Machine Learning Applications in Natural Disastermanagement

Yuxuan Wang *

Miami College, Henan University, Kaifeng, China

* Corresponding Author Email: 13603553921@163.com

Abstract. In the recent century, humans have been using fossil fuels without restraint, causing significant changes in the Earth's atmospheric environment, resulting in increased intensity and frequency of extreme weather/climate events and natural disasters. Natural disasters bring enormous risks and economic losses to human society, and natural disasters are a major challenge that human society have to face. Applying the scientific technology to predict, monitor, evaluate and manage the natural disasters has become a concern for governments, academia, and the general public all over the world. In recent years, the remarkable improvement of science and technology lead to a great concern of the application of machine learning technology of monitoring and managing the natural disasters Based on literature retrieval of professional databases. This article review some research papers related in model algorithms of machine learning, application status and future research directions of machine learning applications in natural disasters such as earthquakes, geological disasters, meteorological disasters.

Keywords: Machine learning; algorithm; disaster management.

1. Introduction

With the springing up of modern science and technology and the rapid expansion of scientific data, natural disaster research has entered a new era characterized by new technologies, new data, and new methods. With a spurt progress in artificial intelligence, humans can use machine learning models to take a series of research, such as the tuning process of the model, the data acquisition process, the remote sensing technology application to accurately predict, monitor, manage earthquakes, geological disasters, meteorological disasters such as floods, hurricanes, storm surges, etc.

2. Overview of Machine Learning Methods

Based on current research literatures, there are usually several machine learning algorithm models applied in the natural disaster management.

2.1. Artificial Neural Network (ANN)

ANN is a widely parallel interconnected network of adaptive simple units, organized to simulate the interactive response of biological neural systems to the real world. It has an input layer, a hidden layer and an output layer, each layer consist of multi- neurons, and neurons in each layer are connected through the weight matrix. The ANN model can train the data to adjust the weight of the connections between each neuron node, which is suitable for event processing with large amounts of data.

2.2. Support Vector Machine (SVM)

SVM is a model used to handle two classifications and multiple classifications. It is a linear classifier in the feature space. The key problem of optimizing SVM is to find the hyperplane with the largest interval between sample points which can be formalized as a problem for solving convex quadratic programming. Moreover, when dealing with the case where the input data is linearly inseparable in the original space, the SVM can introduce kernel functions that map the low-dimensional data to a high-dimensional space for partitioning.

2.3. Decision Tree (DT)

DT is essentially a similitude comparison algorithm based on tree structure. The nodes of numbers represent the judgments of input essential Factors, while the differences in numbers represent the outputs of each judgment result. The final leaf is obtained through multiple output processes, and the nodes represent the final output of the model. DT is commonly used in classification and regression problems.

2.4. Random Forest (RF)

RF is a bagging decision tree-based integration algorithm, which can summarize the prediction results of all decision trees to form the final prediction. Random forests select optimal features for division with different subsets of features each time, introducing nonlinearity to enable more efficient decision boundaries in the feature space. Because of the high integration of this algorithm, it can cope well with the overfitting problems.

2.5. Logic Regression (LR)

LR is a linear classifier that maps the data features to a probability value in the 0-1 interval through the Logistic function, and then obtains the data classification by comparison with 0.5. It is often used to solve the binary classification problems.

$$\mathbf{f(z)} = \frac{\mathbf{1}}{\mathbf{1+e^{-z}}} \tag{1}$$

$$\mathbf{z = w^T x + w_0} \tag{2}$$

2.6. Weight of Evidence (WOE)

WOE is a form of encoding of the original independent variable, which is the logarithm of the proportion of good and bad customers for a value of a character variable or a segment of a continuous variable. It is often used for feature selection and evaluation.

2.7. XGBoost

XGBoost is a measure using to construct supervised regression models, belonging to the Boosting algorithm family Gradient Boosting Decision Tree (GBDT)algorithm framework. XGBoost has significant advantages in handling multivariate nonlinear regression problems such as disaster loss prediction, due to its regularization, parallel computing, and introduction of feature subsampling, which can avoid overfitting while reducing computation.

3. Machine Learning Applications in Natural Disasters

3.1. Earthquake

After an earthquake, due to complex geological conditions and many unpredictable situations, many monitoring and detection tasks cannot be carried out on the ground. Although remote sensing technology has made up for this, traditional remote sensing technology lacks integration with other technologies and still cannot meet practical needs. Giorgia Giardina [1] and his team designed a new damage assessment system which using post-earthquake synthetic aperture radar (SAR) data to address this issue. This method utilizes high-resolution post-earthquake comprehensive SAR data and has been applied to the earthquake which happened in Haiti in 2021 and Grace tropical cyclone.

In 2022, Pankaj Chittora et al [2] proposed different machine learning models to predict the probability of earthquakes. They used artificial neural network, random tree, CHAID, Discriminant, XGBoost Tree and Tree-AS to train, test and evaluate seismic data sets in six regions including North India and Andaman and Nicobar. They use 70% of the data for model training, 30% for model testing, and do the results comparison. In five of the six regions, XGBoost Tree and Tree-AS have achieved more than 90% prediction accuracy. In addition, the researchers also fitted the different factors in the dataset.

An artificial neural network model is proposed by Rienna [3] to predict the number of casualties and the number of damaged buildings in earthquakes, which is trained on data from Indonesian earthquakes. This artificial neural network model sets the training data vector for 1000 iterations and the neurons amount in the output layer and the input layer to 1. In particular, the hidden layer number of neurons is set to 95, this guarantees the value of correlation index gets the best outcome which is 0. 99971.The authors compared the results with those of previous researchers, RADIUS Programme Darpito [4] and Badal et al [5], and found that the model had higher accuracy.

The traditional machine learning source depth detection model has high feature requirements and is vulnerable to factors such as signal-to-noise ratio. Therefore, De-He Yang et al [6] proposed to use convolutional neural network technology to distinguish deep and shallow earthquakes. They assessed 444 microearthquakes occurring in underground caves in South Louisiana and compared the results with other models. The CNN model adopted by the authors has 10 layers, The component of the input layer is a feature map of 16 * 40 pixels, and the number of deep and shallow layers is 1778 times. At the same time, he, in order to select the best characteristics related to the target, adopted filters, wrappers, and principal component analysis (PCA) to select the most relevant classification features. By comparing the results, the author found that although the traditional classification model, such as SVM, RF reliability is weaker than CNN, these models can still have good performance after using PCA conversion technology, which also shows the importance of influencing feature selection and conversion for the results of the model.

3.2. Geologic Hazard (Collapse, Landslide, Debris flow, volcanics ,etc)

Yu Bian et al [7] using multi-source remote sensing data, identified nine important influencing variables, assessed the geologic disaster risk of Atal tunnel on the Qinghai-Tibet Plateau, and compared the accuracy of the four models: weight of evidence (WOE), support vector machine (SVM), logistic regression (LR), and frequency ratio (FR).To accurately obtain the accuracy of the different models, the authors used the hazard percentage included in each risk level, the subject's working characteristic curve (ROC), and the area under the curve (AUC) value. Finally, the WOE-LR model achieved 90.7% accuracy in the risk prediction of the Atal tunnel, which proved that the model was effective and profound.

Rongwei Li [8] Believed that the traditional geological disaster model has limitations, because they do not consider the weight of the evaluation factor, but superposition the information of each evaluation factor, which will have a negative impact on the results. Therefore, he adopted the random forest enabling information algorithm to solve this problem. This experiment optimizes the forest model and assigns them to each target factor.

For volcanic disasters with rich characteristics, the traditional methods cannot felicitously reflect the complex characteristics of many disasters happened in volcano scenarios in remote sensing images. To address this issue, Chengfan Li [9] offer a novel network framework with Swin-T and attention mechanisms to classify volcanic disasters (MI-STA). Swin-T was first used to get multi-aspect features of disaster locations from remote sensing images, and then to integrate channel attention and spatial attention modules. Finally, the importance weight scores of the different instance features are scored by the already trained instance scorer. When extracting features, the researchers converted the remote sensing images into non-overlapping image subblocks, and then projected them into other dimensional space through the linear embedding layer, and then put the projection into the feature extraction stage. After feature extraction, the generated feature vectors are input into the channel attention module, and then the channel global features are generated by sharing convolutional layers and average pooling. Then input it to the spatial attention module to generate the weight parameters, and finally get the output characteristics. Finally, the mloU, F1 and OA reached 82.55%, 86.79% and 90.93%, respectively, which is more reliable than other models.

3.3. Meteorological Disaster (Floods, hurricanes, storm surges, etc)

Hurricanes, floods, and storm surges are meteorological disasters that scholars have studied extensively, and the research mainly focuses on risk assessment, disaster management, and disaster prevention and reduction [10]. With the warming climate and accelerated urbanization, disasters such as rainstorm and flood have become more intense and frequent than ever before.

3.3.1 Flood

Nazim Razali et al [11] built a model for predicting risk of flood happening in the Kuala Krai, Kelantan, Malaysia using machine learning algorithms such as Bayesian Networks (BN), Decision Trees (DT), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), based on the Cross-Industry Standard Process for Data Mining (CRISP-DM). The authors counted the flood data for the five years of 2012-2016, including 1,827 instances and 8 features. In response to the data imbalance, they used the Synthetic Minority Oversampling Technique (SMOTE). Accuracy, precision, recall and f-scores were used as model evaluation indicators for this experiment. From the conventional results, BN results ranked first with 99.94% accuracy, but after applying SMOTE method, DT and KNN achieved better performance. They point out that the BN and SMOTE methods can be more widely applied for future predictions.

A stacking integration algorithm (RF-XGB-CB-LR) is proposed by YANG Huilin et al [12] to realize the assessment of regional flood sensitivity. They pointed out that the stacking models will assign different weights according to the performance of different models, which can not only reduce the error caused by the model limitations, but also the stacking models have stronger generalization ability and can properly cope with the overfitting problem. The experimenter used the data from the Xiangjiang River Basin (XRB) in China, selected 18 flood-sensitive impact factors and used Python to evaluate the weight of the impact factors according on the data, and then tested the modeling. The final stacking model proved the reliability of the finding with a high score of 0.99. They also found that the northeastern and southeastern parts of the Xiangjiang River were vulnerable to flooding, suggesting that the factors were related to low-lying terrain and human activities in addition to rainfall.

Marcel Motta et al [13] Developed a flood prediction system that combines machine learning technology and GIS spatial technology. The authors say that the independent use of machine learning techniques may be influenced by both the limited data availability, and the lack of morphogen performance. Previous scholar Chen [14] has mentioned that the lack of high-resolution topographic and hydrological data will affect the establishment and reliability of flood prediction models. To address this issue, the authors introduce a GIS spatial model for the whole city to identify and detect spatial clusters. Marcel the two techniques were combined to detect the sensitivity to the flooding caused by Storm Elsa and Fabien in 2019. In the modeling phase, the authors evaluated the performance of six machine learning models and found that the random forest model performed the best in this test. Later, he used the hot spots observed by GIS in the historical record to adjust the sensitivity of the model, and finally achieved good robustness.

3.3.2 Hurricane

A hurricane, also known as a tropical cyclone (TC) storm, poses a serious threat to human life and property safety, so it is necessary to use satellite images to detect hurricanes in advance. Soner Kızıloluk et al [15] proposed an AlexNet hyperparameter algorithm optimized by artificial jellyfish search. This method is also known as the hurricane-faster R-CNN-JS. In this experiment, JS algorithm optimized five aspects including learning rate of AlexNet. From the results, the detection performance of the adjusted method is about 10% higher than that of the AlexNet method without JS optimization, and therefore, this model can be considered reliable.

How to efficiently evacuate the affected people in a hurricane has always been a worthwhile research topic. Traditional evacuation prediction model relies on psychology, Prosper K. Anyidoho [16] for this problem proposed a can predict the family evacuation level of enhanced logic regression model, the principle is through low depth decision tree to realize the model of non-linear and interaction function, the author adopted the Katrina and Rita two nearly 20 years the strongest hurricane data set test this model, and compare the results with the previous model. The results show that the enhanced logistic regression model proposed in this experiment is far better than the previous models.

4. Current status and prospects of machine learning application in natural disaster management

In the current context of the lack of basic disaster data and the accumulation of diverse data, how to construct disaster assessment models based on data fusion perspective, use machine learning technology to improve the response speed to natural disasters, prevent and reduce disaster hazards scientifically has become an urgent and necessary practical problem. With the development of 3S technology (Geographic Information System, Remote Sense, Global Positioning System, GIS+RS+GPS), RS and GIS technology have been widely applied in disaster assessment. Remote sensing images are used for automatic mapping after disasters, and machine learning classification technology is used to automatically analyze or label the damage in remote sensing images, extract disaster information data, evaluate the damaged carriers, quantities, ranges, etc., making disaster assessment more convenient and feasible. However, the characteristics of remote sensing mapping in terms of sensor field of view, image scale, scene complexity, etc. may vary, and there is also uncertainty in information extraction, which may restrict the accuracy of loss assessment [11]. Moreover, there are also some drawbacks to using remote sensing technology to detect and monitor natural disasters. Collecting images, feedback images and band information may take a long time, require high information technology processing, and have insufficient interaction with affected populations [17]. Yang and Cervone [18] pointed out using deep learning algorithms to automatically extract relevant information such as damage, disasters, and damages to key infrastructures from images. It is certain that machine learning in the future will be the mainstream method of natural disasters, but the mainstream algorithm model is still facing data factor weight allocation and data capture technology defects, future research should be focused on these several aspects, at the same time might also focus on the current relatively few fields of post-disaster reconstruction recovery.

5. Conclusion

This article cites multiple instances of various algorithms to solve natural disasters. Obviously, machine learning algorithms can now be used for earthquakes, geological disasters, flood prediction, loss assessment, and survey. The data of natural disasters is often complex and difficult to handle, but machine learning models can take advantage of this to self-tune and give feedback. The advantages of machine learning are still its sensitivity to different factor weights and its high efficiency and accuracy in large-scale data processing. It is certain that machine learning in the future will be the mainstream method of natural disasters, but the mainstream algorithm model is still facing data factor weight allocation and data capture technology defects, future research should focus on the several aspects, at the same time also can focus on the current relatively few fields of post-disaster reconstruction recovery.

References

[1] Giardina G, Macchiarulo V, Foroughnia F, et al. Combining remote sensing techniques and field surveys for post-earthquake reconnaissance missions. Bulletin of Earthquake Engineering, 2024, 22(7): 3415-3439.

[2] Chittora P, Chakrabarti T, Debnath P, et al. Experimental analysis of earthquake prediction using machine learning classifiers, curve fitting, and neural modeling. 2022.

[3] Oktarina R, Bahagia S N, Diawati L, et al. Artificial neural network for predicting earthquake casualties and damages in Indonesia//IOP Conference Series: Earth and Environmental Science. IOP Publishing, 2020, 426(1): 012156.

[4] DARPITO K. Earthquake damage assessment: application and verification of the radius method to yogyakarta earthquake 2006. Bulletin of the International Institute of Seismology and Earthquake Engineering, 2011, 45: 85-90.

[5] Badal J, Vazquez-Prada M, González Á. Preliminary quantitative assessment of earthquake casualties and damages. Natural Hazards, 2005, 34: 353-374.

[6] Yang D H, Zhou X, Wang X Y, et al. Mirco-earthquake source depth detection using machine learning techniques. Information Sciences, 2021, 544: 325-342.

[7] Bian Y, Chen H, Liu Z, et al. Geological Disaster Susceptibility Evaluation Using Machine Learning: A Case Study of the Atal Tunnel in Tibetan Plateau. Sustainability, 2024, 16(11): 4604.

[8] Li R, Tan S, Zhang M, et al. Geological Disaster Susceptibility Evaluation Using a Random Forest Empowerment Information Quantity Model. Sustainability, 2024, 16(2): 765.

[9] Li C, Han J, Wu C, et al. Volcanic disaster scene classification of remote sensing image based on deep multi-instance network. Acta Geophysica, 2024: 1-17.

[10] Tan L. Research on Economic Loss Assessment of Urban Rainstorm Flood Disasters-Based on the Perspective of Data Fusion. 2022.

[11] Razali N, Ismail S, Mustapha A. Machine learning approach for flood risks prediction. IAES International Journal of Artificial Intelligence, 2020, 9(1): 73.

[12] Yang H, Yao R, Dong L, et al. Advancing flood susceptibility modeling using stacking ensemble machine learning: A multi-model approach. Journal of Geographical Sciences, 2024, 34(8): 1513-1536.

[13] Motta M, de Castro Neto M, Sarmento P. A mixed approach for urban flood prediction using Machine Learning and GIS. International journal of disaster risk reduction, 2021, 56: 102154.

[14] Chen J, Hill A A, Urbano L D. A GIS-based model for urban flood inundation. Journal of Hydrology, 2009, 373(1-2): 184-192.

[15] Kızıloluk S, Sert E. Hurricane-Faster R-CNN-JS: Hurricane detection with faster R-CNN using artificial Jellyfish Search (JS) optimizer. Multimedia Tools and Applications, 2022, 81(26): 37981-37999.

[16] Anyidoho P K, Ju X, Davidson R A, et al. A machine learning approach for predicting hurricane evacuee destination location using smartphone location data. Computational Urban Science, 2023, 3(1): 30.

[17] Wu K, Wu J, Ye M. A review on the application of social media data in natural disaster emergency management. Prog. Geogr, 2020, 39: 1412-1422.

[18] Yang L, Cervone G. Analysis of remote sensing imagery for disaster assessment using deep learning: a case study of flooding event. Soft Computing, 2019, 23(24): 13393-13408.