Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2021.020105
ISSN 2790-0983 Vol. 2, Issue 1: 38-46

Lightning Warning Methods Based on Machine Learning and Single Station Ground Meteorological Elements

Asia Khan*
LBEF Campus, Nepal
**corresponding author*

Keywords: Machine Learning, Meteorological Elements, Lightning Detection, Early Warning Methods

Abstract: Lightning is a natural phenomenon in nature, and its occurrence rules and effects are very complex. The analysis of meteorological elements plays an extremely important guiding role in lightning prediction. Therefore, this paper uses the method of machine learning to carry out early warning analysis on lightning approaching, in order to obtain weather information in advance and take action to avoid danger. In this paper, the analytic hierarchy process (AHP) and the survey method are mainly used to carry out the correlation analysis on the lightning approaching warning. The survey data shows that in this system, the accuracy rate of lightning early warning using decision tree algorithm can reach about 90%, indicating that machine learning has a good effect in lightning approaching early warning.

1. Introduction

Lightning is a phenomenon of abnormal discharge of potential on the earth surface. In the atmosphere, it interacts with a strong electric field and generates a strong voltage pulse. The superposition and reflection of these pulse waveforms cause a series of changes in electromagnetic strength, polarization effect, etc. Lightning early warning system is a system that applies weather, geographic information system (GIS), global positioning navigation satellite and radar based on single-chip technology. It is the prediction and forecast of meteorological disasters and provides scientific basis for improving the ability of disaster prevention and reduction.

There are many theories about machine learning and lightning approaching warning methods. For example, some scholars study this topic to reduce casualties and economic losses caused by lightning disasters [1-2]. Some scholars have also proposed that CAMS-LNWS adopts a lightning proximity early warning method integrating multiple data, multiple parameters and multiple algorithms, which can automatically generate lightning activity potential prediction results [3-4]. Some scholars put forward a lightning approaching warning method based on the characteristics of

atmospheric electric field in view of the weak applicability of existing lightning warning technology [5-6]. Therefore, this paper studies lightning warning methods and applies machine learning to improve the ability of local weather forecasting and protecting citizens' safety. This research has social benefits and times value.

This paper first studies the demand of the short-time approaching forecast and early warning system, and obtains the basic framework of the system design by analyzing the demand. Secondly, it analyzes the lightning approaching warning prediction based on machine learning, and describes the application of machine learning algorithm in it. Then the lightning approaching warning method based on single station ground meteorological elements is briefly discussed. Finally, the relevant conclusions are drawn through the system design and related investigations.

2. Lightning Approaching Warning Method Based on Machine Learning and Single Station Ground Meteorological Elements

2.1. Demand for Short-Time Approaching Forecast and Early Warning System

Short time proximity prediction system requires fast and accurate response time. In this design, the sensor acquisition module is used to realize the real-time monitoring of the surrounding environment information. Before responding to the early warning signal, it is necessary to first transmit the parameters, positions and other data in the detection area to the microcontroller, and then send an alarm by the buzzer after A/D conversion. In practice, the short-time geological radar velocimeter we designed is developed to measure the velocity, thickness and other information of the surrounding environment at different times near some unknown strata [7-8].

The system requires strong real-time performance. Under abnormal conditions, it can make timely and accurate response and give corresponding accurate information. It can react quickly after analyzing and handling emergencies. The system requires high reliability. The parameters set in the early warning model will change, so a large number of experiments are needed to verify the effectiveness and accuracy of the algorithm, so as to ensure that the performance of the entire system meets the required standards. The short-term proximity prediction system must be real-time and accurate. Under normal working conditions, the detection accuracy of the system should reach the predetermined target. In this paper, it is required to realize fast and effective automatic monitoring of different types of short-time distance measuring instruments. Therefore, it is necessary to design a measuring instrument according to the ranging principle to complete this function. At the same time, it is also necessary to consider whether the components and their parameters used in the sensor structure meet the test conditions, and test their sensitivity, reliability and other performance indicators to ensure that the system can work normally and achieve the intended goals [9-10].

The forecast and early warning system can analyze meteorological data to OpenMap, and display radar, wind direction, lightning, automatic weather station data, etc. The system releases the prediction and early warning signal of the disaster weather falling area by making templates [11-12].

2.2. Lightning Approaching Warning Prediction Based on Machine Learning

The most basic idea in machine learning is to convert the known unknown parameters into the determined model. In a specific environment, the existing data is processed, stored and calculated. Therefore, using computers to complete these tasks requires certain strict requirements. The first

thing to be determined is to use known parameters to obtain unknown quantities. Secondly, we need to know whether these physical information will affect the prediction results. Third, it must meet certain conditions in order to obtain the accuracy value with higher expected accuracy. Fourth, when the required sample is large enough, we can no longer use other types of data for prediction and estimation, otherwise we have to modify and improve the model according to the actual situation [13-14].

Machine learning consists of a small number of training samples. It gets specific information by simulating human cognitive process. Therefore, it is of great significance in processing large amounts of data and detecting anomalies. Neural network is a nonlinear system. Its basic idea is to simplify the complex problem into units or sub models, and transform the input variables into the continuous functional form of parallel structural parameters. At the same time, each neuron can be learned for many times to obtain the global optimal solution, and the weight coefficient and threshold in the perceptron of the entire neural system can be continuously optimized through the adaptive ability. Neural network (BP) is the most widely used network model, which is used to process massive historical data in real time, simulate and identify various types of information resources, and realize the analysis and prediction of a large number of original input sample data. The algorithm has strong learning ability and fault tolerance, and can quickly and efficiently acquire useful knowledge and deal with problems [15-16]. The correlation coefficient of the standardized data is:

$$s_{mn} = \frac{1}{x-1}\sum_{p=1}^{x} a_{pm}a_{pn} \tag{1}$$

Among them, $m,n = 1,2,...,q$. By substituting the standardized original data into the principal component expression, new data of each sample under each principal component can be obtained, that is, the principal component score.

$$\begin{pmatrix} D_{11} & D_{12} & \Lambda & D_{1l} \\ D_{21} & D_{22} & \Lambda & D_{2l} \\ M & M & M & M \\ D_{g1} & D_{g2} & \Lambda & D_{gl} \end{pmatrix} \tag{2}$$

The principal component analysis can be realized by using the principomp function in RSstudio software.

The basic idea of machine learning is to use computer simulation and calculation of a large number of time series data, use the model for statistical analysis, and obtain the change rule of discharge intensity between different positions under lightning impulse. Some methods are commonly used in lightning warning, such as probability distribution function (FAR), random field theory (Monte Carlo method) and Bayesian network prediction algorithm. First, determine a system state or energy level as the signal to be measured and compare it with a certain value. Then the system signal is statistically analyzed to obtain the change rule and time series characteristics of discharge intensity between different positions under lightning impulse [17-18].

The application of machine learning method in lightning warning is to analyze the changes of meteorological elements, geographical distribution characteristics and other influencing factors through computers, and then draw conclusions. The machine learning process includes data preprocessing and state space clustering. The state space clustering needs to be established in

different types of observation environments. For time series classification, it is necessary to consider the influence of weather and other external interference conditions on the sample size, and judge whether appropriate methods can be used to determine whether there is an internal relationship between each parameter in the model. The lightning approaching warning method based on machine learning is mainly to process these high-frequency signals by calculating different frequencies, the same amplitude and polarizability, and obtain an initial state. Then, according to the adaptive filter, the required parameters are modified to determine the data sampling period in each frequency band. Due to the difference between the atmospheric and ground elements, the frequency domain images collected on each band do not have good resolution. This requires adding appropriate thresholds in the subsequent frame structure design to improve the recognition rate and anti-interference ability.

Lightning approaching warning strategies based on machine learning are mainly divided into the following categories. Single parameter regression analysis method, which is based on a certain time, a certain frequency signal as the input sample, determines whether the sample exists in the time series by predicting the independent variable. This statistical model has good global adaptability, but it is widely used because of its complexity and low accuracy. Support vector machine method and grey system theory are combined for discrimination and processing, and single parameter regression analysis is still the most commonly used method when a small number of data are obtained.

2.3. Lightning Approaching Warning Method Based on Single Station Ground Meteorological Elements

Lightning early warning can establish a comprehensive indicator, and select different frequencies, levels and intervals at the monitoring point to collect data for processing and analysis. First, preprocess the collected data. Extract the maximum anomaly intensity of the calculated value when the original observation value is less than or equal to 5W. Then, the critical energy and the total power under the threshold can be calculated by using the weighted summation formula to judge whether there is a discharge phenomenon and analyze the harm of lightning to the ground meteorological elements.

The lightning approaching early warning method based on the meteorological elements of a single station is to calculate the average water vapor in different times according to the atmospheric pressure, wind speed and depth of day to predict the weather change law, and combined with the radar echo signal. By using this model, we can get the number of near ground flashover at normal horizontal distance in each month during thunderstorm period and the nearby surface density, and judge whether there is a positive correlation. The distribution of correlation curve, peak line and anomaly rate are obtained. The lightning approaching warning method based on single station meteorological elements is mainly to obtain the lightning discharge intensity under different weather conditions by collecting data such as lightning strike interval, average voltage amplitude and voltage period. According to the calculation, the maximum frequency of thunderstorm impact in each period can be obtained, so as to judge whether there is a relationship between it and the atmospheric environment. Through acquiring data on the ground and underground, and processing and analyzing the collected data. High frequency pulse signals on the ground are measured by traditional radar rangefinder. The waveform generates a sinusoidal voltage after the high level is measured by the radar rangefinder, causing the receiver resistance to change. When the electromagnetic wave is emitted, the reflectivity increases with the distance.

3. Testing and Investigation of Lightning Approaching Early Warning System

3.1. Structural Design of Lightning Approaching Early Warning System

The lightning short-time approaching early warning system consists of four modules, including data analysis, data collection, early warning analysis, early warning display and prompt, as shown in Figure 1:

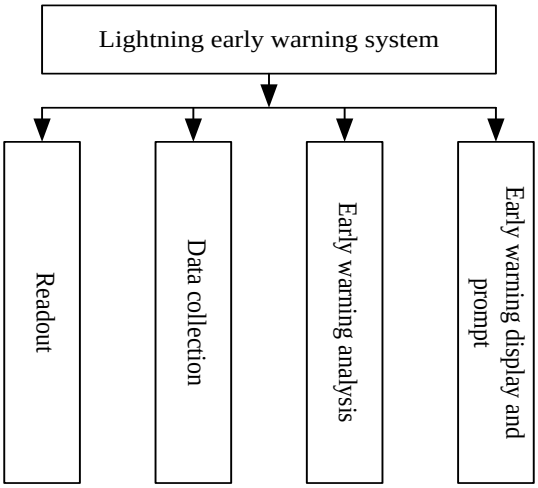

Figure 1. Lightning short-time approaching early warning system

Lightning is based on strong convection, and radar is the most direct observation method for storms, which is related to the characteristics of storms. On the basis of the data analysis of the new generation weather radar and related products, the design mainly focuses on the data used for the new generation weather radar, uses the information advantages of the lightning location network, and uses the specific related meteorological data as an extension, and tries to use the numerical forecast and early warning wind field model to calculate the path of the lightning storm.

3.2. Lightning Warning Test

The background field of early warning is composed of 2-hour cumulative rainfall products, 1-hour cumulative rainfall products and satellite radar factor reflectivity products. The storm path consultation parameter products extrapolate the direction of lightning development in the next 2 hours, and use lightning locators and atmospheric electric field monitors to conduct auxiliary early warning and prediction tests for local and local surrounding areas.

3.3. Verification of Alert Results

This system selects several samples as training samples. According to the analysis of the local lightning potential prediction decision tree, the TS scoring standard is used to preliminarily verify the prediction results of the trial run. The probability of lightning weather in the next 2 hours is shown in "forecast probability", expressed in%. The prediction probability of lightning potential is calculated according to the predicted data, and the probability of lightning occurrence is inferred on 7 probability boundaries, namely 30%, 40%, 50%, 60%, 70%, 80% and 90%.

4. Analysis of Lightning Warning Results

4.1. Local Lightning Potential Forecast in March

This paper makes statistics on the thunderstorm weather forecast data in March and April in this region, and uses machine learning algorithm to make statistics on the relevant data. Table 1 shows the accuracy rate, false alarm rate and false alarm rate of local lightning potential probability in March. Table 1 shows that the accuracy rate increases with the prediction probability.

Table 1. Local lightning potential forecast in March

	Accuracy rate	Empty report rate	Missing report rate
30	55	40.5	4.5
40	59	36.4	4.6
50	72	23.2	4.8
60	78	17	5
70	88	6.9	5.1
80	92	2.3	5.7
90	93.6	0.4	6

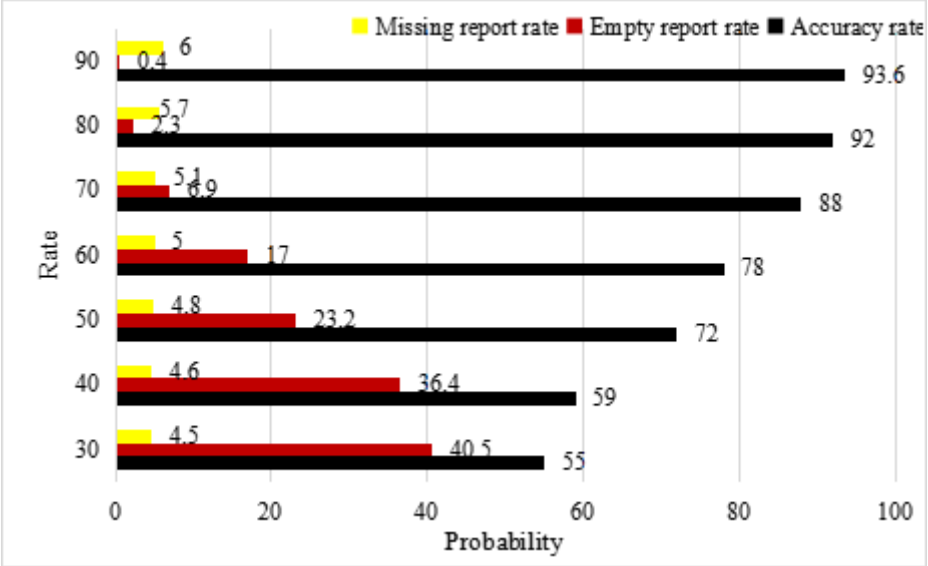

Figure 2. Local lightning potential forecast in March

As shown in Figure 2, when the prediction probability reaches 90%, the accuracy rate reaches 93.6%. When the prediction probability is only 30%, the accuracy rate is only 55%. When the prediction probability reaches 60%, the accuracy rate has reached more than 78%, indicating that the prediction results are good.

4.2. Local April Lightning Potential Forecast

Table 2 shows the accuracy rate, false alarm rate and false alarm rate of local lightning potential probability in March. It can be found from Table 2 that although the accuracy rate is lower than the false rate when the prediction probability is only 30%, the accuracy rate increases with the prediction probability.

Table 2. Local lightning potential forecast in April

	Accuracy rate	Empty report rate	Missing report rate
30	46	49.2	4.8
40	52	43.1	5.1
50	60	33.8	6.2
60	68	24.9	7.1
70	74	18.1	7.9
80	80	11.5	8.5
90	91	0.1	8.9

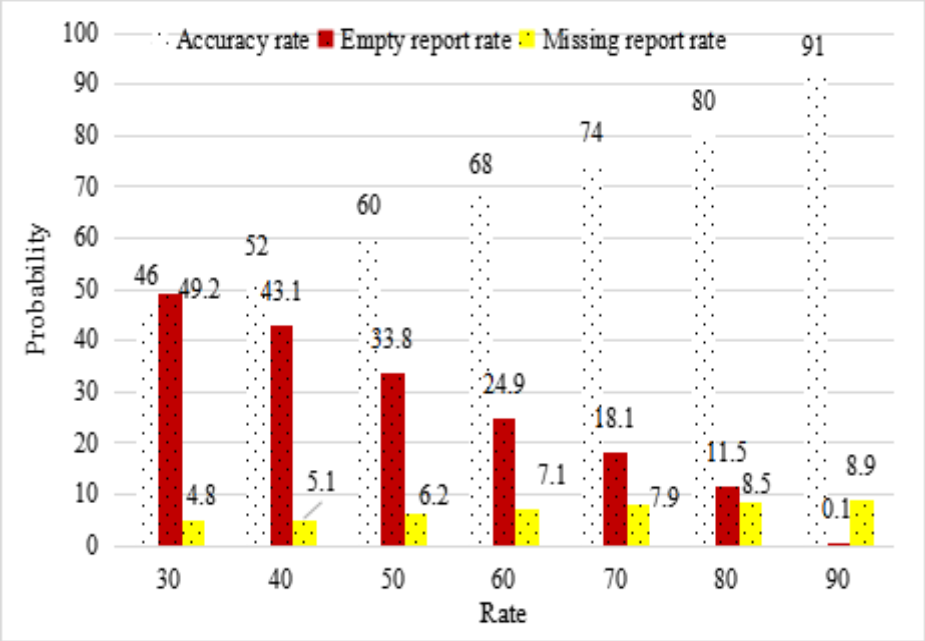

Figure 3. Local lightning potential forecast in April

As shown in Figure 3, we can see that the distribution of lightning potential prediction probability is more than 60%, and the probability of lightning occurrence is high. The use of decision tree has a high accuracy rate for lightning potential prediction, and the effect is significant, especially when it is greater than 70%.

5. Conclusion

Through the monitoring of lightning, this paper uses GIS (spatial analysis and multi parameter comprehensive modeling) technology to establish anomaly rate models of different meteorological elements, atmospheric average intensity weighting and traditional voltage pulse amplitude inversion methods to realize lightning strike proximity recognition. In this paper, the lightning early warning methods based on the ground meteorological elements of a single station are studied and summarized in depth, and combined with practical applications, on the basis of the measured data of a single station, a lightning strike approaching early warning system is established based on regression analysis, using time series correction algorithm and other technologies. Due to the complexity of the detection environment, the relevant experiments in this paper cannot meet the detection requirements under all environmental conditions.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Irshad Ullah, Md Nor Ramdon Bin Baharom, Hussein bin Ahmad, Fazli Wahid, Hakim bin Mahmod Luqman, Zainab Zainal, Bhagwan Das: Smart Lightning Detection System for Smart-City Infrastructure Using Artificial Neural Network. Wirel. Pers. Commun. 106(4): 1743-1766 (2019). https://doi.org/10.1007/s11277-018-5383-4

[2] Pierre Kokou, Philip Willemsen, Mounir Lekouara, Madani Arioua, Andreu Mora, Pieter Van den Braembussche, Emanuele Neri, Donny M. A. Aminou: Algorithmic Chain for Lightning Detection and False Event Filtering Based on the MTG Lightning Imager. IEEE Trans. Geosci. Remote. Sens. 56(9): 5115-5124 (2018). https://doi.org/10.1109/TGRS.2018.2808965

[3] Anastasios Petropoulos, Vasilis Siakoulis, Evangelos Stavroulakis: Towards an Early Warning System for Sovereign Defaults Leveraging on Machine Learning Methodologies. Intell. Syst. Account. Finance Manag. 29(2): 118-129 (2022). https://doi.org/10.1002/isaf.1516

[4] Silvester Dian Handy Permana, Gusti Saputra, Budi Arifitama, Yaddarabullah, Wahyu Caesarendra, Robbi Rahim: Classification of bird sounds as an early warning method of forest fires using Convolutional Neural Network (CNN) algorithm. J. King Saud Univ. Comput. Inf. Sci. 34(7): 4345-4357 (2022). https://doi.org/10.1016/j.jksuci.2021.04.013

[5] Gemma Cremen, Carmine Galasso: A Decision-Making Methodology for Risk-Informed Earthquake Early Warning. Comput. Aided Civ. Infrastructure Eng. 36(6): 747-761 (2021). https://doi.org/10.1111/mice.12670

[6] Maman Abdurohman, Aji Gautama Putrada, Mustafa Mat Deris: A Robust Internet of Things-Based Aquarium Control System Using Decision Tree Regression Algorithm. IEEE Access 10: 56937-56951 (2022). https://doi.org/10.1109/ACCESS.2022.3177225

[7] Hadeer Mahmoud, Mostafa Thabet, Mohamed Helmy Khafagy, Fatma A. Omara: Multiobjective Task Scheduling in Cloud Environment Using Decision Tree Algorithm. IEEE Access 10: 36140-36151 (2022). https://doi.org/10.1109/ACCESS.2022.3163273

[8] Marziye Narangifard, Hooman Tahayori, Hamid Reza Ghaedsharaf, Mehrdad Tirandazian: Early Diagnosis of Coronary Artery Disease by SVM, Decision Tree Algorithms and Ensemble Methods. Int. J. Medical Eng. Informatics 14(4): 295-305 (2022). https://doi.org/10.1504/IJMEI.2022.123921

[9] Chandrashekhar Azad, Bharat Bhushan, Rohit Sharma, Achyut Shankar, Krishna Kant Singh, Aditya Khamparia: Prediction Model Using SMOTE, Genetic Algorithm and Decision Tree (PMSGD) for Classification of Diabetes Mellitus. Multim. Syst. 28(4): 1289-1307 (2022). https://doi.org/10.1007/s00530-021-00817-2

[10] Love Allen Chijioke Ahakonye, Cosmas Ifeanyi Nwakanma, Jae-Min Lee, Dong-Seong Kim: Efficient Classification of Enciphered SCADA Network Traffic in Smart Factory Using Decision Tree Algorithm. IEEE Access 9: 154892-154901 (2021). https://doi.org/10.1109/ACCESS.2021.3127560

[11] Leidiane C. M. M. Fontoura, Hertz Wilton De Castro Lins, Arthur S. Bertuleza, Adaildo G. D'Assunção, Alfredo Gomes Neto: Synthesis of Multiband Frequency Selective Surfaces Using Machine Learning With the Decision Tree Algorithm. IEEE Access 9: 85785-85794 (2021). https://doi.org/10.1109/ACCESS.2021.3086777

[12] Kumpol Saengtabtim, Natt Leelawat, Jing Tang, Wanit Treeranurat, Narunporn Wisittiwong, Anawat Suppasri, Kwanchai Pakoksung, Fumihiko Imamura, Noriyuki Takahashi, Ingrid Charvet: Predictive Analysis of the Building Damage From the 2011 Great East Japan Tsunami Using Decision Tree Classification Related Algorithms. IEEE Access 9: 31065-31077 (2021). https://doi.org/10.1109/ACCESS.2021.3060114

[13] Ferdinand Bollwein, Stephan Westphal: A Branch & Bound Algorithm to Determine Optimal Bivariate Splits for Oblique Decision Tree Induction. Appl. Intell. 51(10): 7552-7572 (2021). https://doi.org/10.1007/s10489-021-02281-x

[14] Muhamad Hasbullah Mohd Razali, Rizauddin Saian, Yap Bee Wah, Ku Ruhana Ku-Mahamud: An Improved ACO-Based Decision Tree Algorithm for Imbalanced Datasets. Int. J. Math. Model. Numer. Optimisation 11(4): 412-427 (2021). https://doi.org/10.1504/IJMMNO.2021.10041068

[15] Firoozeh Karimi, Selima Sultana, Ali Shirzadi Babakan, Shan Suthaharan: Urban Expansion Modeling Using an Enhanced Decision Tree Algorithm. GeoInformatica 25(4): 715-731 (2021). https://doi.org/10.1007/s10707-019-00377-8

[16] Matheus Guedes Vilas Boas, Haroldo Gambini Santos, Luiz Henrique de Campos Merschmann, Greet Vanden Berghe: Optimal Decision Trees for the Algorithm Selection Problem: Integer Programming Based Approaches. Int. Trans. Oper. Res. 28(5): 2759-2781 (2021). https://doi.org/10.1111/itor.12724

[17] Rajendra Mahla, Baseem Khan, Om Prakash Mahela, Anup Singh: Recognition of Complex and Multiple Power Quality Disturbances Using Wavelet Packet-Based Fast Kurtogram and Ruled Decision Tree Algorithm. Int. J. Model. Simul. Sci. Comput. 12(5): 2150032:1-2150032:23 (2021). https://doi.org/10.1142/S179396232150032X

[18] Vinay Arora, Rohan Singh Leekha, Inderveer Chana: An Efficacy of Spectral Features with Boosted Decision Tree Algorithm for Automatic Heart Sound Classification. J. Medical Imaging Health Informatics 11(2): 513-528 (2021). https://doi.org/10.1166/jmihi.2021.3287