CarbonNet: Enterprise-Level Carbon Emission Prediction with Large-Scale Datasets

Jinghua Tang[1], Nan Fang[1], Lanqing Yang[2(✉)], Yuqiao Pei[2], Ran Wang[2], Dian Ding[2], Yu Lu[2], and Guangtao Xue[2]

[1] Shanghai Voicecomm Technology Co., Ltd., Shanghai, China
[2] Shanghai Jiao Tong University, Shanghai, China
yanglanqing@sjtu.edu.cn

Abstract. The precise prediction of carbon emissions is crucial to combat global climate change and foster sustainable development. Conventional carbon emissions forecasting usually relies on limited records of national or regional levels, which is coarse and compromises its accuracy. Furthermore, the data collection requires lengthy statistical cycles and high costs, making it too costly to provide timely feedback from the forecasting. Additionally, they are targeted at specific fields thus hard to construct universal models. To overcome these challenges, we propose *CarbonNet*, a novel firm-level carbon emission prediction scheme. To build large-scale firm-level datasets, we crawled carbon emission data and reporting data (e.g., financial statements) of 3346 companies over 31 years containing 688 data fields, and combined them together. A preprocessing scheme is proposed to aggregate data with different statistical intervals or sources, and many outliers. A factor-analysis-based features extraction scheme is proposed to build a generalized forecasting model for different types of companies. A machine learning scheme is proposed for big data mining and long-term forecasting. We evaluated *CarbonNet* on real-world datasets. Results show that it achieves a median relative error of 0.25, outperforming others by 22%. The corresponding carbon emissions dataset has been made publicly available to advance related research.

Keywords: Enterprise-level carbon emissions prediction · Factor analysis · Big data mining

1 Introduction

In recent years, climate changes have attracting increasing attention. According to the United Nations, excessive carbon emissions are leading to global warming, with the earth's temperature rising by 1.5 °C in the next decade, causing 20% of species facing extinction [1]. Reducing carbon emissions has become an urgent issue all over the world. In order to reduce carbon emissions, accurate carbon emission forecasts are required, which can not only help provide timely warnings to avoid exceeding carbon emission standards, but also help carbon emitters to plan ahead for carbon emission trading and accelerate energy transformation. Accurate carbon emission forecasts have become an important way to reduce carbon emissions.

D.-S. Huang et al. (Eds.): ICIC 2024, LNCS 14873, pp. 411–422, 2024.
https://doi.org/10.1007/978-981-97-5615-5_33

There have been a lot of existing works attempting to make carbon emission forecasting. They use statistical learning methods [2–5] or machine learning methods [6–13] to predict the following carbon emissions of a region, thus enabling timely adjustments for regional-level carbon emission decisions. However, there are limitations with existing carbon emission methods: 1) Data granularity is too coarse. Traditional carbon emission methods typically use macro-level data sources, such as data from national, regional, and park-level statistical offices. They are difficult to be used for adjusting the carbon emissions of micro carbon emitters (e.g., specific companies). 2) Slow feedback on carbon emission adjustment. As it takes a lot of time to collect, summarize, generate reports, and make decisions using these macro-level data, they are insensitive to real-time carbon emissions fluctuations. This delay weakens its ability to guide real-time carbon emission policy adjustments. 3) Few carbon emission factors considered. The emission behavior of actual carbon emitting entities may be affected by various factors. Using only few carbon emission indicators often leads to poorer forecasting results. For the same reason, they need to make assumptions about data distribution, etc., and it is difficult to construct a generalized model. 4) Small data scale. For data-driven models, the carbon emission prediction accuracy using macro data sources is limited by the scale of data.

In response to these challenges, in this paper, we propose *CarbonNet*, a novel method for enterprise-level carbon emission prediction. To solve the problem of scarcity of enterprise-level carbon emission data, we crawled the carbon emission-related data of 3364 companies and calculated the corresponding carbon emissions; to include more factors that may affect a company's carbon emissions into consideration, we crawled the statement data, basic information, etc. of these companies and integrated them together; to solve the problem of irrelevant factors affecting the prediction performance of carbon emissions, we used factor analysis for feature extraction and validated it using a visualization process. Finally, we employed the XGBoost model for large-scale data mining. The scientific contribution of this study thesis is threefold:

- We built a large-scale enterprise-level carbon-emission prediction dataset. We built a big dataset containing 3346 companies over 31 years, and 688 data fields. It's built by manually crawling the carbon emissions and company reporting data of them. We have made the dataset openly available to further advance research in this field.
- We proposed a factor-analysis based features carbon emissions extraction scheme to build a generalized carbon emission model for different types of companies.
- We evaluated *CarbonNet* on 3346 enterprises to demonstrate that it can effectively improve the performance of enterprise-level long-period carbon emission prediction, with the median absolute relative error (MARE) of 0.25, outperforming others by 22%. The corresponding carbon emissions dataset has been made publicly available to advance related research.

2 Related Work

2.1 Carbon Emission Prediction with Statistical Models

Traditionally, there are many existing works using statistical models for carbon emission predictions. For example, Grubert [2] presents a model for utility-specific annual average emissions projections in the United States' electricity sector through 2050. Grubb

[14] reviews 89 scenarios from 12 different models regarding CO2 emissions in China up to 2030. Van Ruijven [15] introduces a global simulation model for the steel and cement industries, providing projections of energy use and CO2 emissions. Gao et al. [3] introduce a novel fractional grey Riccati model that integrates the Environmental Kuznets Curve hypothesis and the differential information principle for predicting CO2 emissions. Kargupta [5] explores the challenge of controlling greenhouse gas emissions, with a focus on the transportation sector. These methods are generally based on regional or national carbon emissions statistics and assume that carbon emissions follow certain assumptions, which may fail when it comes to corporate carbon emissions. In comparison, *CarbonNet* was directly developed on a finer-grained enterprise-level dataset; Besides, the data-driven learning strategy also makes it rely less on specific assumptions.

2.2 Carbon Emission Prediction with Machine Learning Models

There are also many existing works that try using machine learning models for carbon emission predictions. For example, Meng [6] utilizes the logistic function and three industry-specific parameter estimation algorithms to model and forecast CO2 emissions from fossil fuel combustion. Du [7] utilizes a Logistic model to predict carbon emissions in Chinese provinces from 2011 to 2020. Agbulut [8] employs deep learning, support vector machine, and artificial neural network algorithms to forecast transportation-related energy demand and CO2 emissions in Turkey. Ahmed [10] investigates influencing factors of carbon emissions in China and India, utilizing a machine learning approach, particularly the long short-term memory (LSTM) method. Huang [11] employs grey relational analysis, principal component analysis, and the long short-term memory (LSTM) method to identify influential factors, providing a theoretical basis for carbon emission reduction strategies in China. Sun [12] explores the relationship between green finance and carbon emissions through a correlation analysis model based on big data and machine learning. Yang [13] emphasizes achieving net zero carbon emissions and outlines how data mining, leveraging transportation big data, can address gaps in medium and heavy vehicle electrification. However, these machine learning methods are based on data with coarse granularity (usually regional or national data) and few dimensions (carbon emission data and several industrial data fields), which are less generalizable when applied directly to enterprise-level data. In comparison, *CarbonNet* takes into account a number of indicators related to carbon emissions (e.g., enterprise type, debt size, etc.) and extracts important factors from these indicators, which enables it to make more accurate carbon emission forecasts.

3 Data Collection and Preprocessing

3.1 Collection of Carbon Emission Data

This study innovatively combines financial statements and carbon emission data of publicly listed companies to construct a comprehensive dataset for subsequent carbon emission forecasting models. The carbon emission data, spanning from 1990 to

2021, was crawled from China Data Online [16], reflecting disparities among different companies. Notably, the temporal data from 1990 to 2001 are recorded in six-month intervals, while data from 2002 to 2021 are documented quarterly. The dataset encompasses 10 data fields, including total carbon emissions, emissions from fossil fuel combustion, biogenic fuel combustion, fugitive emissions from raw material extraction, fugitive emissions from oil and natural gas systems, indirect carbon emissions from electricity transmission, emissions from production processes, emissions from solid waste incineration, emissions caused by wastewater treatment, and emissions resulting from land-use changes (e.g., deforestation for industrial use). The "total carbon emissions of listed companies" field, used in this paper as the enterprise carbon emission data, is derived by summing various types of carbon emission data, which can be formulated by: TotalCarbonEmissions = Combustion/EnergyFuelEmissions + ProductionProcessEmissions + WasteEmissions + LandUseChangeEmissions We have open-sourced the dataset.[1]

Table 1. Data Fields in Quarterly Corporate Reports.

Category	Representative Fields	Fields
Financial Summary	Earnings Per Share, Net Assets	212
Income Statement	Operating Incomes, Operating Costs	128
Balance Sheet	Total Current Assets, Total Liabilities	156
Cash Flow Statement	Cash Flow from Operations, Cash Flow from Investments	105
Shareholder Equity	Total Shareholder Equity, Dividend Per Share	79

3.2 Collection of Company Reporting Data

Concurrently, enterprise report data was gathered via web crawling from corporate disclosure websites of listed companies, with a temporal dimension extending from 1990 to 2022, and exhibiting inter-company variations [17]. The reporting data fields are categorized into four types: financial summary data, income statement data, balance sheet, cash flow statement, and shareholder equity data. As shown in Table 1, the financial summary includes metrics such as Earnings Per Share (EPS), Net Asset Value Per Share (BPS), and Return on Equity (ROE). The income statement data comprises total operating revenue, total operating costs, and operating profit. Cash flow data include cash flows from operating, investing, and financing activities. For each category, the data fields span from 79 to 212, with a total of 688 data fields.

[1] https://github.com/CarbonNet2023/CarbonNet/

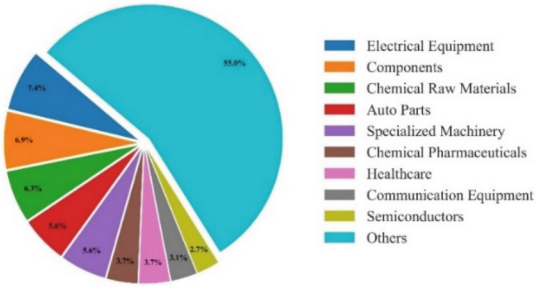

Fig. 1. Industry types of 3346 companies.

3.3 Collection of Company Basic Data

In addition to the primary financial and emission metrics, our research undertook a comprehensive data-gathering approach to encapsulate various auxiliary factors potentially impacting corporate carbon footprints. Utilizing the Scrapy web-crawling framework, we extracted additional corporate parameters of 8 data fields including the specific industry sectors, employee headcounts, geographical locations of the enterprises, etc. For example, Fig. 1 shows the industry types of all 3346 companies. This supplementary data provides a broader context for each enterprise, contributing to a more holistic understanding of the environmental impact reports.

3.4 Data Preprocessing

The preprocessing stage comprises several components, like data integration, anomaly detection and mitigation, temporal data segmentation through sliding windows, and the partitioning of data into training and validation sets.

Anomaly Detection: We combine box plots and the 3σ rule for outlier detection in each data column. To address the 3σ method's sensitivity, we enhance accuracy by combining the box plot method with weighted upper and lower bounds. The specific formula is provided below: Lower Inner Fence (LIQ): $\text{LIQ} = \text{Q1} - 1.5 \cdot \text{IQR}$, , and Upper Inner Fence (UIQ): $\text{UIQ} = \text{Q3} + 1.5 \cdot \text{IQR}$ Lower and Upper Boundaries with Weighted Ratios (ratio): $\text{Low} = 3 - \sigma \cdot \text{ratio} + \text{LIQ} \cdot (1 - \text{ratio})$, $\text{High} = 3 - \sigma \cdot \text{ratio} + \text{UIQ} \cdot (1 - \text{ratio})$, $\text{ratio} = 0.3$ For outliers and gaps in company-specific time-series data, we utilize moving averages for missing values and the median for outlier replacement, ensuring precise reflection of underlying characteristics.

Window Sliding: In selecting the sliding window size, experiments to compare various lengths for accuracy. The results show that the optimal length of the time sliding window is 4 using four quarters of the company's historical data in the current year as the data window for predicting the next quarter's data. At the same time, we also specially process the data at the boundary of the training and test sets to ensure that there are no data leakage problems.

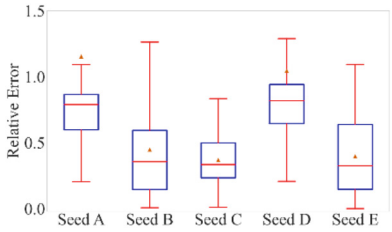

Fig. 2. Prediction errors by different division of companies randomly. For each, 80% companies used for training and others for test.

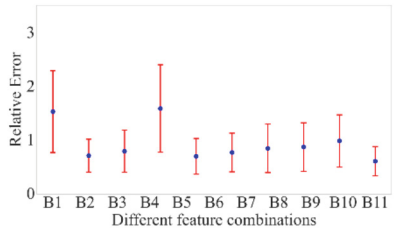

Fig. 3. Prediction errors v.s. different factor sets: using carbon emission only (B1), or 4 selected factors (B2-B11) correspondingly.

Data Segmentation: The integrated dataset includes carbon emission data and financial reports from 3,364 companies. In this study, 80% of the companies are randomly selected for training and 20% for testing.

4 Preliminary Study

In order to perform enterprise-level carbon emission forecasting, a dataset of enterprise carbon emissions needs to be constructed. In constructing this dataset, we explore the following key issues.

a) Is it feasible to construct a generic carbon emissions model using only the carbon emissions data of a company? To answer the first question, we conducted an experiment using the carbon emission data of the 3364 companies. In the experiments, we used SVM models to predict future carbon emissions with a training data length of 4 quarters and a prediction scale of 1 quarter. Note that we always randomly select 80% of the companies' for training and the others testing. Five different random seeds are used to make cuts to construct different datasets A1-A5, while Fig. 2 shows the prediction errors on A1-A5 correspondingly. It shows for same model but different cuts, there is a 78.4% difference in average error between the best performing Seed C dataset and the worst performing Seed A dataset. This suggests it difficult to construct a generalized accurate carbon emission model for all firms using carbon emissions alone. This is mainly due to the huge differences in corporate industries and company sizes among different companies, as shown in Fig. 1; this also inspires us that more comprehensive company data needs to be integrated in order to develop a generalized prediction model.

b) Does the inclusion of more company information always improve model performance? In order to construct a more generalized model, we crawled the above company's corporate statements and other 688 pieces of company-related information. In experiments, in addition to the carbon emission data, we tried to construct the dataset by randomly adding 3 of the 688 dimensional corporate information. By selecting 10 different random number seeds, the B1-B11 datasets were obtained, where B1 indicates that only carbon emission information was used. The SVM model was also taken to predict each dataset, and the prediction errors are shown in Fig. 3. It shows that although

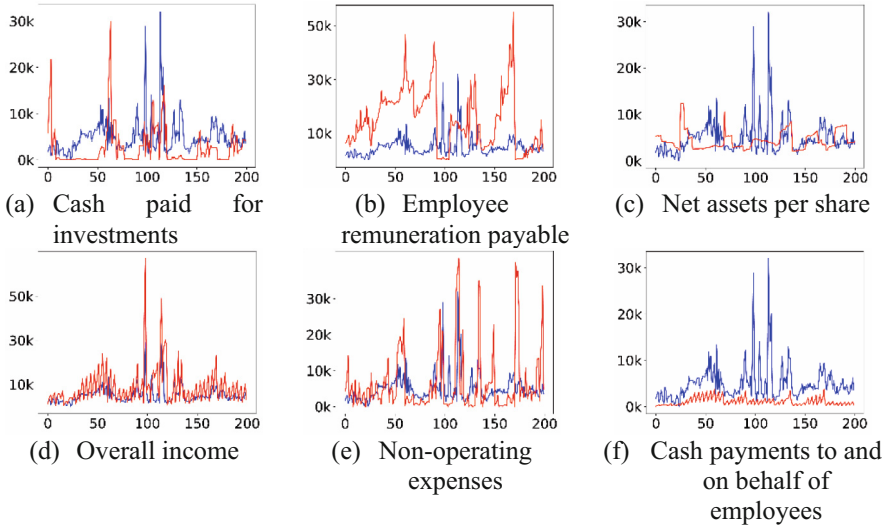

Fig. 4. Correlation of different variables with carbon emission. X axis denotes different times (each point for a season). Y axis denotes the carbon emissions of a company.

the error is smaller than B1 on most of the datasets, increasing the dimension of the factors on dataset B4 instead makes the error larger, even up to 12% higher.

c) Why does adding more enterprise data not improve model performance? To explore this issue, as shown in Fig. 4, we drew six of the factors and carbon emissions on the same graph as a comparison, the result suggested that some business information, such as total business revenue is in line with the trend of carbon emissions changes, while other factors, such as Cash paid for investments, are basically uncorrelated with the trend of carbon emissions. This can be also attributed to the prevalence of outliers or uncorrelated factors in certain industry data, which, once included, can affect the overall prediction accuracy.

In summary, a universal carbon emission model cannot be constructed by using only carbon emission information or using all data of a company, which may also reduce the accuracy of the model. Therefore, we propose to incorporate selected and reconstructed industry information to construct a generic and highly accurate carbon emission prediction model. In the following sections, we will discuss the related system architecture design and algorithm in detail.

5 System Design

As shown in Fig. 5, the system is divided into 3 steps: data collection and preprocessing; factor analysis based feature extraction and XGBoost based data mining.

5.1 Factor Analysis with Stepwise Regression

There are 686 data fields in total. To link this data to carbon emissions, regression analysis with the Pearson correlation coefficient is employed for dimensionality reduction.

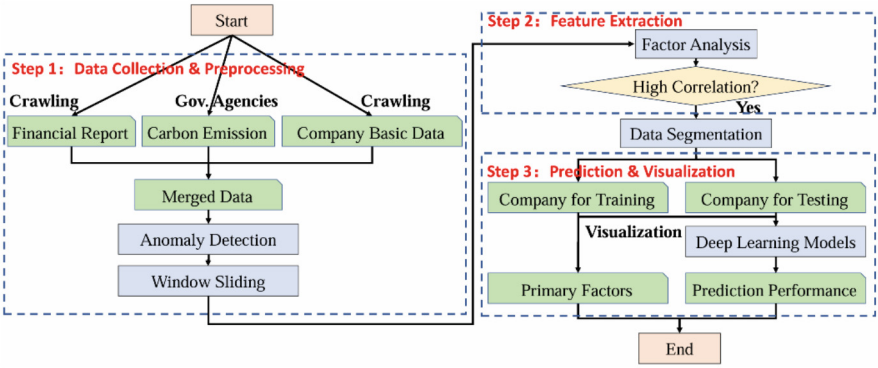

Fig. 5. *CarbonNet* framework.

For analyzing financial reports with high dimensionality and covariance, the variance inflation factor (VIF) is used with a threshold of 10. The stepwise regression method is applied for efficient correlation analysis and to eliminate highly covariant features [18]. The combined algorithm of Stepwise Regression, integrating both Forward Selection and Backward Elimination, is outlined in Algorithm 1:

Algorithm 1 Combined Stepwise Regression

Initialize the model with a set of candidate variables $X = \{X1, X2, \ldots Xn\}$.
Define the dependent variable Y.
Initialize empty model for Forward Selection or full model for Backward Elimination.
While model improvement $Li > Th$ **do**
  if Forward Selection then
    Test addition of each variable $X_i \notin$ model.
    Add the variable X_i with biggest p-value.
  end if
  if Backward Elimination **then**
    Test removal of each variable $X_i \in$ model.
    Remove the variable whose exclusion most improves the model.
  end if
  Update the model and evaluate its performance.
end while
Finalize the model when no further significant improvement is found.

In this algorithm, Y represents the dependent variable, and X_i are the independent variables. The model iteratively adds or removes variables from the set X, adjusting their coefficients β_i in the regression equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

where β_0 is the intercept, β_i are the coefficients, and ε is the error term. The objective is to find the variables that best predict Y.

5.2 Extreme Gradient Boosting

We employ XGBoost (Extreme Gradient Boosting) for mining the large-scale dataset, which is known for its efficiency, flexibility, and portability [19]. The algorithm is formalized as follows:

$$\text{Obj}(\Theta) = L(\Theta) + \Omega(\Theta)$$

where $\text{Obj}(\Theta)$ is the objective function to be minimized, $L(\Theta)$ is the training loss function, and $\Omega(\Theta)$ is the regularization term. Θ represents the parameters of the model. In this algorithm, $l(y_i, \hat{y}_i)$ represents the loss function computed between the actual value y_i and the predicted value $\hat{y}_i$. The gradient and hessian are used to guide the construction of trees, effectively optimizing the loss function. The algorithmic steps are encapsulated in Algorithm 2:

Algorithm 2 XGBoost

Given a dataset with n examples and m features.
Initialize the model with a constant value: $f_0(x) = \text{argmin}_\gamma \sum_{t-1}^{n} l(y_i, \gamma)$.
for $t = 1$ to T **do**:
  Compute the gradient and hessian of the loss function: $g_{ti} = \partial_{\hat{y}_{(t-1)i}} l\left(y_i, \hat{y}_{(t-1)i}\right), h_{ti} = \partial^2_{\hat{y}_{(t-1)i}} l\left(y_i, \hat{y}_{(t-1)i}\right)$.
  Fit a tree to the gradients and hessians and add it to the model: $f_t(x) = f_{t-1}(x) + \text{learn}(\{g_{ti}, h_{ti}\}_{i=1}^{n})$.
  Update the model: $\hat{y}_{ti} = \hat{y}_{(t-1)i} + f_t(x_i)$.
end for
The final model is $\hat{y}_T$.

In this paper, for the XGBoost model, the max depth is set to 6, the estimators number set to 50, and the learning rate is 0.01.

6 Evaluation

The *Median Absolute Relative Error* (MARE) measures the accuracy of predictions in a regression model. This measure is robust against outliers, providing a typical error magnitude. It is defined as the median of the absolute values of relative errors between the actual values y_i and the predicted values $\hat{y}_i$:

$$\text{MARE} = \text{median}\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right)$$

6.1 Overall Performance

Figure 6 and Fig. 7 demonstrate the overall performance of the system. Figure 6 presents the effectiveness of different algorithms in predicting corporate carbon emissions. The

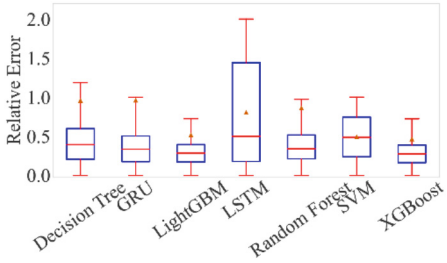

Fig. 6. Performance of 7 algorithms.

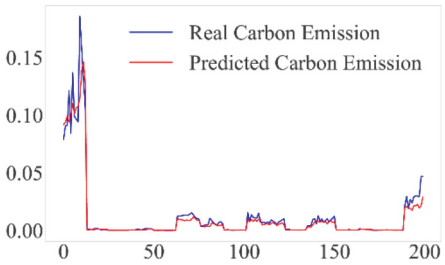

Fig. 7. Regression using XGBoost.

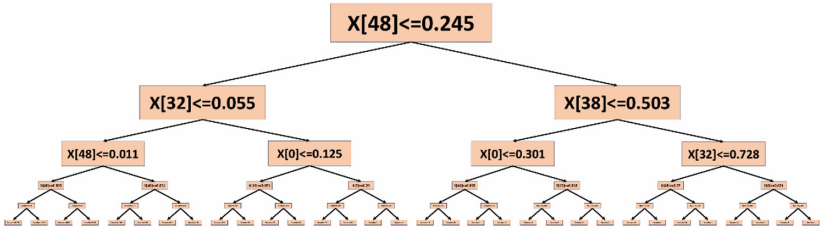

Fig. 8. Visualization using Decision Tree. x[48] denotes carbon emission of last season, x[32] denotes carbon emission of 2 seasons ago, and x[38] denotes the total liabilities of the last season.

results indicate that the XGBoost model outperforms others, achieving MARE of only 0.25, which is a 22% improvement over the next best method. This superior performance is attributed to XGBoost's ability to better integrate various factors and provide longer prediction timelines. Figure 7 displays the regression results using the XGBoost with a median regression error of only 34093 tons. Notably, this predictive accuracy uses data from 80% of companies for training. This underscores *CarbonNet*'s capability to effectively meet the carbon emission prediction needs of different companies.

6.2 Visualization with Decision Tree

As shown in Table 2, the analysis reveals 6 important factors, like total operating income and liabilities, highly correlated with carbon emissions. Also, as seen in Fig. 8, not only previous carbon emissions, but also the financial data of some companies (x[38]), also play a major role in the decision-making process (the second layer).

6.3 Efficiency of Proposed Schemes

Efficiency of Adding In Company Data: As shown in Fig. 9, experiments verify that adding factors with higher correlations improves model performance. The model used factors selected predicts better results than carbon emission univariate. The relative error of carbon emission univariate prediction is 1.52 and the relative error of our prediction is 0.51. Compared to the carbon emission univariate model, our accuracy is improved by 66.4%.

Table 2. Correlation of Various Factors with Carbon Emissions.

Factors	Correlation	Factors	Correlation
Overall income	0.878	Cash payments to and on behalf of employees	0.782
Total liabilities	0.700	Net profit	0.699
Non-operating expenses	0.595	Non-operating profit	0.351

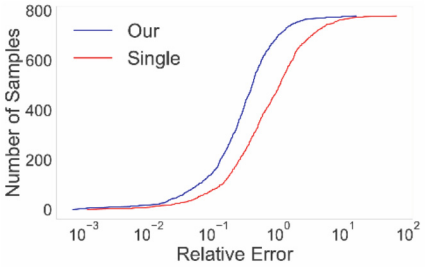

Fig. 9. Prediction errors by 2 schemes. Single for using carbon emission only.

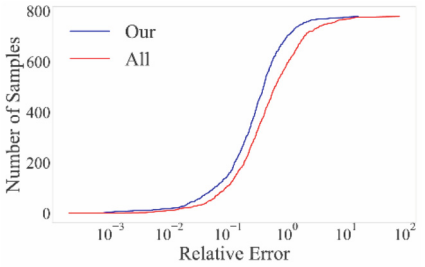

Fig. 10. Prediction errors by 2 schemes. All for using all data.

Efficiency of Factor Analysis Schemes: As shown in Fig. 10, the experiment verifies that screening the factors, removing the less relevant ones improves the model performance. The relative error of the holovariate prediction result is 1.12, while the relative error of our prediction is 0.51. Compared to the model using all variables, our accuracy improved by 54.4%.

7 Conclusion

In this paper, we propose *CarbonNet*, an enterprise-level carbon emissions forecasting scheme using enterprise carbon emissions and report data. We built a big dataset containing 3346 companies over 31 years. We extracted the reported essential factors that are related to carbon emissions. Experimental shows that *CarbonNet* achieves a median absolute relative carbon emission error of 25%, median regression error of 34093 tons, which demonstrates that *CarbonNet* is very promising in helping companies plan ahead for carbon emissions and adjust their carbon behaviors.

References

1. AR6 Climate Change 2021: The Physical Science Basis —IPCC, November 2023. Accessed 15 Nov 2023
2. Grubert, E.: Emissions projections for us utilities through 2050. Environ. Res. Lett. **16**(8), 084049 (2021)

3. Gao, M., et al.: A novel fractional grey riccati model for carbon emission prediction. J. Clean. Prod. **282,** 124471 (2021)
4. Dong, F., et al.: Regional carbon emission performance in china according to astochastic frontier model. Renewable Sustainable Energy Rev. **28**, 525–530 (2013)
5. Kargupta, H., et al.: The next generation of transportation systems, greenhouse emissions, and data mining. In: Proceedings of the 16th ACM SlGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1209–1212 (2010)
6. Meng, M., Niu, D.: Modeling co2 emissions from fossil fuel combustion using the logistic equation. Energy **36**(5), 3355–3359 (2011)
7. Du, Q., et al.: Forecast carbon emissions of provinces in china based on logistic model. Resour. Environ. Yangtze Basin **22**(2), 143–151 (2013)
8. Agbulut, U.: Forecasting of transportation-related energy demand and co2 emissions in turkey with different machine learning algorithms. Sustainable Prod. Consumption **29**, 141–157 (2022)
9. Mo, X., Li, M., Li, M.: Predicting abnormal events in urban rail transit systems with multivariate point process. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 41–53. Springer (2022)
10. Ahmed, M., Shuai, C., Ahmed, M.: Influencing factors of carbon emissions and their trends in China and india: a machine learning method. Environ. Sci. Pollut. Res. **29**(32), 48424–48437 (2022)
11. Huang, Y., Shen, L., Liu, H.: Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in china. J. Clean. Prod. **209**, 415–423 (2019)
12. Sun, C.: The correlation between green finance and carbon emissions based on improved neural network. Neural Comput. Appl. **34**(15), 1239912413 (2022)
13. Yang, M., et al.: Data mining challenges and opportunities to achieve net zero carbon emissions: focus on electrified vehicles. In: Proceedings of 2023 SIAM International Conference on Data Mining (SDM), pp. 953–956. SIAM (2023)
14. Grubb, M., et al.: A review of Chinese co2 emission projections to 2030: the role of economic structure and policy. Climate Policy **15**(sup1), S7–S39 (2015)
15. Ruijven, V., et al.: Long-term model-based projections of energy use and co2 emissions from the global steel and cement industries. Resour. Conserv. Recycl. **112**, 15–36 (2016)
16. Wang, H., Liu, J., Zhang, L.: Carbon emissions and assets pricing-evidence from Chinese listed firms. China J. Econ **9**, 28–75 (2022)
17. Seasonal Company Reporting Data, November 2023. Accessed 15 Nov 2023
18. Gandhmal, D.P., Kumar, K.: Systematic analysis and review of stock market prediction techniques. Comput. Sci. Rev. **34,** 100190 (2019)
19. Chen, T., et al.: XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, August 2016