Proceedings of the 3rd International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/140/2025.21274

Application of Machine Learning and Deep Learning in Weather-related Forecasting

Zhixi Ye[1,a,*]

[1]*Computer Science, Hainan Bielefeld University of Applied Sciences, Danzhou, Hainan, China*
a. zhixi.ye.23@stu.hainan-biuh.edu.cn
**corresponding author*

Abstract: Weather plays a broad and decisive role in many areas. Its volatility can disrupt traffic and endanger lives. It is therefore imperative to accurately predict its impact. Improved forecasting accuracy can aid multi-industry decision making. Traffic authorities can control traffic in advance; Agriculture can adjust its strategy in time; Resource allocation can be optimized in the energy sector. The rise of machine learning and deep learning technologies has opened up new prospects for weather-related forecasting. This article takes a methodical look at the application of machine learning and deep learning to traffic and weather forecasting, dissecting the details of model construction, data processing processes, and performance evaluation metrics. By comparing the advantages and disadvantages of each model, it provides ideas for model improvement, powerfully guides future research direction, and lays the foundation for building a more accurate prediction framework. The research direction of this thesis has far-reaching theoretical and practical value.

Keywords: Machine learning, deep learning, weather prediction, model application.

1. Introduction

Traffic conditions are highly affected by weather conditions. It exerts a profound and extensive influence on many aspects of human society. For example, heavy rain may cause water on the road, leading to traffic congestion and even traffic accidents. Foggy weather will seriously affect visibility, causing traffic jams and so on. Accurate prediction of weather conditions affecting traffic is of great significance for traffic management departments to clear traffic, issue early warnings, and make reasonable travel planning for drivers. In this context, with the development of artificial intelligence technology, models of different network structures have been applied to the field of weather forecasting, hoping to grasp the impact of weather on traffic more accurately. With the power of data mining and pattern recognition, machine learning, relying on its advanced data mining and excellent pattern recognition algorithm, can acutely capture key information clues from massive, heterogeneous and dynamic meteorological data, deeply deconstruct the internal logical correlation between weather elements and complex traffic phenomena, and provide strong support for forecasting models. With its rich hierarchical and well-structured neural network architecture, deep learning can automatically extract higher-order abstract features contained in data, effectively deal with nonlinear problems under the influence of multi-factor coupling in complex weather situations, and build a solid foundation for accurate prediction. The two complement each other and jointly promote the historic leap of weather forecasting from a traditional empirical model to an intelligent data-driven model. Therefore, different models will be discussed here to see which combination works best.

This review includes model principle, data processing, performance evaluation and application scenarios. The architecture principles of machine learning multiple linear regression (MLR), decision tree and deep learning convolutional neural network (CNN) are reviewed to reveal their unique mechanism of processing meteorological data. This review is of great significance, which can expand the boundary of meteorological forecasting knowledge and lay a foundation for model innovation. In practice, it provides convenience for accurate decision-making in the industry, such as accurate traffic dredging, agricultural scientific planning, and reasonable allocation of energy, so as to enhance the overall vitality of the social economy and open a new chapter of intelligent meteorology.

2. Machine Learning Approaches

The linear regression model is the cornerstone of exploring the relationship between weather factors and traffic congestion [1]. In a study by Lee et al., the MLR model was developed [2]. This data contains 48 weather forecast factors such as temperature, humidity, cloud cover, rainfall and wind speed, as well as six dummy variables representing the working day, and traffic information (traffic speed of each section, section ID, time and speed, etc.). Data collected from Korea Meteorological Administration. The data is preprocessed and the weather data is converted from multi-time scale forecast data to daily frequency data suitable for analysis. The model was constructed with traffic congestion score as the dependent variable and meteorological factors and weak dummy variables as the independent variables. A full regression model with 54 variables was first established, and then the variable elimination method was used to screen the variables with significant influence to determine the optimal model. The model was optimized to improve the explanatory power and prediction accuracy during training. The verification set was used to evaluate and monitor the performance to prevent overfitting. With MAPE as the index, MAPE was 0.059 (94.1% accuracy rate) in 2013 and 0.152 (84.8% accuracy rate) in 2014. The reliability of traffic congestion prediction from a weather perspective is verified.

Decision trees can provide managers with a clear and intuitive logical framework. Decision trees play an important role in weather-related forecasting research. For example, in the study of Liu et al. on regional landslide hazard warning models, decision tree algorithms were used in the comparison of multiple machine learning algorithms [3]. In the sample preparation stage of model construction, the team analyzed geological and rainfall elements in depth: according to the geological environment (slope range, slope direction change, formation lithology, etc.), the feature database was constructed, and the data of daily rainfall and previous effective rainfall were accurately collected to form the feature set; Based on the spatial and temporal constraints of positive sample innovation, negative samples are randomly sampled, the buffer radius is set to sample the outer space of buffer, and the positive and negative samples (about 1:2) are randomly selected at a specific time period (such as the multi-year flood season). Superposition positive and negative samples and feature database to extract attributes in-depth, strictly clear data to process missing values, outliers and unified feature dimensions, and build a high-quality training sample set. The decision tree constructs a hierarchical tree-like decision structure based on rich and diverse sample attributes. In the training session, the sample set is scientifically divided according to 80% training and 20% testing, and then the rigorous learning training and optimization journey is started. The model performance was comprehensively evaluated by the 50-fold cross-validation method, which combined accuracy (up to 0.937, measuring the proportion of samples predicted correctly), ROC curve (intuitively presenting the dynamic balance of sensitivity and specificity of the model) and AUC value (0.933, quantifying the overall advantages and defects of the model, the closer the AUC to 1, the better). Although it is slightly inferior to the random forest algorithm (accuracy of 0.963 and AUC 0.986), the Bayesian optimization

DOI: 10.54254/2755-2721/140/2025.21274

algorithm is used to fine-adjust parameters and effectively improve the prediction accuracy and generalization ability of the decision tree model, so that it can maintain stable performance under complex and changeable geological and meteorological conditions.

3. Deep Learning Approaches

CNN are mainly used in images and videos, and can effectively extract spatial features. For example, Sunardi et al. adopted a dataset containing weather conditions (such as temperature, humidity, wind speed, etc.) and historical travel time data in the model construction of weather prediction [4]. Data are collected from meteorological agencies and aviation management systems to ensure the accuracy and completeness of data. The collected data is then preprocessed to ensure data quality and consistency. This includes dealing with data issues such as missing values, outliers, etc., to make the data fit into the input of the CNN model. The convolutional layer extracts advanced data features by a special filter or kernel and detects the mode features of input data. The pooling layer reduces the dimensionality and efficiency, and reduces the risk of overfitting; The optimal route change prediction results are outputted through the integrated processing information of the full connection layer. 1000 data sets were used in the experiment, including wind speed, direction, temperature, rainfall, visibility and other parameters. With the increase of training rounds, the training loss decreases, indicating that the model continues to learn the data pattern, and the data fitting ability of route change becomes stronger. At the same time, the loss is monitored and verified to ensure that the model generalizes well to unknown data and prevents overfitting. The accuracy rate of predicting the optimal route change was 92%, which confirmed that CNN can accurately identify and adapt the route change mode under different weather, laying a solid foundation for traffic planning to accurately navigate according to weather.

Feedforward Neural Network (FNN) is a type of artificial neural network with a simple structure and easy construction. For example, Betkier et al., use weather conditions (such as rainfall intensity (7 categories), fog, storm, etc.), road data, population density data, etc., in the construction of the weather prediction model [5]. Model construction method: FNN with a multi-layer perceptron structure is adopted. The input layer neurons receive a variety of information, including road type, time, event, weather condition, population density, etc. The hidden layer uses a hyperbolic tangent function as the activation function to prevent excessive influence of extreme values, and the output layer uses a linear activation function to facilitate model extrapolation. Model training takes Mean squared error(MSE) as an error function and optimizes model parameters by minimizing error. For the display of results, MSE is 0.0472, indicating that the model has a certain accuracy in estimating traffic factor parameters, and the error is within an acceptable range, which can better capture the changing trend of traffic conditions.

Long Short-term Memory Network (LSTM)is a special type of recurrent neural network. It is mainly used to process and predict important events with very long intervals and delays in time series or sequence data [6]. A dataset from the Kaggle platform, encompassing daily weather details of the Delhi region such as date, average temperature, humidity, wind speed, and average pressure, was utilized for model construction. Subsequently, the data underwent preprocessing: a novel feature, "humidity pressure ratio", was generated. Thereafter, the date column was dissected into three distinct columns representing the year, month, and day. Ultimately, these data were normalized to ensure uniformity and suitability for subsequent analysis and model training. LSTM is used to construct the prediction model, and the LSTM model class is defined first. Its structure includes two layers of bidirectional LSTM, with 6 input features (month, day, humidity, wind speed, average air pressure, humidity to air pressure ratio), 128 hidden units, and the adjustment of "batch_first = True" to fit the data format. The LSTM layer is then connected to the Dropout layer with a probability of 0.2 to prevent overfitting, and two fully connected layers are connected: the former maps the bidirectional

DOI: 10.54254/2755-2721/140/2025.21274

LSTM output to 64 dimensions, the latter shrinks to 8 dimensions, and finally passes through the output linear layer to 1 to predict the daily temperature, completing the 200-round training process. Root mean square error (RMSE) is calculated to measure the performance and the RMSE value is 0.052093. The lower RMSE shows the accuracy of the model in predicting the average daily temperature, which can accurately capture the regularity and trend of meteorological data, and provide strong support for accurate meteorological forecasting.

4. Transformer and hybrid methods

In the field of advanced computer network model research, many innovative methods have emerged to improve the processing power and representation effect of models on complex data. For example, Jaderberg et al., proposed to segment the input image into "visual sentences" and smaller "visual words" [7]. The internal and external transformer blocks work together to process features and attention, maintain spatial information through unique location coding, reveal its advantages of improving performance with a small number of computing resources, and build network variants of different scales to flexibly adapt to tasks. Then, Han et al., proposed that the Point Transformer layer can operate in the local neighborhood based on vector self-attention, innovative position coding can enhance perception by using coordinate information, and the carefully constructed network architecture includes multi-stage downsampling and transition module to realize feature fusion [8]. Ablation studies accurately determine the optimal settings for key parameters such as the number of neighbors, location encoding, and type of attention. Finally, Zhao et al., proposed that the Spatial Transformer module can be composed of a positioning network, parametric sampling grid and micro-image sampling components [9]. The positioning network outputs transformation parameters, the sampling grid is deformed according to parameters, and the sampling mechanism ensures gradient return. This module can be embedded in any position of CNN. Multiple modules can be deployed in serial or parallel according to tasks, significantly improving model space transformation and feature extraction capabilities in image classification, co-localization, and spatial attention tasks, reducing the impact of data space changes, and achieving higher accuracy and efficiency.

5. Limitations and future prospects

As far as the current situation is concerned, in machine learning, MLR can only assume a linear relationship between dependent variables and independent variables, which may not be the case in practice and cannot capture nonlinear features in data. The decision Tree Model is prone to overfitting risk, and its convenience is limited and its feature interaction processing ability is weak. In deep learning, the CNN model has a complex structure, excessive training parameters, is easy to overfit, and requires a large amount of data. If the data is insufficient or poor in quality, the model performance will be affected, the calculation cost will be high, and the training and reasoning time will be long. FNN expression ability is limited, difficult to deal with complex nonlinear problems, prone to gradient disappearance or explosion problems, resulting in training difficulties, high requirements for feature engineering, need to manually extract and select effective features, the required cost and difficulty is too great. LSTM has a complex structure and many parameters, which are prone to overfitting phenomenon, long training time, high computational complexity, and a large number of labeled data are required for training, resulting in high cost. In the hybrid approach, the computing resource cost of the transformer is too high, and the long distance dependence on capture capability has great limitations.

In the future, Transformer and CNN can be combined to accurately locate key areas, lock target positions and focus important areas, enhance model spatial processing and feature extraction capabilities, and improve prediction accuracy and efficiency, or CNN and LSTM can be combined.

CNN is effective at capturing spatial features in weather data, while LSTM is good at handling long-term dependencies in time series data [10]. The combination of the two can not only extract spatial features from satellite cloud images, weather radar images and other data but also make full use of time information in historical weather data, so as to find out the spatiotemporal laws in weather data more comprehensively and improve the accuracy of prediction.

6. Conclusion

This review systematically explores the multivariate models and technical paths of machine learning and deep learning in weather-related forecasting. In the field of machine learning, MLR and decision tree models have poor prediction accuracy due to easy overfitting and poor feature interaction processing. In the field of deep learning, CNN, FNN and LSTM have certain advantages in prediction, but there is still room for improvement due to complex structure, easy over-fitting based on massive data, high training cost due to complex structural parameters, large consumption of computing resources, and limited capture of long-distance dependence. Looking to the future, integration has become a key trend. Transformer and CNN can work together to accurately anchor key areas and strengthen feature extraction. Together, CNN and LSTM can integrate spatial and temporal information to dig deep into meteorological spatiotemporal laws, comprehensively improve prediction accuracy and efficiency, and continue to expand the depth, breadth and practical value of weather-related prediction technologies.

References

[1] Li, H. (2022). Machine learning methods. Tsinghua University Press.

[2] Lee, J., Hong, B., Lee, K., & Jang, Y. J. (2015, December). A prediction model of traffic congestion using weather data. In 2015 IEEE International Conference on Data Science and Data Intensive Systems (pp. 81-88). IEEE.

[3] Liu, Y., Fang, R., Su, Y., & Xiao, R. (2021). Research on regional landslide disaster warning model based on machine learning. Journal of Engineering Geology, 29(1), 116-124. https://doi.org/10.13544/j.cnki.jeg.2020-533

[4] nardi, S., Sinambela, M., Humaidi, S., & Situmorang, M. (2024, November). Application of convolutional neural networks (CNN) for optimizing route changes based on dynamic weather conditions and travel time. In Proceedings of the National and International Seminar of the Faculty of Engineering and Computer Science, Dharmawangsa University (Vol. 1, No. 1, pp. 183-197).

[5] Betkier, I., & Oszczypała, M. (2024). A novel approach to traffic modelling based on road parameters, weather conditions and GPS data using feedforward neural networks. Expert Systems with Applications, 245, 123067.

[6] herstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306.

[7] Jaderberg, M., Simonyan, K., & Zisserman, A. (2015). Spatial transformer networks. Advances in Neural Information Processing Systems, 28.

[8] Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., & Wang, Y. (2021). Transformer in transformer. Advances in Neural Information Processing Systems, 34, 15908-15919.

[9] Zhao, H., Jiang, L., Jia, J., Torr, P. H., & Koltun, V. (2021). Point transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 16259-16268).

[10] Zhang, Y. (2023). Interactive spatial co-location pattern mining based on machine learning (Doctoral dissertation, Guilin University of Electronic Technology).