Related Work

Sean Klein, Kevin Russell, Ataish Nehra Northeastern Data Science Capstone
Supervised by: Dr. Fatema Nafa

1 Literature.

1.1 Problem Addressed:

Climate Change is one of the most important concerns for humanity that led to an increase in frequency of natural disasters. David Phiri's contribution [1] towards predicting natural disasters in Zambia has made an impact towards understanding and detecting natural disasters. Sang-Jin Park researched similar machine learning algorithms [2] to predict coastal flooding risk under climate change impacts. However, a more recent article on deep neural networks to predict global climate change [3] has shown promise in furthering predictions beyond traditional machine learning models.

1.2 Methodology:

Long Short-Term Memory: In the formulation of our LSTM models for time series forecasting, particularly in the context of climate prediction, our methodology has been significantly informed and shaped by the foundational principles and applications discussed in pivotal scholarly works. Two such pieces of literature that have notably influenced our approach include "A long short-term memory-based model for greenhouse climate prediction" by Liu et al. [4], and "Understanding LSTM–a tutorial into long short-term memory recurrent neural networks" by Staudemeyer and Morris [2].

The study by Liu et al. [4], published in the International Journal of Intelligent Systems, introduces an LSTM-based model tailored for the accurate prediction of greenhouse climate variables. This work underscores the LSTM's capacity to capture complex temporal dependencies within environmental data, a characteristic paramount to forecasting in natural disaster scenarios. By applying an LSTM model to greenhouse climate prediction, Liu et al. demonstrate the efficacy of this approach in environmental monitoring, which parallels our project's aim to predict natural disasters using climate-related time series data. Their methodology, involving the integration of LSTM networks for predictive modeling in climate-sensitive environments, has provided a blueprint for harnessing LSTM architectures in our forecasting tasks.

Further grounding our understanding of LSTM networks, the tutorial by Staudemeyer and Morris [5] offers an in-depth exploration of LSTM recurrent neural networks and their advantages for long-term dependency modeling. This comprehensive guide through the operational mechanisms of LSTM networks elucidates the theoretical underpinnings necessary for implementing effective LSTM models. The clarity provided by Staudemeyer and Morris in articulating the nuances of LSTM networks, including their structure and function, has been instrumental in informing our model design and optimization strategies.

Leveraging the insights from these studies, our models are built upon the robust capabilities of TensorFlow Keras, facilitating the development of LSTM networks designed to forecast complex sequential patterns in climate data. This methodology, inspired by cutting-edge research and practical applications of LSTM in climate prediction and understanding, positions our work within the broader discourse on LSTM's transformative potential in time series forecasting.

Attention: The second models for the project was an LSTM model with an added attention transformer. The attention layer first proposed by google in the article 'Attention is all you Need'[6] is a transformer architecture that is built using attention mechanisms.This type of model has been shown to be able to solve similar time-series forecasting problems, the article 'Time Series Forecasting Using Attention Mechanism' [7] shows how to implement an attention neural network solve this kind of problem. The model that was build for this project was built independently from this article, however, the findings of this article helps to justify the choice of this architecture. The architecture of the model was built using TensorFlow Keras [8], a python library, created by google engineer Francois Chollet, that helps to simplify the process of constructing neural networks.

Forward Forward: Neural networks has traditionally used gradients calculated by backpropagation [9] to correct errors, but the Forward-Forward algorithm [10] has shown promise in overcoming the limitations tied to backpropagation, which provides a more biological learning algorithm mimicking the frontal cortex. This new algorithm has paved way towards different applications of the algorithm [11]–[13]; extending the original concept, such as sleep deprivation, feature extraction for skin lesions, or optical neural networks. However, Gandhi's most recent work [14] extended the Forward-Forward algorithm beyond vision to perform sentiment analysis classification. In this extension, Gandhi shows that a singular linear layer can learn information from a concatenation of Forward-Forward activation layers to classify sentiment of words. A slightly similar work by Brenig [15] shows that the Forward-Forward algorithm can extend to self-supervision by using a linear classifier to train on the frozen backend of the Forward-Forward network.

1.3 Application Area:

In the natural disasters area, machine learning is revolutionizing the detection and frequency of natural disasters with climate change data [1], [2] by leveraging K-Nearest Neighbors, Random Forest, Support Vector Machines, and other algorithms to predict natural disasters. EL-HABIL attempted to improve natural disaster prediction by using a LSTM cell to predict global climate change [3].

2 Key Findings.

2.1 Highlights & Results:

Unfortunately, David Phiri's, Sang-Jin's, and EL-HABIL's works used classifications to predict natural disasters in regions of the world whereas our work predicts the increased frequency of natural disasters for the top 30 countries. Their metrics used Accuracy, AUC, and classification evaluation methods for model evaluation whereas our work focuses on Mean Squared Error and Mean Absolute Error for model evaluation. We couldn't find any related research papers on predicting total disaster frequency as a time-series or continuous value prediction.

2.2 Benchmarks and Datasets:

The data was sourced from the International Monetary Fund's six public datasets on climate change and natural disasters [16]. The IMF collaborated with multiple different reputable organizations in order to collect and aggregate the data into an easy to access web page. The Surface Temperature dataset was sourced from the Food and Agriculture Organization Corporate Statistical (FOASTAT)[17] Database and collected by the National Aeronautics and Space Administration's[18] Goddard Institute; Carbon Dioxide Concentrations were provided by the National Oceanic and Atmospheric Association(NOAA)[19] Global Monitoring Laboratory; Sea Level Changes were from the NOAA's Laboratory for Satellite Altimetry; Forest and Carbon data was collected by the IMF in collaboration with FOASTAT; Land Cover is from FOA land cover with additional work done by the IMF; and the Natural Disasters data is sourced by EM-DAT [20] which is an international database specifically for disasters and their epidemiology.

Unfortunately, the IMF dataset is not a commonly used dataset in the Data Science community as there are currently a handful of articles available on google scholar. However, Tsui-Fang Hu and team did utilize the dataset to forecast inflation under globalization [21]. In his paper, they utilized Root Mean Squared Error and Mean Absolute Error as benchmarks to evaluate their ANN model. Besides this article, we couldn't find others in the Data Science community that leveraged this climate change data. However, typical Data Science articles on time-series and continuous values has traditionally used Mean Absolute Error [22], Root Mean Squared Error [23], and/or Mean Absolute Percentage Error [23].

3 Gaps and Limitations.

3.1 Related Research:

Although David Phiri's performed extensive research with various machine learning models, he used oversampling, SMOTE, that is known to introduce noise and bias towards the fitting process. There wasn't a trial of using under-sampling, which has been known to improve the fitting process for class-imbalance datasets. Sang-Jin Park does perform similar research to David Phiri's research, but he does not list whether the training used cross-fold validation or train/test/valid dataset validations. Additionally, there is very little evidence to whether the information is over-fitted or under-fitted. EL-HABIL's architecture of LSTM does use State of the Art architecture, but the results seem too good to be true. There is little evidence that the model was fitted on a true validation/test dataset, and the lack of additional metrics like a confusion matrix, F1 Score, and etc... prove that there are gaps in the validation process.

3.2 Dataset:

One of the limitations of this project was the dataset[16]. The 6 datasets used had limitations of the amount of years and rows available for analysis, as well as countries not being accounted for datasets like Sea Levels and CO Concentrations. Sea Levels had to be manually adjusted for align and CO Concentrations had to be dropped as a whole. The challenge with the sea levels made it hard to generalize the project to include more countries. This made the dataset limited, so the models did not have a large amount of data to work with. This means that noise and outliers from the dataset has a greater impact on the prediction.

3.3 Long Short-Term Memory:

Despite the considerable advancements in time series forecasting using LSTM networks, as detailed in works by Yu et al. [24], Liu et al. [4], and Natel de Moura et al. [25], a recurrent theme in the gaps and limitations revolves around the customization and optimization of LSTM architectures for specific forecasting scenarios, especially under changing climate conditions. Yu et al.'s comprehensive review [24] underscores the LSTM's prowess yet hints at the challenge in fine-tuning and architecturally adapting LSTMs for nuanced applications. Liu et al. [4] exhibit a tailored LSTM application for greenhouse climate prediction, showcasing the potential yet not fully exploring LSTM's adaptability across varied climate scenarios. Similarly, Natel de Moura et al. [25] evaluate LSTM's efficacy in discharge prediction under climate variability, indicating room for enhanced LSTM models that can dynamically adjust to a broader spectrum of environmental factors. Addressing these gaps, our work sought to refine LSTM model architecture and training processes, aiming for a model that not only learns from historical data but also adapts more fluidly to emerging climate patterns. This involved a deeper investigation into LSTM's internal mechanisms and leveraging advanced techniques for model optimization, thereby pushing the boundaries of LSTM's application in climate-related forecasting.

3.4 Attention:

One limitation of the Attention Neural Network was its construction using TensorFlow Keras[7]. There is a trade off when using keras: its ease of use is in sacrifice of having more control over the models architecture and design. A python library such as PyTorch[26], may have been a better choice for control but it is a harder to use library; and since Sean was newer to working with Neural Networks, he decided to use the easier to implement library.

3.5 Forward-Forward:

The Forward-Forward has had several interesting extensions beyond vision classification. The most interesting of extensions is Gandhi's work [14] in extending the Forward-Forward algorithm towards sentiment analysis classification.

However, the method is limited as a classification problem that uses one-hot encoding of correct (positive data) and incorrect (negative data) classification labels. It is not a true sentiment analysis of providing a continuous value polarity. There has yet been a true method to extend towards predicting continuous values with the Forward-Forward algorithm. Additionally, the method still relies on correct labels to hot-encode positive goodness to learn information from the word2vec input.

4 Our Work.

Our work deviates from David's, Sang-Jin's, and EL-HABIL's work of classifying natural disasters by predicting Total Disasters, which is a time-series and continuous problem instead of a classification problem. Predicting Total Disasters, therefore, provides awareness of natural disasters through magnitude and time. This provides different evaluation metrics and data pre-processing to the models that we fit. Additionally, we fit different kinds of Neural Networks that are unique to the models fit by David's, Sang-Jin's, and EL-HABIL's work.

Our LSTM model represents an innovative approach to time series forecasting, particularly in predicting climate-related natural disasters. This model is an evolution of traditional neural network methodologies, leveraging the advanced capabilities of Long Short-Term Memory (LSTM) cells to effectively learn and remember long-term dependencies in sequential data. Distinguishing itself from conventional models, our LSTM model integrates a sophisticated architecture that includes multiple LSTM layers to deepen the learning process, coupled with dropout layers to prevent overfitting. This structure ensures a robust understanding of complex temporal patterns within the data, making it exceptionally suited for the nuanced task of forecasting natural disaster occurrences based on climate indicators. In a bid to optimize accuracy and efficiency, the model underwent rigorous training and validation phases, employing state-of-the-art optimization algorithms and loss functions tailored for time series analysis. The culmination of these efforts is a highly reliable LSTM model that not only outperforms standard forecasting techniques but also opens new avenues for leveraging deep learning in environmental science and disaster preparedness strategies.

The attention neural network is a modernization of the classic neural network models that incorporates a newer type of mechanism that leverages attention during the learning process. The model created is an exploration of this architecture's ability to train and predict on a unique type of time series problem. Additionally, the attention neural network explored regularization with dropout to simplify the network, which improves upon the Attention All you Need's Multi-Attention Head mechanism and LSTM cell combination.

The Forward-Forward algorithm addresses Ghandi's issue by extending the Forward-Forward algorithm to predict continuous values. The Forward-Forward Continuous network does not rely on correctly guided targets with inputs for positive data. With this new Forward-Forward algorithm extension, a true prediction of continuous values can be performed with comparable Mean Squared Error and Mean Absolute Error of a traditional neural network with similar architecture.

5 Up to Date References

[1] D. Phiri and C. Chembe, "Predicting climate change related extreme natural disasters using machine learning in zambia," *Zambia ICT Journal*, vol. 6, no. 1, pp. 18–24, Dec. 2022.

[2] S.-J. Park and D.-K. Lee, "Prediction of coastal flooding risk under climate change impacts in south korea using machine learning algorithms," *Environmental Research Letters*, vol. 15, no. 9, Sep. 2020.

[3] B. Y. EL-HABIL and S. S. ABU-NASER, *GLOBAL CLIMATE PREDICTION USING DEEP LEARNING*, vol. 100, no. 24, pp. 4824–4838, Dec. 2022.

[4] Y. Liu, D. Li, S. Wan, *et al.*, "A long short-term memory-based model for greenhouse climate prediction," *International Journal of Intelligent Systems*, vol. 37, no. 1, pp. 135–151, 2022.

[5] R. C. Staudemeyer and E. R. Morris, "Understanding lstm– a tutorial into long short-term memory recurrent neural networks," *arXiv preprint arXiv:1909.09586*, 2019.

[6] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, *Attention is all you need*, 2017. arXiv: `1706.03762 [cs.CL]`.

[7] subhradipta Choudhury. "Time series forcasting using attention mechanism." (), [Online]. Available: `https://www.analyticsvidhya.com/blog/2023/06/time-series-forecasting-using-attention-mechanism/`. accessed on: 03/20/2024.

[8] F. Chollet. "Keras: The high-level api for tensorflow." (), [Online]. Available: `https://www.tensorflow.org/guide/keras`. accessed on: 03/20/2024.

[9] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by backpropagating errors," in *Nature*, vol. 323, 1986, pp. 533–536.

[10] G. Hinton, *The forward-forward algorithm: Some preliminary investigations*, 2022. arXiv: `2212.13345 [cs.LG]`.

[11] M.-T. Lică and D. Dinucu-Jianu, *Sleep deprivation in the forward-forward algorithm*, 2023. arXiv: `2310.18647 [cs.AI]`.

[12] A. Reyes-Angulo and S. Paheding, *The forward-forward algorithm as a feature extractor for skin lesion classification: A preliminary study*, 2023. arXiv: `2307.00617 [cs.CV]`.

[13] I. Oguz, J. Ke, Q. Wang, *et al.*, *Forward-forward training of an optical neural network*, 2023. arXiv: `2305.19170 [cs.LG]`.

[14] S. Gandhi, R. Gala, J. Kornberg, and A. Sridhar, *Extending the forward forward algorithm*, 2023. arXiv: `2307.04205 [cs.LG]`.

[15] J. Brenig and R. Timofte, *A study of forward-forward algorithm for self-supervised learning*, 2023. arXiv: `2309.11955 [cs.CV]`.

[16] I. M. Fund. "Climate change data." (), [Online]. Available: `https://climatedata.imf.org/pages/climatechange-data`. accessed: 03.07.2024.

[17] U. Nations. "Foa: Food and agriculture organization of the united nations." (), [Online]. Available: `https://www.fao.org/faostat/en/#home`. accessed on: 03/14/2023.

[18] U. S. F. Government. "Nasa: National aeronautics and space administration." (), [Online]. Available: `https://www.nasa.gov/`. accessed on: 03/14/2024.

[19] U. D. of Commerse. “Noaa: National oceanic and atmospheric administration.” (), [Online]. Available: `https://www.noaa.gov/`. accessed on: 03/14/2024.
[20] CRED and T. W. H. Organization. “Em-dat: The international disaster database.” (), [Online]. Available: `https://www.emdat.be/`. accessed on: 03/12/2024.
[21] T.-F. Hu, I. G. Luja, H. Su, and C.-C. Chang, “Forecasting inflation under globalization with artificial neural network-based thin and thick models,” in *World Congress on Engineering and Computer Science, USA*, Citeseer, 2007, pp. 909–914.
[22] K. Iuchi, R. Miyazaki, G. C. Cardoso, K. Ogawa-Ochiai, and N. Tsumura, “Remote estimation of continuous blood pressure by a convolutional neural network trained on spatial patterns of facial pulse waves,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Jun. 2022, pp. 2139–2145.
[23] X. Song, Y. Liu, L. Xue, *et al.*, “Time-series well performance prediction based on long short-term memory (lstm) neural network model,” *Journal of Petroleum Science and Engineering*, vol. 186, p. 106 682, 2020.
[24] Y. Yu, X. Si, C. Hu, and J. Zhang, “A review of recurrent neural networks: Lstm cells and network architectures,” *Neural computation*, vol. 31, no. 7, pp. 1235–1270, 2019.
[25] C. Natel de Moura, J. Seibert, and D. H. M. Detzel, “Evaluating the long short-term memory (lstm) network for discharge prediction under changing climate conditions,” *Hydrology Research*, vol. 53, no. 5, pp. 657–667, 2022.
[26] A. Paszke. “Pytorch.” (), [Online]. Available: `https://pytorch.org/`. accessed on: 03/20/2024.