Predicting Global Power Consumption Based on Historical Trends using Long Short-Term Memory Networks

1st Alexander Pellegrino
Department of Electrical and Computer Engineering
Stevens Institute of Technology
Turnersville, United States of America
apellegr1@stevens.edu

2nd Anthony Paolantonio
Department of Electrical and Computer Engineering
Stevens Institute of Technology
Hoboken, United States of America
apaolant@stevens.edu

***Abstract*—Electricity is one of the basic needs for modern life. Without a functioning power grid we lack heat for our homes, clean water to drink and to wash ourselves with, and much of our current food production. As we as a species continue to increase our reliance on electronic devices to keep our society running, the demand for energy has increased rapidly around the world. Though the invention of newer, more energy efficient devices has helped make some progress in slowing that demand, we still put more and more strain on our grid each year. In this project, our group will attempt to model the changing demand for energy over time on a global scale, and based upon this model predict its future growth. For this we will make use of Recurrent Neural Networks, specifically the Long Short-Term Memory Model. While our model is still in its early stages, LSTMs have shown great promise in predicting trends as their nature allows them to more effectively overcome the "vanishing horizon" problem - where a Neural Network favors newer data over data it saw long ago. We aim to take advantage of this property to better model global power consumption in the future.**

I. Introduction

For well over a century, humanity's reliance and usage of electricity has continued to grow. Electricity has become integral to the way the world functions and, without its consistency and availability, much of the world's infrastructure would crumble. As a society that's almost entirely dependent on electricity, it is not only historically interesting, but incredibly important to evaluate how energy consumption has grown and will continue to grow moving forward. By predicting future trends, we can better scale our energy infrastructure for the future in order to avoid the cost of continual refinement (the "add as you go" approach), allowing us to instead build a grid which will be robust for years to come.

With this context in mind, this paper will discuss a method to develop a model based upon past energy consumption which can then be used to predict future electricity consumption trends. This model will utilize the Global Electricity Statistics (1980-2021) data set obtained from Kaggle[1]. With the help of this set, over forty years worth of data will be processed and used to train the machine learning model. Since the data set is now somewhat outdated, we can evaluate our model by comparing its predictions for 2022 and 2023 against the true global energy consumption for these recent years.

The aforementioned machine learning model will be based on the Long Short-Term Memory (LSTM) Network Model, a specific type of Recurrent Neural Network (RNN) that addresses the vanishing gradient problem encountered with most RNNs. This problem states that most RNNs eventually "forget" long-term dependencies within data despite its potential relevance to predictions. For example, forgetting words towards the beginning of a paragraph by the time it reaches the end in a Large Language Model. Even the now-famous ChatGPT exhibited this problem in its first public release, requiring a large amount of refinement in order to mitigate the issue. Even now, the solution is imperfect and by having a long enough conversation with the chatbot one can start to see it for themselves, with earlier sentences being forgotten as new prompts are given.

Similar projects have been attempted before. After all, figuring out energy consumption for the future is a matter of national importance to many governments, not just a fleeting interest for college research. One example of this is a study which focused on using the Support Vector Regression algorithm in the Support Vector Machines (SVM) model to predict the future energy consumption of China[2]. While the major difference which separates LSTMs from SVMs is the LSTM's ability to maintain long term memory of features, that is not the only benefit it offers over the existing approach. LSTMs are inherently designed with an orientation towards linear data, a trait which is not as well emphasized in SVMS. Other methods used to compute similar predictions include Multiple Linear Regression, Random Forest Regressors, Decision Tree Regressors, ARIMA modeling, and Gradient Boost Regressors. These machine learning algorithms are useful in terms of predicting trends based off training data, but do not have the same complexity as a Neural Network which allows for extrapolation of minor data points even across a large time frame.

II. Related Work

Despite this being one of the first projects published attempting to apply LSTMs to the prediction of electricity consumption, there exist a plethora of alternate studies which make use of other machine learning algorithms to achieve a similar solution. Meng Z, Sun H and Wang X published a paper entitled ”Forecasting Energy Consumption Based on SVR and Markov Model: A Case Study of China” in 2022 which made use of Support Vector Regression (SVR) to create Support Vector Machines[2]. These SVMs are however not designed inherently for linear data, have no mode of retaining long-term dependencies, and are less complex than a neural network, meaning that they inherently have a lesser ability to reason about subtle changes in data.

Another related study titled ”Modeling Energy Consumption Using Machine Learning” made use of a combination of Multiple Linear Regression, Random Forest Regression, Decision Tree Regression, Gradient Boost Regression, Support Vector Machines, Random Forest, K-Nearest Neighbor (KNN), and Deep Learning models in a hybrid approach[3]. While this hybrid algorithm is more advanced in terms of complexity, this does not always correlate to a better approach. While the complexity of the model does account for the subtle features discussed earlier, and the model did perform well, this massively hybrid approach means many moving parts and likely requires massive computational power for training. The more additional features one adds onto a model, the more one experiences diminishing returns and the higher the likelihood of failure due to a single component being out of line or an issue merging the resulting data. The hypothesis set to be proven throughout this report is that an LSTM-based approach has the potential to provide similar results in a much more concise form. However, it is also important to take into consideration the usage of different data sets, as the study was conducted using the U.S. Department of Energy Industrial Assessments Centers' data from 1981 until 2013, meaning one cannot directly compare the results with the prior study as it is, at this point, quite outdated and used a different source of information. Still, this project should be able to take advantage of both a decade more data and newer models in the hopes of finding an approach which is an improvement in simplicity, accuracy, or both, relative to the prior hybrid model.

Others who have attempted the challenge of predicting energy trends with Machine Learning include Gori and Takanen (2004)[4], who used a modified EDM model to predict the energy consumption of Italian industry, household and service. Ediger and Akar (2007)[5] used the ARIMA and seasonal ARIMA methods with energy data from Turkey from 2005 to 2020. Zong and Roper (2009)[6] suggested an ANN model for estimating Korea's energy demand. These tried and true general statistical models show great promise and with a fraction of the computational cost of Deep Learning, however they too suffer from the lack of granularity and long-term memory that exists within LSTMs. Lastly, these studies are once more specific to a target region or country, and do not provide a global snapshot as this study aims to accomplish. However, they are some of the best comparisons available for simpler approaches that still provide reasonable accuracy, which will likely be implemented in at least one of these approaches in order to compare their output of the data set against the LSTM model.

III. Our Solution

This project involves the use of a Long Short-Term Memory type Recurrent Neural Network to predict future global energy consumption based on historical trends ranging from 1980 to 2021. It will be used to predict the energy consumption for the years 2022 and 2023 according to the International Energy Agency[7] for testing. In addition to the primary LSTM model, two of the standard models for time series analysis will be used, ARIMA and Support Vector Regression, to compare the LSTM's performance against them on the same data.

A. Description of Dataset

The Global Energy Statistics (1980-2021) dataset is a fairly simply laid out one, with the data arranged into 44 columns and 1,611 rows. The first column corresponds to a country. The second states which statistic is being provided and can be used for pruning. In our case, it is used to prune the net energy consumption of the country so we can filter down to a global average. The third column tells what region the country is a part of. This corresponds to power grids, not necessarily specific continents, but doesn't really impact our training and is thus dropped. All remaining columns, 4-44 make up the numerical portion of the data, displaying statistics for each year from 1980 to 2021. Every country is repeated 7 times, with its first appearance correlating to its net power generation in billions of kilowatt-hours. The second is its net consumption, the third the imported energy, the fourth the exported. The fifth is the net import or export in total. The sixth corresponds to the maximum amount of electricity in millions of kilowatt-hours that the country is capable of producing. The seventh and final value is how much energy is lost in transmission or waste energy by the country. The net consumption of the country is the main data point being used, so the data is filtered down to only these entries for training.

While the first three columns are always properly formatted in this particular data set, the numerical data must be handled cautiously, as in some cases countries had unknown power consumption, marked as – in the data set. There are also a few empty cells which result in NaN (not-a-number) values. In others the country was known not to have any energy infrastructure, leaving 0s in the data. These cases need to be handled in order to avoid bias or parsing invalid data (since – for example is not a number). Since our current approach uses a sum, the 0s are simply allowed to propagate to the total as they won't impact the overall result. The – are consequently replaced with 0s during the initial file parsing. However, this approach may need to be modified if we go back and decide to add country-specific data to the model. Because countries have wildly varying power draws based on their level of development, the global sums must be calculated prior to normalization, lest larger countries drown out smaller ones or smaller countries be unfairly biased in training. Several models were attempted without this, as summation loses granularity, however all proved to introduce far more error than accuracy.

B. Machine Learning Algorithms

Rather than a hybrid model or a purely mathematical one, structures which are often used for the prediction of electricity usage over time, it was decided to stick with a singular LSTM network for the sake of this project. Partially, this is to reduce the complexity of the project within the duration of the semester and to allow the algorithm to run faster. However, the focus of this paper lies also in evaluating the performance of an LSTM for this particular problem and gauging whether or not it is sufficient in predicting trends, as there was not much research to be found on the topic, and none of it was from recent years. LSTMs are highly optimized for use on sequential data, making them an ideal candidate for the scenario of looking at yearly values in an attempt to model future trends. Because the nature of the data (yearly samples) makes it discrete, it allows easy use of a Neural Network. In addition, the recurrent nature of the LSTM model also means it is less likely to forget older data as it learns on newer data. The algorithm should effectively be able to view the entire series and adjust accordingly. Since the TensorFlow and Keras libraries are being used for this, all parameters will be left as their default values in Keras. After this, experimentation with tweaking values such as the loss function, optimization function, and normalization will be done once consistent results from the program are achieved. These default values are, at the time of writing: a singular hidden layer with 64 neurons and a tanh activation function, followed by a Dense layer. The sparse categorical cross-entropy function will be utilized to deal with loss, and the stochastic gradient descent function for optimization.

The second algorithm being implemented for comparison is the ARIMA model. While not a Machine Learning model in the strictest of senses, it is still one of the most commonly used for time series analysis due to its robustness[5], relatively high degree of accuracy, and ease of use. This makes it an ideal candidate for comparison, as a considerable amount of time will not have to be invested into heavily optimizing a secondary model while still getting fairly accurate results. In addition, since it is one of the simplest approaches to implement, it will be highly telling whether the LSTM model can make more accurate predictions or not. If the LSTM network cannot outperform even the simple ARIMA implementation, then it doesn't make sense to pursue it further in this particular case. Most users would likely end up going with the simpler and faster approach instead to begin with, so for the LSTM approach to be worth using it MUST have a higher degree of accuracy in order to warrant the time it would take to implement.

The third and final algorithm which will be implemented, again as a benchmark and as a sanity check, is the Support Vector Machines model, powered by the Support Vector Regression algorithm. This is an approach which has been used successfully in the past for this exact same problem on a smaller scale[2] and thus should make a good baseline for comparison; one which, in theory, should provide a good middle ground between the speed and simplicity of the ARIMA model and the large complexity and effort of the LSTM approach. It is slower than ARIMA but faster than training the LSTM, more complex than ARIMA but less complex than LSTM, and should provide middling accuracy. However, it can also be quite error prone and may prove less accurate than either approach without careful tuning and optimization, such as via Grid Search.

C. Implementation Details

For this project, the first step was to properly prune and format our data in order to predict global trends. To do this, all the rows in the dataset correlating to details other than net consumption of electricity within countries were filtered out and dropped. This left only the data pertaining to energy consumption, allowing a summation to provide a relatively accurate view of global demand which could be used to train the model. Missing cells or strings notating countries without a power grid were handled by backfilling the data in those cells with 0s. Because the goal is to calculate a net global power consumption, 0s can be safely used to provide a numeric value which won't influence the outcome. This is due to the model's final prediction being a summation, and because a country without a power grid inherently doesn't consume any power. Once the data was properly pruned, the initial model was trained on the raw dataset. However, it proved to be highly unstable. Of the seven distinct regions, six of them would consistently fail to train, resulting in NaN outputs, with only one region providing any numeric data. The first attempt to rectify this involved normalizing the data to value between 0 and 1, as large jumps, such as between low consumption rural countries and first world countries like the USA, can cause instability in an LSTM network due to the "exploding gradient" problem. (Where attempting to correct for jumps causes the model to overshoot and run off the end of the valid range of data.) After applying min-max normalization, the model became a bit more stable, but still five of the seven regions would consistently yield NaN results. To further stabilize the model, several different optimization functions were attempted. The default was stochastic gradient descent, however the adam optimizer proved far more robust, allowing three or four regions to train depending on the run and starting seed value. Still, the model was too unstable to provide consistent output. Adjustments were made to the learning rate and batch size, however neither had any significant impact on the stability of the model, with the only notable difference being that reducing the batch size below 64 had a considerable impact on the training speed, causing it to spike drastically. Through careful filtering the cause of the remaining instability turned out to be a few NaN values introduced into the dataset during the normalization process applied earlier. The ordering of the operations was shifted around so that normalization happened prior to NaN corrections, defaulting to treating NaN values as 0s for the normalization process. This allowed the model to train successfully across all regions. However, despite successfully training, the model experienced unstable output, giving results anywhere from negative values through predictions expecting energy to triple in a single year then return to normal the next. Further experimentation with parameters proved ineffective in rectifying this in the current format of the data. Thus the data was simplified in order to facilitate easier pattern identification for the model. Rather than splitting the data into separate regions and then merging the final predictions, the data was summed prior to training the model, giving a net sum for each year. This approach proved far more stable at the cost of losing regional granularity, finally allowing the model to consistently give fairly accurate outputs often within $\tilde{5}\%$ of the actual global energy consumption for 2022 and 2023 according to the International Energy Agency [7]. Through testing different activation functions, it was found that switching from Stochastic Gradient Descent to the Hyperbolic Tangent Function brought the accuracy to within $\tilde{3}\%$ on average, and adjusting the epochs, learning rate, and batch size eventually landed a final combination of 75 epochs, a batch size of 16, and a learning rate of 0.001 resulting in an average error of around $\tilde{1}\%$, with only seconds of training time required on high-end consumer hardware. (The model was trained using Keras 2.15.0 on the NVIDIA RTX 4090 GPU.)

The second model was far simpler to implement, with the ARIMA functionality provided by the statsmodels python library and the data already properly formatted using the same approach as for the LSTM model. For the sake of consistency, the summed data was used in this approach as well, to ensure that all models were trained using the same input data. To optimize the ARIMA model, the Autocorrelation and Partial Autocorrelation functions were calculated on the dataset for the base case, first differential, and second differential. By observing the properties of the data, it was determined that p, d, q values of 1, 1, and 0 were the most optimal, though several other values were tested experimentally to confirm. With these optimal values the ARIMA model consistently outputs the same results with every run, with an error of around $\tilde{1}.1\%$. This put it within reach of the LSTM model's accuracy, though the final tuned LSTM performed slightly better in almost every case, only very rarely having comparable error in the worst training runs. Still, the LSTM never showed a higher level of error than ARIMA once tuned, and does still have merit when the absolute smallest margin of error is desirable. For less precise cases, ARIMA still holds its rank near the top.

The third and final model was also fairly simple to implement, with Support Vector Regression functionality provided by the scikit-learn python library. Once more the pre-formatted data was taken from the LSTM model to ensure consistency across all tests. Of all the models tested during this paper, this one stood out for the ratio of its difficulty to work with versus its accuracy on this type of problem. Whereas ARIMA worked almost effortlessly, quickly, and could give the LSTM network a run for its money in terms of precision, the SVR-based SVM model required extremely careful formatting of the data, use of a custom scaler, and necessitated the use of Grid Searching to find the optimal hyperparameters for the model. Even after all this, it still provided relatively abysmal results compared to the other two models, consistently yielding predictions with a margin of error greater than 55%. No amount of tweaking the scaling, increasing the range of hyperparameters checked in the Grid Search, or re-shuffling the data was able to improve this outcome.

IV. Comparison

While basic details of the models were provided above, the final performance of each model after much careful refinement and fine tuning shows one which clearly trails behind the other two for this use case, and two that show serious contention for the top spot depending on the user's focus on speed of implementation versus level of accuracy. In terms of accuracy, the overall winner was the LSTM network model by a narrow margin, as it consistently performed with the least margin of error relative to the true values. However, getting there required hours of tuning hyperparameters, refining the input data, and ensuring proper normalization and rescaling of the normalized outputs back to human-legible format. ARIMA on the other hand was a close second, never falling more than half a percent behind the LSTM network in terms of its prediction accuracy despite requiring very little tuning, far less computational power, and having far few parameters to adjust. If the user doesn't need the absolute cutting edge of precision, it is likely still one of the fastest and computationally cheapest algorithms to get relatively high accuracy estimations for future values in time series data. The Support Vector Regression based Support Vector Machines model is the only one which, for this particular use case, proved to have little merit. Regardless of the amount of tweaking and refinement made to the input parameters, it never started outputting predictions with an acceptable margin of error, while at the same time proving more difficult to implement and work with than ARIMA. While still simpler than the LSTM network approach, the comparatively high amount of effort to get it working paired with the relatively low prediction accuracy invalidates it as a contender for this particular dataset.

V. Future Directions

The current solution is fairly rudimentary and does not utilize the full wealth of information available from the dataset, instead aiming for a merged approach. It analyzes the overall trend of global power consumption in one lump sum for each year, while the dataset provides information on a country-by-country basis. While it would add further complexity to the model, something which this paper aimed to mitigate in order to provide a better alternative to existing hybrid approaches, bringing in this added information could provide the model with deeper insights which it could use to more accurately predict the trend as a whole by predicting the future growth, reduction, or stagnation of individual countries. It would also help to show trends such as how some countries gained access to electricity later than others, meaning that if there are examples of things such as "booms" that typically occur a certain number of years after a country gets access to electricity, the model would be able to predict those. (Though this is just an example and not something visually observable just by looking at the data. Such a trend would be within the hidden layer if found.) It would also be wise to test the LSTM implementation provided against hybrid models, as in this paper it was mainly compared against solo models for the sake of conciseness and due to limitations in time and resources. The end goal is to prove its effectiveness as a replacement for hybrid approaches, and while percentage wise it seems promising, a definitive proof cannot be made without direct comparison on the same dataset.

VI. Conclusion

The findings in this paper show that LSTM networks do indeed have promise in terms of predicting energy usage on a global scale. While not beating the conventional ARIMA model by a significant margin when both are highly optimized, an LSTM does still grant an increased level of accuracy in its predictions over ARIMA, while maintaining the need to only train and optimize a single model, as opposed to the multiple often used in past hybrid approaches[3]. This offers benefits in training speed, ease of maintenance, and ease of improvement over hybrid approaches while still giving a high level of accuracy which could be utilized for certain applications. While the improvement is marginal and ARIMA can still certainly be useful when a simple and non-computationally-expensive approach is needed, for those times when as much accuracy is needed as possible with as little training as possible LSTMs could properly fill that niche.

REFERENCES

[1] A. Jethwa. Global energy statistics (1980-2021). [Online]. Available: https://www.kaggle.com/datasets/akhiljethwa/global-electricity-statistics

[2] S. H. Meng Z and W. X. Forecasting energy consumption based on svr and markov model: A case study of china. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fenvs.2022.883711/full

[3] P. T. Sarswatula SA and P. V. Modeling energy consumption using machine learning. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmtec.2022.855208/full

[4] F. Gori and C. Takanen. Forecast of energy consumption of industry and household & services in italy. [Online]. Available: https://art.torvergata.it/handle/2108/50421

[5] V. S. Ediger and S. Akar. Arima forecasting of primary energy demand by fuel in turkey. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0301421506002291?via%3Dihub

[6] W. G. Zong and W. E. Roper. Energy demand estimation of south korea using artificial neural network. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421509003218?casa_token=nzpIlDC0Il4AAAAA:MGpX3aIF0RoPRCfIC4eW1IBHzf3uBfmO7CQ7UaBJ38J5aCn_dNBA4v1Eqq5Z98Vw3U7zYljw

[7] International energy agency energy system. [Online]. Available: https://www.iea.org/energy-system/electricity