Forecasting Carbon Emissions in China's Provinces Based on Graph Neural Networks

Xiao FANG, Hanyu GONG, Mingze GONG

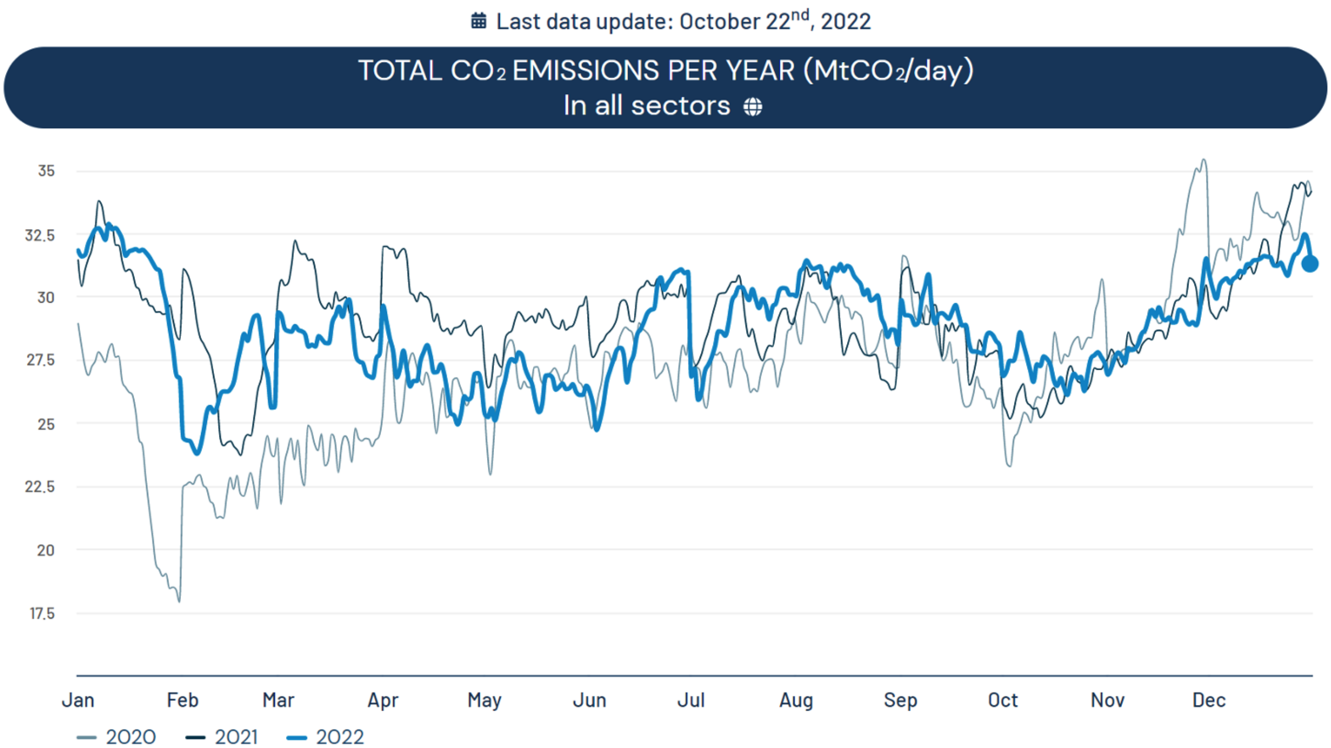

Figure 1: Total Carbon Emissions across Investigated Sectors

ABSTRACT

This is a coursework project submitted to the course *Foundation of Data Science and Analytics*. The project aims to forecast the carbon emissions in China's provinces based on Graph Neural Networks. The project is divided into three parts: data preprocessing, model training, and model evaluation. The data preprocessing part includes data cleaning, data integration, and data visualization. The model training part includes the construction of the graph neural network and the training of the model. The model evaluation part includes the evaluation of the model and the analysis of the results. The project is implemented in Python and the source code is available at here.

KEYWORDS

Emission prediction, Graph Neural Networks, Carbon Emissions, China's Provinces

1 INTRODUCTION

Global warming, far from being a recent phenomenon, can be perceived from another point as a persisting dilemma that continues to impact both anthropogenic development and the natural ecosystem. The primary agent provoking global warming is attributed to the emissions of greenhouse gases. Empirical studies confirm that carbon dioxide holds the dubious distinction of being the most abundant greenhouse gas in the atmosphere, contributing a staggering 72% to global warming [1].

However, it is noteworthy to mention that China ranks as the preeminent emitter of carbon dioxide on a global scale, discharging in excess of 6 billion tonnes of carbon dioxide annually. Hence, addressing this crisis, intrinsically connected to the existential fate of mankind, became a priority for China. As a concrete commitment to this endeavor, President Xi Jinping, in September 2020, proclaimed China's aim to "reach a peak in CO2 emissions by 2030 and accomplish carbon neutrality by 2060". The challenge of achieving carbon neutrality is multifarious, necessitating a holistic approach, encompassing policy, economy, culture, and technology. This paper opts to focus on the forecasting of carbon emissions, a crucial foundational element for strategic decision-making. Proficient predictions furnish invaluable data that bolsters informed decision-making. Conversely, if the prognosis proves inaccurate, ensuing plans may fall into the domain of impracticality [2].

The stakeholders who stand most directly impacted by these emissions include governments, investors, and researchers. Government bodies, equipped with foresight into future carbon emissions, can effectuate more meaningful change in climate policy, emergency development, and global cooperation. Conversely, researchers and investors, informed by predictive results, can more effectively design mechanisms such as the Emissions Trading System, Carbon Pricing System, and related technologies. Consequently, the act of forecasting carbon emissions carries significant implications for subsequent research.

Since the year 2011, endeavors have been made to utilize logistic equations in order to prognosticate China's carbon emissions [3]. At present, the primary methodologies employed for the prediction of CO2 emissions can be categorized into three principal clusters, specifically, statistical analysis models, non-linear intelligent models, and grey prediction techniques [4]. Statistical models offer ease of application, yet they necessitate the collection of ample historical data before the models can undergo training. Conversely, machine learning frequently outperforms in forecasting relative to conventional statistical methods.

This study implements the Graph Neural Network (GNN) approach, presenting multiple advantages: Firstly, GNNs are capable of effectively modelling spatial and temporal correlations. Secondly, GNNs have the capacity to integrate additional contextual data, such as socio-economic and policy factors. Thirdly, GNNs can yield an interpretable model structure, thus enabling researchers to derive insights into the relationships between various factors and their subsequent impact on carbon emissions[5].

At this juncture, GNNs have not been extensively explored in the context of predicting carbon emissions. Hence, future research should pivot its focus towards delving into the application of GNNs in carbon emissions forecasting. The unique benefits offered by GNNs should be leveraged to construct models that are both highly accurate and interpretable.

2 DATA DESCRIPTION

Transitioning now to our research endeavor, the initial phase entailed data acquisition from CarbonMonitor CHINA, a repository that chronicles the carbon emission statistics for the past five years. We amassed more than 200 thousand data points spanning from January 1, 2019, to December 31, 2022, systematically segregating the data into 31 distinct states and five sectors respectively.

Subsequently, we performed a rudimentary data analysis, classified according to temporality, sector, and state. A review of the past four years reveals that carbon emissions peaked in the year 2021. Furthermore, a substantial variation in carbon emissions was discerned across different sectors. Additionally, the states of Hebei and Shandong emerged as the leading contributors to the nation's carbon emissions.

3 METHODOLOGY

In this section of the term paper, we systematically explore the technical aspects of three distinct machine learning methodologies that are used in our study: the Adaptive Graph Convolutional Recurrent Neural Network (AGCRN), Multi-Layer Perceptron (MLP), and Fully Connected Long Short-Term Memory (FC-LSTM).

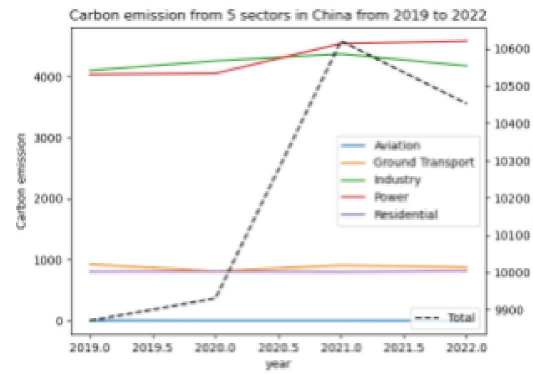

Figure 2: A graphical representation of emissions from five different sectors.

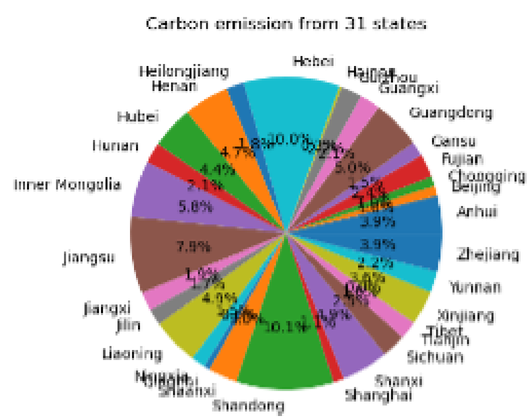

Figure 3: A graphical representation of emissions from 31 different states.

4 EXPERIMENT

In this section, we present the experiment of these three models, AGCRN, MLP and FC-LSTM.

4.1 Evaluation Metrics

To compare the model performance, we implement MAE, MAPE and RMSE to measure these three models.

MAE: MAE measures the average absolute difference between the predicted and actual carbon emissions. It evaluates the model's ability to make accurate predictions.

$$MAE(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

MAPE: MAPE measures the average absolute percentage difference between the predicted and actual carbon emissions. It is used to evaluate the accuracy of the model's predictions relative to the

Table 1: Parameter settings

Parameter/Setting	Value
Tp	1d
Tr	10d
Loss function	L2Loss
Optimizer	Adam
Percentage of training data	70%
Percentage of validation data	15%
Percentage of test data	15%
Epochs	100
Number of runs	1

Table 2: Average performance comparison of different approaches.

Method	MAE	MAPE	RMSE
AGCRN	0.0151	2.7%	0.0263
MLP	0.0699	8.8%	0.1010
FC-LSTM	0.5447	412.7%	0.4984

actual values.

$$MAPE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} \frac{||y_i - \hat{y}_i||}{||y_i||}$$

RMSE: RMSE is the square root of the MSE and is used to measure the standard deviation of the errors made by the model.

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

4.2 Simulation Results

In the simulation section, the parameter settings are listed in Table. 1, we split the datasets into 70% training data, 15% validation data and 15% test data.

In Table. 2, we exhibit the average performance comparison of different approaches. As you can see, the AGCRN performs the best in the three methods. MAE is 0.0151, MAPE is 2.7% and RMSE is 0.0263.

The AGCRN model outperformed the MLP and FC-LSTM models in our experiment, primarily due to its ability to capture the complex spatiotemporal dependencies of carbon emissions. The graph-based convolutional neural network structure of the AGCRN model enables it to extract features from the spatial and temporal domains simultaneously, resulting in more accurate predictions. In contrast, the MLP and FC-LSTM models are traditional machine learning models that rely on linear regression and LSTM networks, respectively, to forecast carbon emissions.

4.3 Performance Evaluation

In the experiment, we also visualize the prediction result and evaluate the performance, as shown in Fig. 4 and Fig. 5.

From these two pictures of Guangdong carbon emission prediction, we can apparently find that the AGCRN has better prediction

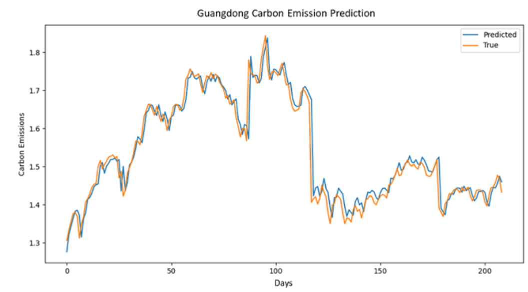

Figure 4: AGCRN.

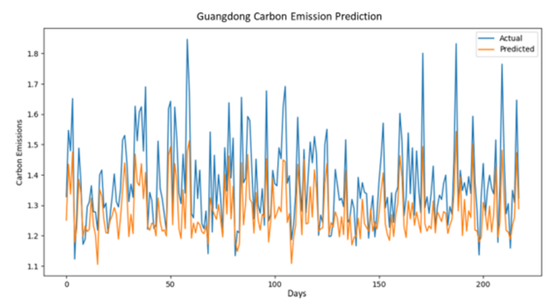

Figure 5: MLP.

results than MLP, which further proves that the AGCRN method is more suitable for carbon emission prediction.

In conclusion, the AGCRN model's superior performance in predicting carbon emissions can be attributed to its unique graph-based convolutional neural network structure, which allows it to capture the intricate spatiotemporal dependencies of carbon emissions. This study's findings offer valuable insights into the development of more accurate and effective models for predicting carbon emissions, which can inform policymakers and stakeholders in their efforts to reduce carbon emissions and mitigate climate change.

5 CONCLUSION AND FUTURE WORK

In this course project, we presented three models, AGCRN, MLP, and FC-LSTM, for predicting the daily carbon emissions of 31 Chinese provinces using historical data from January 1st, 2019 to December 31st, 2022. Our results demonstrated that the AGCRN model outperformed the MLP and FC-LSTM models, indicating its robustness in predicting carbon emissions accurately.

The findings of this project offer a new approach to enhance the precision of carbon emission prediction. The AGCRN model's superior performance can be attributed to its ability to capture the complex spatial and temporal dependencies of carbon emissions. Our study provides valuable insights into the development of more accurate and effective models for predicting carbon emissions, which can inform policymakers and stakeholders in their efforts to reduce carbon emissions and mitigate climate change.

Future work: This project model can help carbon market stakeholders grasp the future trend of the carbon market more accurately and provide a reference for policymakers and investors in decision-making. However, the quality and availability of the carbon emission data are low, which makes it hard to improve the accuracy rapidly.

In future work, we can consider more factors to improve the accuracy of carbon emission prediction such as weather, seasonality, and economic conditions. What's more, it is necessary to improve the computational complexity.

REFERENCES

[1] Ping Li, Congjun Rao, Mark Goh, and Zuqiao Yang. Pricing strategies and profit coordination under a double echelon green supply chain. *Journal of Cleaner Production*, 278:123694, 2021. ISSN 0959-6526. doi: 10.1016/j.jclepro.2020.123694.

[2] Forecasting Chinese CO 2 emissions from fuel combustion using a novel grey multivariable model | Request PDF. URL https://www.researchgate.net/publication/317647208_Forecasting_Chinese_CO_2_emissions_from_fuel_combustion_using_a_novel_grey_multivariable_model.

[3] Ming Meng and Dongxiao Niu. Modeling CO2 emissions from fossil fuel combustion using the logistic equation. *Energy*, 36(5):3355–3359, 2011. ISSN 0360-5442. doi: 10.1016/j.energy.2011.03.032.

[4] Mingyun Gao, Honglin Yang, Qinzi Xiao, and Mark Goh. A novel fractional grey Riccati model for carbon emission prediction. *Journal of Cleaner Production*, 282:124471, 2021. ISSN 0959-6526. doi: 10.1016/j.jclepro.2020.124471.

[5] Teg Alam and Ali AlArjani. Forecasting CO $_2$ Emissions in Saudi Arabia Using Artificial Neural Network, Holt-Winters Exponential Smoothing, and Autoregressive Integrated Moving Average Models. In *2021 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP)*, pages 125–129, Jakarta, Indonesia, 2021. IEEE. ISBN 978-1-66541-641-2. doi: 10.1109/ICT-PEP53949.2021.9601031.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009