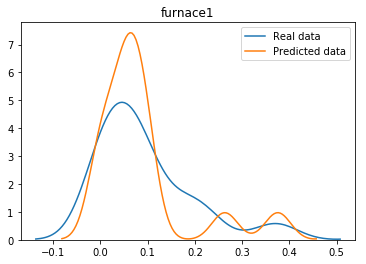
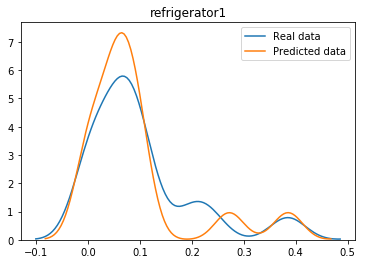
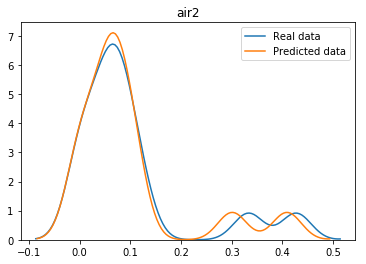
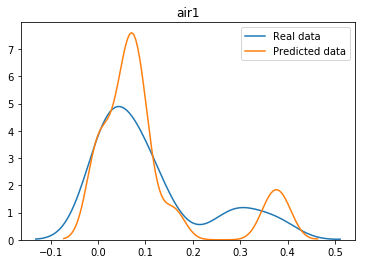
The models where trained using Texas and Colorado as training data and the California data was used to test the models.

Fig1. Training loss for Variational Autoencoder



The VAE models was trained for 30 epochs. During training the model performs very well at the beginning as the loss function begins to reduce. By the 15th epoch the loss begins to reduce.

Fig2. Sample of the result from the test set and the prediction from 4 appliances

****

From the figure above it is clear that model is performing well as the predicted results tries to match the distribution of the real data.

Fig3. Training loss for Generative Adversarial Network

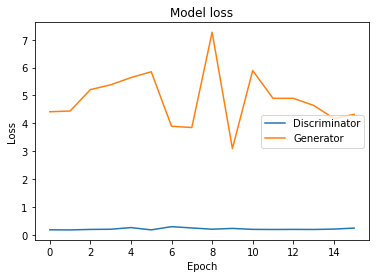
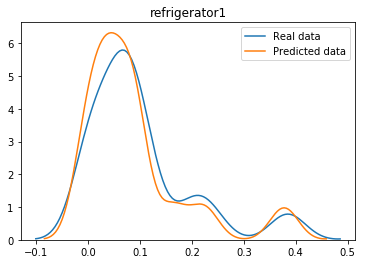
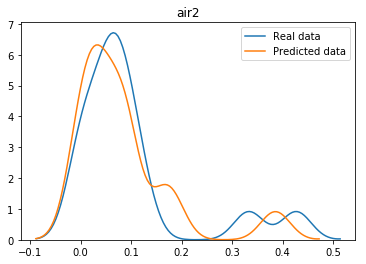
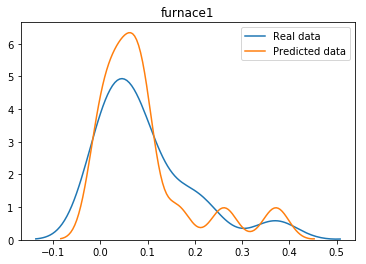
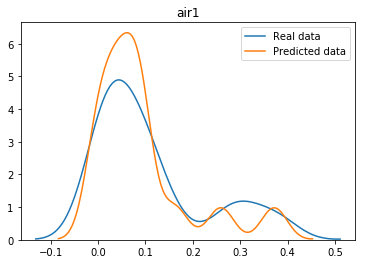
****

Fig4. Sample of the result from the test set and the prediction from 4 appliances

****

Despite the poor result of the GAN loss function the predicted results tries to match the distribution of the real data.

**Conclusions and Future Work**

This paper proposed a method to generate synthetic household electrical appliance energy usage scenarios using Variational Autoencoder and GANs.The trained model is able to produce synthetic that follow the original data distribution without memorizing and that can be used for further analysis.

In the future, explore the use of Conditional GANs and Conditional VAEs to produce data conditioned on a particular region/state. Also, add categorical variables such as building type, and city to generate them as they would be useful for forecasting and understanding the demographics of the regions.

**Reference:**

1. Ehrhardt-Martinez, K. *et al.* *Advanced metering initiatives and residential feedback programs: a meta-review for household electricity-saving opportunities* (American Council for an Energy-Efficient Economy Washington, 2010).

2. ElexonLoad profiles and their use in electricity settlement (2013)(London)  
 Retrieved from: <http://www.elexon.co.uk/wpcontent/uploads/2013/11/load_profiles_v2.0_cgi.pdf>[Google Scholar](https://scholar.google.com/scholar_lookup?title=Load%20profiles%20and%20their%20use%20in%20electricity%20settlement&publication_year=2013&author=Elexon" \t "_blank)

3. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In NIPS’2014.

4. Diederik P Kingma and Max Welling. Auto-encoding variational Bayes. ICLR, 2014.

5. Tim Salimans, Diederik Kingma, and Max Welling. Markov chain monte carlo and variational inference: Bridging the gap. ICML, 2015.

6. Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in Deep Generative models. In ICML, 2014.

7. Tejas D Kulkarni, William F. Whitney, Pushmeet Kohli, and Josh Tenenbaum. Deep convolutional inverse graphics network. In NIPS, 2015.

8. Diederik P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. In NIPS, 2014. 6. Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Rezende, and Daan Wierstra. Draw: A recurrent neural network for image generation. In ICCV, 201

9. Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In NIPS, 2015. 8. Jacob Walker, Carl Doersch, Abhinav Gupta, and Martial Hebert. An uncertain future: Forecasting from static images using variational autoencoders. In ECCV, 2016.

10. Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In NIPS, 2015.

11. Jacob Walker, Carl Doersch, Abhinav Gupta, and Martial Hebert. An uncertain future: Forecasting from static images using variational autoencoders. In ECCV, 2016.

12. Mirza, M. & Osindero, S., 2014. Conditional Generative Adversarial Nets. , pp.1–7. Available at: <http://arxiv.org/abs/1411.1784>.

13. Generating Residential PV Production and Electricity Consumption Scenarios via Generative Adversarial Networks Rui Tang, Philip H.W. Leong, Jonathon Dore and Anthony Vassallo

14. Salimans, T., I.Goodfellow, Zaremba, W., Cheung, V., Radford, A., and Chen, X. Improved techniques for training gans. In Proc. NIPS, pp. 2234–2242, 2016.

15. . Press, O., Bar, A., Bogin, B., Berant, J., and Wolf, L. Language generation with recurrent generative adversarial networks without pre-training. In arXiv preprint arXiv:1706.01399v2, 2017.

16. 2015. Pecan street inc., dataport.

17. Gissella Bejarano, David DeFazio & Arti Ramesh 2019. Deep Latent Generative Models For Energy Disaggregation