Drought Prediction using Deep Learning Methods

Shamika Dhuri
Robotics & AI for Agriculture
Carnegie Mellon University
Pittsburgh, PA, USA
sdhuri@andrew.cmu.edu

***Abstract*—Drought prediction is essential for farmers looking to prevent against crop loss due to drought, especially as climate change is increasing the likelihood of drought. After a review of previous literature, it seemed as though combining different datasets for drought prediction could prove to be useful. However, it was found that this was not the case. It was found that an LSTM architecture was idea for predicting drought based on Palmer Drought Severity Index data.** [1] [2]

I. Introduction

Drought is increasingly becoming a problem for farmers in the recent years. As can be seen from Figure 1, the level of drought in the US over the past 20 years has significantly increased [3]. Without water, agriculture suffers and farmers are unable to produce enough food to financially provide for themselves and to provide for the nation's food needs. Because of this, drought prevention and methods for dealing with drought are heavily researched. Farmers can choose drought resistant crops, plant cover crop, change their growing season slightly, or create more climate controlled environments to grow in to prevent drought from devastating all of their crops [4].

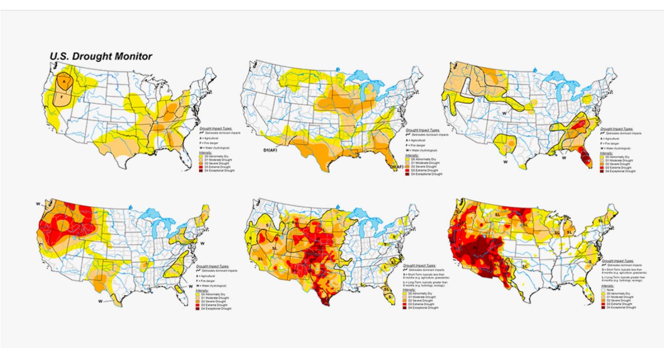

Fig. 1. US drought monitoring as it changes over years. Darker colours are areas with more drought.

Though drought prevention and resistant measures are available, without knowing when or where a drought will occur, farmers are unable to put these measures in place early enough for them to be useful. By the time that drought is obvious, it is often too late to start planting drought resistant crops or build climate controlled environments. The solution here is predicting drought using environmental metrics so that measures can be taken well before a drought occurs.

II. Review of Literature

Before deciding the baseline model architecture and the data to be used for training, review of literature was done to determine what the state of the art was at the moment.

A. Data

Drought prediction seems to be most effectively done by choosing a metric that measures something in the environment, like precipitation or soil moisture, and using past data for an area to predict the future values of that metric. Then, once values are predicted, the metric is used to determine how likely a drought is to occur in that area.

Though several metrics can be used, the most popular in literature seem to be the Palmer Drought Severity Index, or PDSI [6], [9], and the Standardized Precipitation Index, or SPI [7], [8]. This is reasonable as they are both easily publicly available.

B. Model Architectures

After determining which datasets to use for prediction, literature was examined to determine which method of prediction yielded the best results. Several statistical [9] and machine learning methods were used for this task in previous literature. Machine learning methods were found to be more robust. Within machine learning, there were a variety of models used. Some of of the more popular ones were random forest ensemble methods [6], hidden Markov models [7], support vector machines [11], Long sort-term memory networks [8], and Artifical Neural Networks [10]–[12]

C. Baseline Model

In previous literature, it seems as though often, Machine learning methods were used on isolated datasets where only one metric was used to predict drought. However, several different metrics exist. They are just not being used by the same models.

After reviewing literature, though artificial neural networks with PDSI data were the most popular, it was determined that the baseline model for this paper would be an LSTM using PDSI data as it seemed like this combination was the best that previous literature had found. The idea with this was to first

[1]Code and data for this project can be found here: https://github.com/ShamikaD/Drought-Prediction.git

[2]Information about experimentation run can be found here: https://wandb.ai/sd-team/drought_prediction-ablations/table?nw=nwusersdhuri

improve on the LSTM model for prediction and then choose another metric to add to the training data to improve on the prediction accuracy. It is logical to conclude that training on more data is likely to give better results. The second metric chosen was the second most popular metric, SPI.

III. Data and Data Processing

The two datasets that were used for this project were the Standardized Precipitation Index (SPI) and the Palmer Drought Severity Index (PDSI). Both are made available by the US government and cover the years from 1895 to 2016 [13], [14]. SPI data is better for shot term prediction as it is based only on precipitation and is collected fairly frequently [5]. PDSI data is better for longer term prediction as it is calculated less frequently and from a multitude of factors, including surface air temperature using a physical water balance model [2].

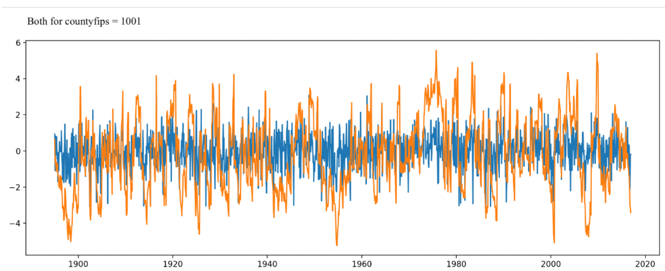

Fig. 2. Represented as time series data, the above figure shows how the PDSI data (orange) and SPI data (blue) for the county with the ID 1001 look when plotted over the time period given by the dataset

As can be seen from Figure 2, which shows both SPI and PDSI data from the country 1001 as time series data, SPI data has significantly more variance. This is reasonable as SPI data is collected more frequently, which means the curve will be less smooth when plotting the data. This also means that the SPI data is more variable from year to year because the same exact time over several years may have significantly different amounts of precipitation even though the general time of year is similarly likely to have a drought. Additionally, because SPI does not include factors other than precipitation while PDSI is calculated from a multitude of factors, including surface air temperature using a physical water balance model [2], even if the soil is still moist or there is an abundance of ground water, the index may still predict a drought when PDSI may not.

High PDSI and SPI values indicate that there is not likely to be a drought in the area while low numbers indicate that there is likely to be a drought. As can be seen, PDSI data is smoother and has less variation over short time frames.

IV. Methods

A. Baseline Model

After review of literature, it was decided that the baseline model should be an LSTM (long short-term memory) network with 2 LSTM layers trained on PDSI data to predict PDSI for the future. Training was done on the first portion of the time series data and validation was done on the held out last (or

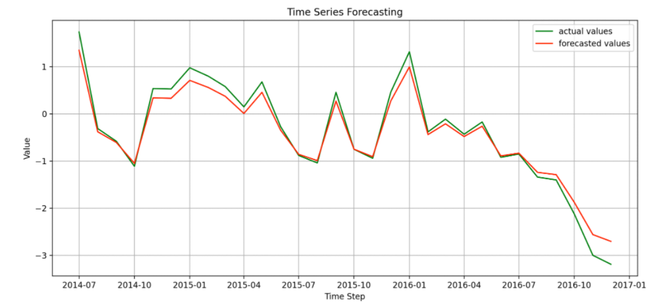

Fig. 3. Baseline model performance

closest to the current date) of the time series data. The results of this model for a short time period can be seen in Figure V-A and the same model on a longer sequence for prediction can be see in Figure 4.

B. Improved PDSI model

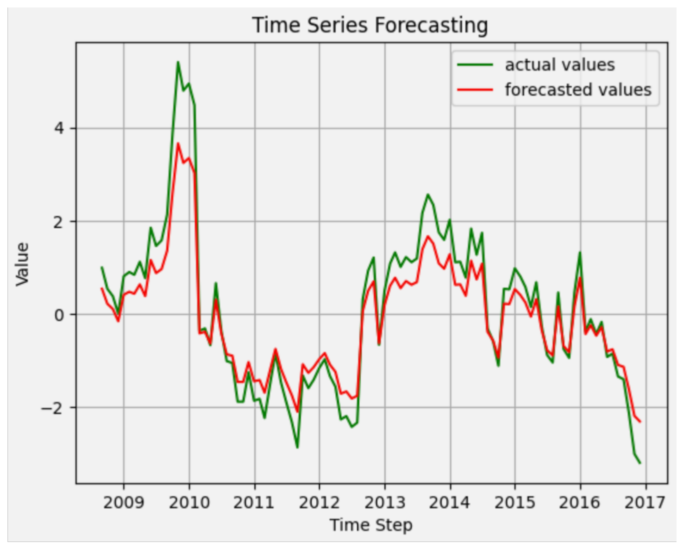

Fig. 4. Baseline model performance on longer time sequence

Improving on the PDSI model, 2 more LSM layers were added along with 0.5 dropout before the fully connected layer. This improved performance significantly as can be seen in Figure 5. It was found during experimentation here that 20 epochs or less were ideal for training as over fitting to the training data was a major problem in all models. Dropout helped with this to some extent, but no enough that training could continue much further.

C. Combined SPI and PDSI model

After looking at how the model performed using only PDSI data for training, it was decided that adding more data for training could improve model performance. The same LSTM model was used, but both SPI and PDSI data were used for training. Then, two predictions were created, one predicting the SPI using the combined data, and another predicting the

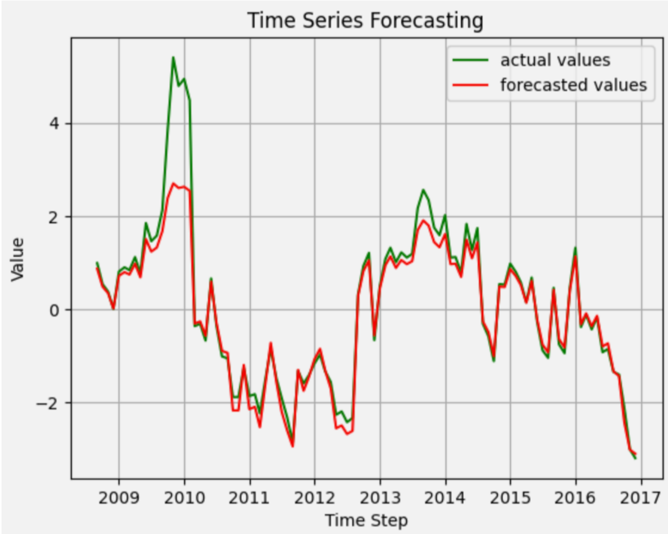

Fig. 5. Improved model performance on longer time sequence

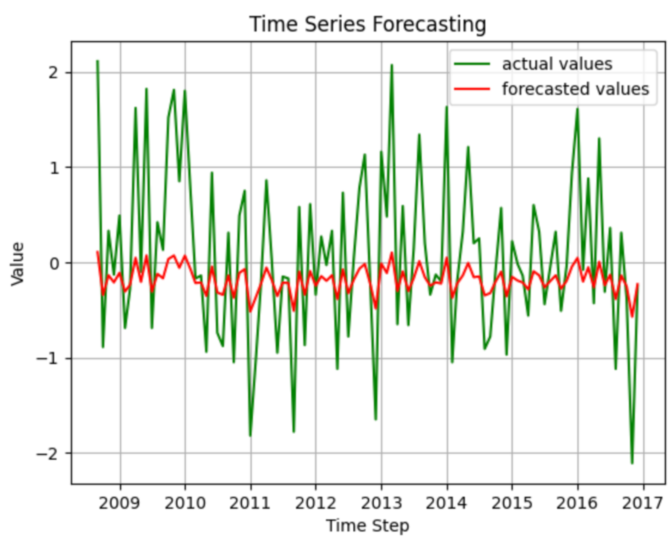

Fig. 6. SPI model performance

PDSI using the combined data. Both performed worse than expected.

D. Final Model

As the results of the PDSI model were the best, it was decided that the final model for this paper would be the best PDSI model. The model was a LSTM network. It was found that a network with four LSTM layers and a dropout of 0.5 before the fully connected layer performed the best for PDSI data prediction. An architecture diagram of the network can be found in Figure 8.

V. Results

A. Baseline Model

The baseline model, as can be seen from Figure , gave fairly accurate predictions over a short time period. Validation

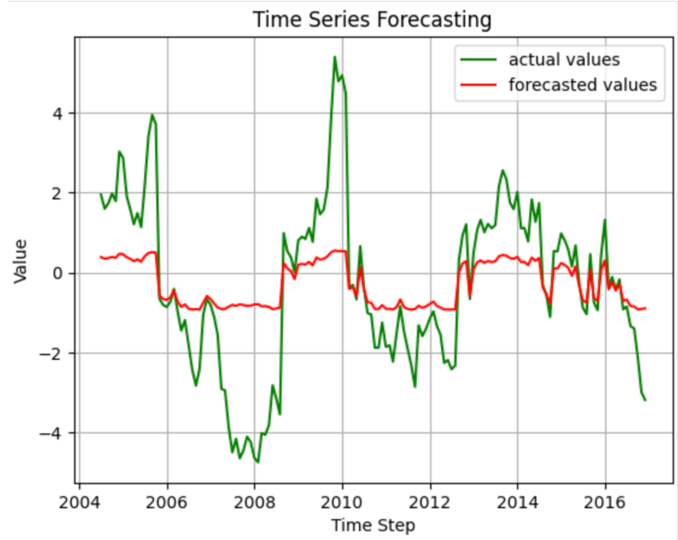

Fig. 7. Improved model performance on PDSI with SPI data added for training.

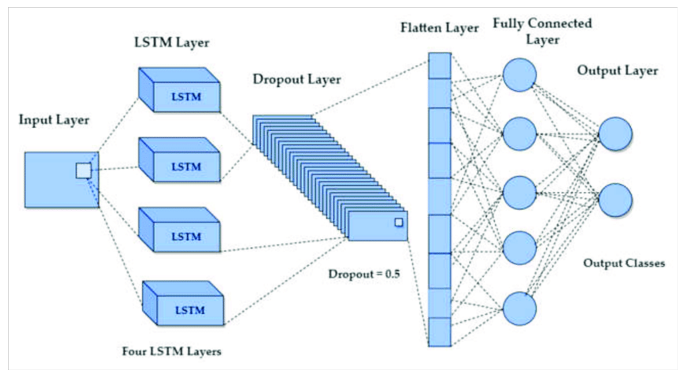

Fig. 8. LSTM model architecture

was only done on 30 data points in this Figure and the predictions were compared to the real data at the end of the time series. The model's performance was not ideal because it only performed well for short term prediction. Having longer term prediction would be more beneficial for drought prediction as it could allow farmers more time to prepare for potential droughts. Additionally, for when PDSI claimed that droughts were very likely (the index was a vert low number), the model did not predict the drought as well as it could have. This could cause issues if farmers were to use this model depending on what threshold of PDSI numbers they were keeping for deciding when a predicted drought was worth preparing for.

B. Improved PDSI model

Looking at Figures 4 and 5, it can be seen that model performance increased significantly. The mean squared error loss for validation was 0.8418 for the baseline model and it decreased to 0.2309 for the improved model. Additionally, as can be seen from the figures, the model tended to perform better on predicting valleys in the data (times of drought), even though it performed slightly worse on peaks (times where drought is very unlikely to occur). This error is preferred to the opposite because prediction of high humidity is not necessary while accurately predicting drought is the main goal of the model.

C. Combined SPI and PDSI model

Using combined data for predicting SPI and PDSI performed worse than expected. The results from these experiments can be seen in Figure 6 and 7. It is suspected that the SPI data significantly decreased performance of the model because of the amount of variance that it introduced. This made it so that the model was more likely to try and guess the average SPI or PDSI score rather than trying to fit a peak or a valley because the likelihood of a peak or a valley being in the opposite direction of a prediction was high. So, the loss would be minimized when a score of near 0 was always predicted. The MSE (mean squared error) loss for the model when predicting SPI was 0.6508 and when predicting PDSI was 0.5126.

VI. Conclusion

In conclusion, this paper shows that PDSI data is sufficient for drought prediction with an optimized model. Additionally, adding additional metrics like SPI for model training does not seem to improve model performance. For future experimentation, this methodology should be tried again with metrics that are more similar and have less variance, like the Crop Moisture Index (CMI) [1].

Acknowledgment

This paper was written as a final project for a Carnegie Mellon University course, Robotics and AI for Agriculture (16765), offered by the University's Robotics Institute.

References

[1] "Measuring Drought." Measuring Drought — National Drought Mitigation Center, drought.unl.edu. Accessed 25 Apr. 2024.

[2] "Climate Data Guide." Palmer Drought Severity Index (PDSI) — Climate Data Guide, 19 Aug. 2023, climatedataguide.ucar.edu/climate-data/palmer-drought-severity-index-pdsi.

[3] "Map Archive." Map Archive — U.S. Drought Monitor, droughtmonitor.unl.edu/Maps/MapArchive.aspx. Accessed 25 Apr. 2024.

[4] "Drought Resistant Practices." Drought Resistant Practices — USDA Climate Hubs, www.climatehubs.usda.gov. Accessed 25 Apr. 2024.

[5] "Global Gridded Standardized Precipitation Index (SPI) from CMORPH Daily." Drought.Gov, www.drought.gov

[6] Liu, Yan, and Yeonsang Hwang. "Improving drought predictability in Arkansas using the ensemble PDSI Forecast Technique." Stochastic Environmental Research and Risk Assessment, vol. 29, no. 1, 27 July 2014, pp. 79–91, https://doi.org/10.1007/s00477-014-0930-3.

[7] Paulo, Ana A., and Luis S. Pereira. "Prediction of SPI drought class transitions using Markov chains." Water Resources Management, vol. 21, no. 10, 6 Jan. 2007, pp. 1813–1827, https://doi.org/10.1007/s11269-006-9129-9.

[8] Poornima, S., and M. Pushpalatha. "Drought prediction based on SPI and Spei with varying timescales using LSTM recurrent neural network." Soft Computing, vol. 23, no. 18, 6 June 2019, pp. 8399–8412, https://doi.org/10.1007/s00500-019-04120-1.

[9] Dehghan, Sheida, et al. "Prediction of meteorological drought in arid and semi-arid regions using PDSI and SDSM: A case study in Fars Province, Iran." Journal of Arid Land, vol. 12, no. 2, Mar. 2020, pp. 318–330, https://doi.org/10.1007/s40333-020-0095-5.

[10] Banadkooki, Fatemeh Barzegari, et al. "Multi-timescale drought prediction using new hybrid artificial neural network models." Natural Hazards, vol. 106, no. 3, 9 Feb. 2021, pp. 2461–2478, https://doi.org/10.1007/s11069-021-04550-x.

[11] Mokhtarzad, Maryam, et al. "Drought forecasting by Ann, ANFIS, and SVM and comparison of the models." Environmental Earth Sciences, vol. 76, no. 21, 28 Oct. 2017, https://doi.org/10.1007/s12665-017-7064-0.

[12] Bari Abarghouei, Hossein, et al. "Prediction of drought in dry lands through Feedforward Artificial Neural Network abilities." Arabian Journal of Geosciences, vol. 6, no. 5, 25 Nov. 2011, pp. 1417–1433, https://doi.org/10.1007/s12517-011-0445-x.

[13] "Palmer Drought Severity Index, 1895-2016 - Dataset by US-HHS-Gov." Data.World, 22 Mar. 2021, data.world/us-hhs-gov/c74010cf-8b86-4be4-bd25-5366d8004d82.

[14] "Standardized Precipitation Index, 1895-2016 - Dataset by US-HHS-Gov." Data.World, 22 Mar. 2021, data.world/us-hhs-gov/ea3a3271-8895-410c-8cc7-2b214496969b.