Fuzzy Systems and Data Mining IX
A.J. Tallón-Ballesteros and R. Beltrán-Barba (Eds.)

doi:10.3233/FAIA231102

Global Warming Analysis Based on Long Short-Term Memory and Extreme Gradient Boosting Feature Engineering Models

Haoxuan LIU[a], Lanhua ZHANG[a] and Yujuan LI[a,1]
[a] *School of Medical Information and Engineering, Shandong First Medical University and Shandong Academy of Medical Sciences, Tai'an, 271016, China*

Abstract. Global warming brings negative impacts to people's lives. In order to explore the factors that contribute to global warming and predict temperature changes, engineering models using Long Short-Term Memory and Extreme Gradient Boosting algorithms were built. With the global year-average temperature data after preprocessing by Python, the mathematical models were computed to discover the factors that influence the global temperature change with temperature parameters and fluctuations of anomalies, including the global average temperature, carbon dioxide concentration, solar activity, etc. With the multi-variable linear differential equation, we conclude that the global climate is determined by a variety of factors. The results implied that a high concentration of greenhouse gases had a very limited impact on climate change, and solar activity had the greatest impact on global temperature.

Keywords. Global warming, XGBoost model, LSTM model, gray correlation analysis, feature engineering

1. Introduction

Global change is the term used to describe worldwide changes in the functioning of the Earth system due to natural and human factors, including changes in atmospheric and ocean circulation, biogeochemical cycles, the water cycle, the carbon cycle, resources, land use, urbanisation and economic development. Global warming is a prominent symbol of global change, and the Greenhouse Effect, resulting from massive emissions of greenhouse gases caused by human activities [1].

The causes of global warming include both natural factors and human activities. Solar activity, volcanic activity and multi-scale vibration within the climate system may affect global or regional temperature changes [2]. The influence of solar activities on climate change also includes ultraviolet radiation and solar magnetic field. One view is that with the solar activity cycle, solar irradiance changes little but ultraviolet radiation changes greatly, and the influence of ultraviolet radiation changes may be amplified through the absorption of the ozone layer in the middle atmosphere. Another

[1] Corresponding Author, Yujuan LI, School of Medical Information and Engineering, Shandong First Medical University and Shandong Academy of Medical Sciences, Tai'an, 271016, China; E-mail: 13605487932@163.com.

view is that cosmic rays contribute to the formation of condensation nuclei, and condensation nuclei can increase cloud cover [3].

Global warming is caused by human activities mainly through changes in the properties of the underlying surface, changes in certain components of the atmosphere [4], and anthropogenic heat release. However, natural factors have also had a great influence on the inter-annual temperature changes in the last hundred years [5].

The essence of machine learning is to simulate human learning through computers to find and learn the hidden laws behind them from a large amount of data, so as to achieve the purpose of simulation and prediction [6]. It has been widely used in commercial fields such as search engines, machine translation, spam filtering, speech recognition and other commercial fields, and also provides new methods for academic research in climate research [7, 8] and other fields. In recent years, with the increasing maturity of machine learning techniques, temperature prediction techniques based on machine learning have also been developed.

Long Short-Term Memory is a neural network that has the ability to remember long and short-term information. LSTM was first proposed by Hochreiter & Schmidhuber [9] in 1997 resulting in a more systematic and complete LSTM framework with the rise of deep learning. Another deep learning algorithm is XGBoost which was proposed by Tianqi Chen in 2016 [10] based on GBDT.

In the field of meteorology, Le [11] et al. achieved spatial and temporal prediction of air pollution at the second level with high accuracy by using LSTM with convolutional neural network to model a wide range of environmental variables. Nidhin [12] et al. constructed the ConvLSTMSR model by combining ConvLSTM with SR module to analyse the most climatically diverse India with better ability to predict extreme events. Seyed Matin Malakouti [13] used LSTM to predict the climate characteristics with a variety of LSTM types. Li [14] et al. combined three algorithms of linear regression, random forest and Extreme Gradient Boosting to establish a model for estimating biomass based on forest types. There are advantages in aboveground estimation, and the XGBoost and RF models significantly improve the estimation accuracy.

In order to further characterise the global warming scenario and to assist non-specialists in understanding global climate change trends, we disucuss the global temperature change and the related factors affecting the temperature using LSTM and XGBoost based on feature engineering analysis and other models for mutual validation to provide some references to ameliorate the global climate change and to achieve the healthy and green sustainable development strategy.

2. Methods

2.1. Model Assumptions

Suppose that the Earth's ecosystem can remain relatively stable in the future, and there will be no major or serious plate movements and geological activities. Assuming that there will be no major breakthroughs in human science and technology, and the main energy used in the future will still be similar to the existing energy sources. Assume that no new anthropogenic and non-anthropogenic factors affecting the Earth's climate will emerge in the short term. Suppose that the data used in this paper are true and reliable and can accurately reflect the rule of global climate change.

2.2. Data sources and Algorithm Models

In order to make a comprehensive and detailed analysis of the global warming problem, we construct LSTM and XGBoost feature engineering analysis model based on temporal features to carry out controlled experiments. Visualisation of the data can lead to a clearer and more explicit conclusion related to global warming.

The datasets are the Berkeley dataset (http://berkeleyearth.lbl.gov/) and the NOAA dataset (https://psl.noaa.gov/data/) from which data on temperature changes in 100 cities from 1833 to 2013. 81 of them are cities in the northern hemisphere and 19 are cities in the southern hemisphere. Some temperature data of some months or years is missing, and the Lagrangian interpolation method was used to interpolate the missing values. It is worth mentioning that there is a unit of measurement of temperature anomalies in meteorology as the distance level. The distance level is the difference between one of a series of values and the mean, divided into positive and negative. The mean temperature distance level is the difference between a series of mean temperatures and the total mean temperature. An increase in the mean temperature level is an indication that this difference has increased and that there is an anomaly in the temperature for the corresponding time period.

2.3. Work flow of our works

For LIST model, we standardize the data, build the model structure, define three LSTM layers, and output a fully connected layer. We adopt the average temperature of each month in the first 5 years as the input parameter of the model with the input dimension of (60, 1).

For XGBoost model, a feature engineering model based on time series and machine learning algorithm is designed. We take the average temperature of each month around the world to reflect the global temperature. Then we calculate the average monthly temperature values. To filter classification problems, machine learning strategies can be used. After filtering the collected data, we use the XGBoost algorithm in machine learning for testing.

About the realization in programming, the data set is read first, and then we use Tensorflow and Keras libraries in deep learning to predict it. The data sets are used for training and regression prediction. After that, the time series analysis is carried out to derive the model parameters and formulas. Finally, we compare the accuracy of the two models by calculating R2_score.

3. Results

With the LSTM model, the forecast results simulating the past temperature to predict the temperature from October 2022 to December 2100 are as follows (Fig. 1, Fig. 2):

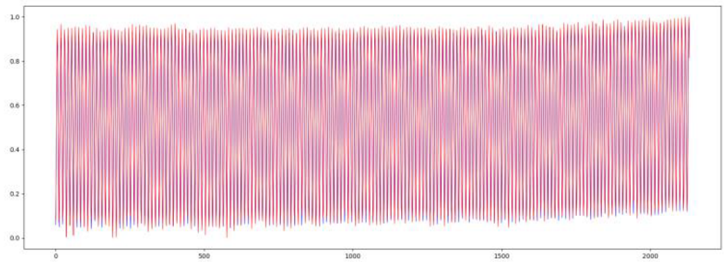

Figure 1. LSTM's past temperature fitting map

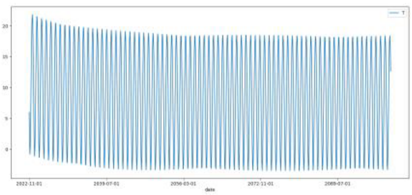

Figure 2. LSTM's temperature forecast image

The LSTM model is a very good fit to reflect the temperature change. According to the forecast results, the average monthly temperature in 2050 and 2100 will be below 20 degrees Celsius. Using Python programming, the temperature around the year 2135 was predicted to reach 20 °C. The visualization results are as follows (Fig. 3):

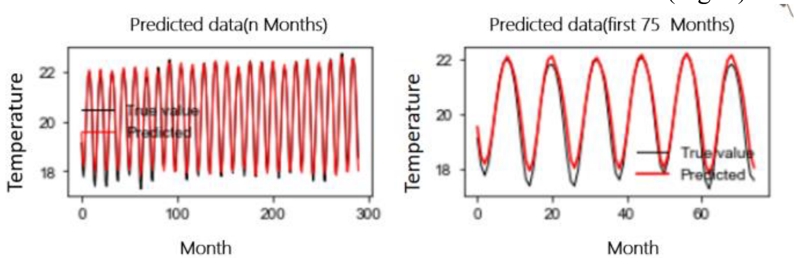

Figure 3. The temperature prediction in 2135 with LSTM model

For XGBoost feature engineering based on ARIMA, the system automatically searched for the optimal parameter based on AIC information criterion with the temperature variable. The model result was the ARIMA model (0,1,1) test table based on 2 different data (Fig. 4).

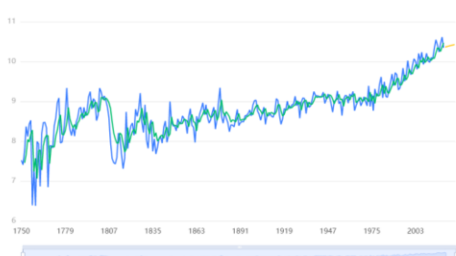

Figure 4. The simulation results of ARIMA model

Based on the time series-feature project, the searched data set is used. The prediction results are as follows (Fig. 5):

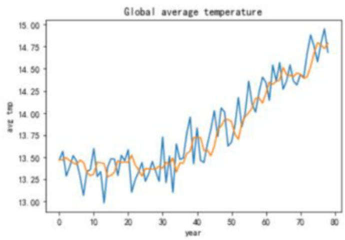

Figure 5. The training effects of XGBoost model

After the training, we can get a more accurate estimate of future global temperature with the XGBoost algorithm. The prediction results are as follows (Tab. 1):

Table 1. The prediction results with XGBoost model

Date	Time series	Feature Engineering
2050	15.205377	15.2560791
2100	15.968084	16.0983494

It is concluded that the average temperature at the global observation point in 2050 or 2100 does not reach 20.00°C.

Combining the running global temperature with future temperature projections, we can conclude that global temperatures have been and will continue to rise.

4. Discussion

According to the future global temperature forecast by LSTM and XGBoost feature engineering models, the global temperature will reach 20℃ around 2135. The temperature will not exceed 20℃ before then. Compared with model accuracy, the R2 value of LSTM is 0.97, which is higher than 0.95 of XGBoost feature engineering. Therefore, LSTM has higher accuracy in long-term temperature prediction than XGBoost feature project. At the same time, both models confirm that global temperatures have been and will continue to rise.

Through the quadratic curve fitting of K sequence and E2 sequence, the resulting parameters are shown in Table 2, and the fitting degree is obviously high. It is clear that the last hundred years the CO2 concentration is relatively strict rise. This is very different from the wild fluctuations in temperature changes, that is, the effect of greenhouse gases on the climate will be much smaller than previously estimated. Although the greenhouse gases will contribute to global warming, it is not feasible to describe global temperatures in terms of greenhouse gas concentrations alone.

Table 2. Concentration fitting table of greenhouse gas concentration growth curve

parameters	K	E2
R^2	0.999	0986
p	313.643	282.873
a	0.012	00.083
b	0.796	0.002

In addition to greenhouse gases, solar activity also has an effect on temperature. The influence of solar activity on global temperature is mainly reflected in the change of irradiance. To quantify the total solar energy received at the top of the atmosphere, one of the most commonly used metrics is total solar irradiance (TSI). Therefore, we use TSI as a variable and use TSI sequence to describe solar activity. We choose the T I sequence made by Kopp et al. (hereinafter referred to as T sequence for reference) and the sequence made by ourselves as J sequence. Pearson coefficient was used to show the correlation between TSI and global temperature. Comparing T series and J series, the Pearson correlation coefficient of the two sequences was 0.540, indicating a certain correlation with global temperature change. Compared with solar activity, the Pearson correlation coefficient of J series and T series is 0.846, indicating a high correlation between TSI and global mean temperature. Since solar activity has a large influence on the global average temperature, it should be limited to about 11 years based on the results. At the same time, according to the solar anomaly sequence data from 1854 to 2004, it can be found that the direction of J-M series and T-M series is the same in most areas, but the variation trend is slightly different in some areas, which may be affected by other factors such as geological activities. Using TSI to examine solar activity and discuss its relationship with global temperature changes, we can learn that solar activity has a high correlation with long-term changes in global temperature, and that solar activity has increased significantly over the past 100 years, which is consistent with global temperature increases. This suggests that solar activity has a significant influence on global temperature change.

The epidemic also has an impact on global warming. Energy-related CO2 emissions rose 6% to 36.3 billion tonnes in 2021, the largest year-on-year increase in energy-related CO2 emissions on record. The 2021 rebound reverses a decline of nearly 1.9 billion tonnes in emissions due to the 2020 pandemic. On top of this, CO2 emissions in 2021 are about 180 million tonnes higher than pre-pandemic levels in 2019. As global warming worsens, massive emissions of greenhouse gases will reduce the benefits of previous carbon cuts. Therefore, we believe that while COVID-19 related shutdowns have mitigated global temperature change to some extent, the impact of subsequent resumption of work on global temperature change has not only made up for the reduced emissions during the pandemic, but the emissions will further exacerbate global warming.

Based on the analysis of the global surface air temperature and the CO2 concentration TSI series in recent years, we draw a conclusion different from the traditional view:

(1) In the periodicity of global temperature, the current global temperature is at the peak of the wave and the change of global temperature in the last hundred years may not be caused by the increase of greenhouse gas concentration.

(2) Global climate is determined by a combination of factors, none of which can fully describe changes in temperature.

(3) Significant increases in greenhouse gas concentrations caused by human activities since the industrial revolution have had a very limited impact on climate change. It can neither determine the trend of global climate change nor fully describe global climate change.

(4) Solar activity has the greatest influence on global climate, and TSI has a strong correlation with the sliding average series of temperature.

5. Conclusions

In this paper, the LSTM and XGBoost feature engineering models were constructed for the prediction and analysis of global temperature from the perspective of global warming. It is concluded that the global temperature will not be more than 20 degrees before 2100, but the temperature will continue to rise. The mathematical methods and models used have undergone several iterations and the processing logic is excellent.

The use of Lagrange interpolation to interpolate the missing values in the data can improve the prediction accuracy of the data.

Based on the analysis results, we have found that emissions of greenhouse gases such as carbon dioxide were only part of the causes of rising global temperatures, and that solar activity has the greatest effect on global temperature. Other natural factors, such as geological and oceanic activity, also have some effect on temperature, but the effect cannot yet be estimated due to uncertainties. At the same time, the research and discussion in this paper is still incomplete. Although solar activity is strongly correlated with global temperature, there are still some periods unclear such as the negative correlation, moreover, the reasons for that are not yet known. Meanwhile, what effect CO2 concentration has on global temperature and how many other natural factors play roles in global temperature change need to be further discussed.

Acknowledgements

This research was supported by the Students' project for innovation and entrepreneurship training program of Shandong First Medical University and Shandong Academy of Medical Sciences (No.2022104391602), the Youth Science Foundation Cultivation Support Program of Shandong First Medical University and Shandong Academy of Medical Sciences (No. 202201-036) and the Tai'an Science and Technology Innovation Development Project (No. 2022GX019).

References

[1] Change IPOC. Climate Change 2014—Impacts, Adaptation and Vulnerability: Regional Aspects. Cambridge University Press, Cam-bridge, 2014.
[2] Ring MJ, Lindner D, Cross EF, e t al. Causes of the Global Warming Observed since the 19th Century. Atmospheric and Climate Sciences, 2012, 02(4). DOI:10.4236/acs.2012.24035.
[3] Stover SK, Mcarthur LB, Mabry ML. Presenting Global Warming and Evolution as Public Health Issues to Encourage Acceptance of Scientific Evidence. Bioscene, 2013, 38(2):3-10.
[4] Change IPOC. Climate Change: The IPCC Scientific Assessment. Press Syndicate of the University of Cambridge, New York, 1990.
[5] Berlie AB. Global Warming: A Review of the Debates on the Causes, Consequences and Politics of Global Response. Journal of Geography, 2018(2018):144-164.
[6] Uche-Anya E, Anyane-Yeboa A, Berzin TM, et al. Artificial intelligence in gastroenterology and hepatology: how to advance clinical practice while ensuring health equity. Gut, 2022, 71(9):1909-1915.
[7] Mi Xiaoxi, Tang Aitao, Zhu Yuchen, et al. Machine learning: A potential powerful tool for materials science research. Materials Reports, 2021, (15): 1-18.
[8] Ma Tinghuai, Mu Qiang, Tian Wei, et al. Research on meteorological data mining. Journal of Wuhan University of Technology, 2010, 32(16): 110-114.
[9] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8):1735-1780.
[10] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. ACM, 2016. DOI:10.1145/2939672.2939785.
[11] DUC LE V, CHA SK. Real-time Air Pollution prediction model based on Spatiotemporal Big data. The ternational Conferece on Big data, IoT and Cloud Computing, Guangzhou, 2018.
[12] Harilal N, Singh M, Bhatia U. Augmented Convolutional LSTMs for Generation of High-Resolution Climate Change Projections. IEEE Access, 2021(9):25208-25218.
[13] Malakouti, Seyed. Utilizing time series data from 1961 to 2019 recorded around the world and machine learning to create a Global Temperature Change Prediction Model. Case Studies in Chemical and Environmental Engineering. 2023,7:100312.
[14] Li Y, Li C, Li M, et al. Influence of Variable Selection and Forest Type on Forest Aboveground Biomass Estimation Using Machine Learning Algorithms. Forests, 2019, 10(12). DOI:10.3390/f10121073.