Overcoming Extreme Weather: The Insurance Industry's Road to Recovery

Kexin Wang*

Business School, Xi'an International Studies University, Xi'an, 710100, China

*Corresponding author: 18208768160@163.com

Abstract. In recent years, there has been an increase in extreme weather events globally, posing challenges to the economy and the property insurance industry. The property-casualty insurance industry needs to adapt to enhance its underwriting capacity and service levels to cope with climate change risks. In order to assess whether insurance companies should underwrite in areas facing an increase in extreme weather events, this paper establishes a risk assessment model. At first, high-frequency indicators are selected as the basis for judgment, and their correlation analysis is performed. And the Bayesian formula was used to estimate the probability of occurrence of a specific extreme weather event. Next, an LSTM model was constructed using historical loss data for predicting the amount of compensation loss in different regions. Then, the TF-IDF algorithm was used to calculate the weights of the events according to their impact on the regions, so as to calculate the comprehensive risk scores of the regions. Finally, the quantitative data of the indicators are used as inputs to the SVM model to categorize the risk of different regions and provide a scientific basis for insurers' pricing and decision-making.

Keywords: Insurance industry; Extreme weather; Risk assessment model; LSTM model.

1 Introduction

Since the Industrial Revolution, anthropogenic emissions of carbon dioxide have steadily increased global surface temperatures and significantly altered terrestrial precipitation patterns, the warming effect that has increased the frequency and intensity of extreme weather events (Newman, R., & Noy (2023)[1]; Raihan, 2023[2]; Geng et al., 2023[3]). In recent years, there has been an increase in the number of extreme weather events around the globe, which can not only have an obvious impact on global economic development, but also create a crisis for the property insurance industry and property owners (Zhong et al., 2023[4]; Kaur, 2024[5]).

Whether now or in the past, whether at home or abroad, research on the development of the property insurance industry has always been an important part of insurance development theory (Zhou et al., 2023[6]). The core of the development of the property insurance industry lies in its business, and the improvement of its underwriting capacity

J. Liao et al. (eds.), *Proceedings of the 2024 2nd International Conference on Digital Economy and Management Science (CDEMS 2024)*, Advances in Economics, Business and Management Research 292,
https://doi.org/10.2991/978-94-6463-488-4_51

and services determines the success or failure of the industry's development (Osariemen & Benedicta, 2023[7]). Nowadays, with the occurrence of extreme weather events, it is of great significance to study how insurance companies should adjust their property insurance business.

Considering the background information and restricted conditions identified in the problem statement, we need to solve the following problem: a model needs to be built to help an insurance company determine under what circumstances an insurance company can underwrite a certain area under extreme weather conditions; and what impact the homeowner can have on the underwriting.

2 Risk Assessment Model

2.1 Preparation of Model

In order to build a reliable model, this paper collects meteorological data (including precipitation, humidity, etc.), geographic data, socio-economic statistics, and disaster statistics of the relevant regions. Due to the large amount of data in the world, this paper chooses two regions, the United States and China, for model demonstration, and some of the acquired data are visualized as figure 1:

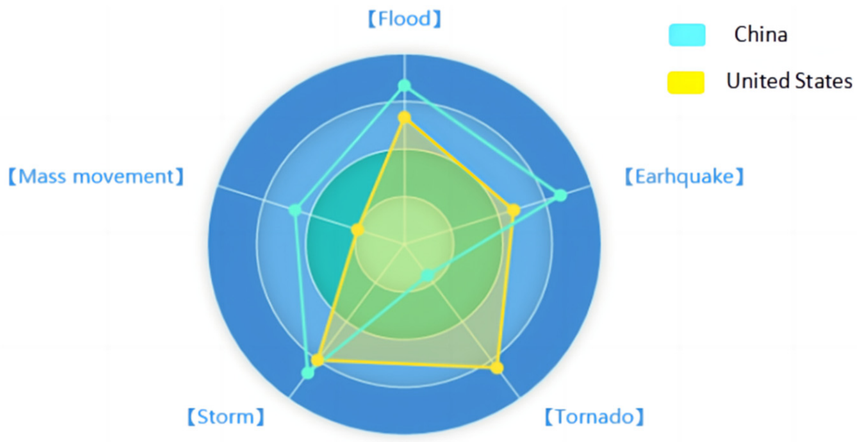

Fig. 1. Radar Map of Natural Disasters in China and the United States Region

2.2 Index Selection

In this paper, I select the indicators that appear more frequently to determine whether an insurance company writes policies in a certain region from three aspects, namely, geographic factors, economic factors, and the ability of property owners to pay, respectively. The specific indicators selected are shown in the figure 2:

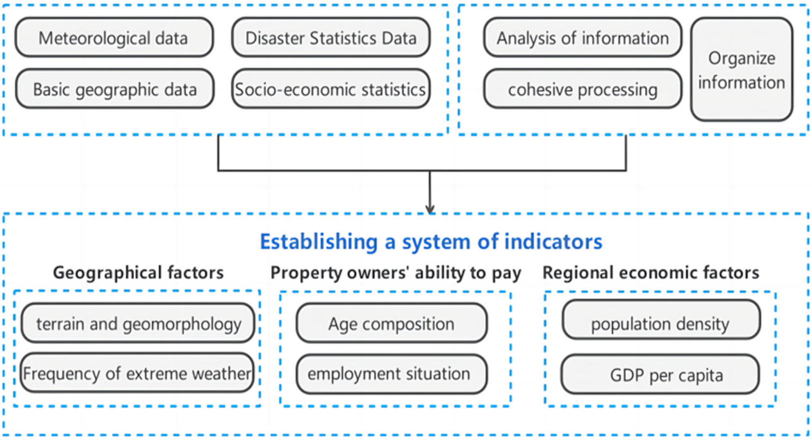

Fig. 2. Flowchart for the Identification of Indicators

2.3 Relational Model

The Spearman correlation coefficient is a statistical measure of the correlation between two sets of independently and identically distributed data (Ali, 2022[8]). In order to determine the relationship between the indicators, this paper establishes a correlation analysis model to determine the degree of correlation between the indicators selected above through the Spearman coefficient.

2.4 Probability of Occurrence of Extreme Weather Events

The Bayesian formula is shown below:

$$P(A_i \mid B) = \frac{P(A_i B)}{P(B)} = \frac{P(A_i)P(B \mid A_i)}{P(B)} \tag{1}$$

In this paper, $P(A_i|B)$ is the probability that extreme weather A_i occurs when data B is observed, where i denotes different types of extreme weather events. $P(B|A_i)$is the probability that data B is observed given that time A_i occurs, and $P(B)$ is the probability that data B is observed.

2.5 LSTM Prediction Model

In order to more comprehensively predict the frequency of future extreme weather in the selected region, this paper considers a Long Short-Term Memory (LSTM) neural network model (Kumari & Toshniwal, 2021[9]). The LSTM model adopts a gated output model, including three gates (input gate, forget gate, output gate) and two states (long unit state, short hidden state). The core of which is the cell state, which is used for information transfer. Its specific structure is shown in figure 3:

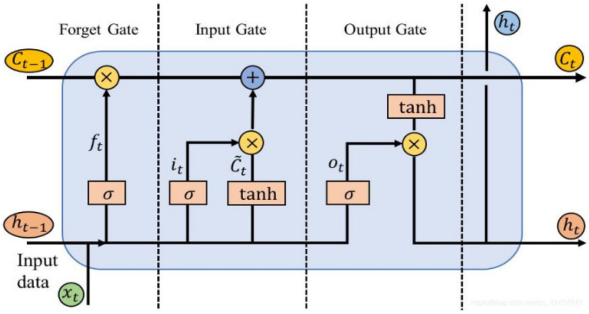

Fig. 3. Structure of the LSTM Model

2.6 Risk Scores with Weights

The composite risk score for an area can be calculated by considering the event weights, probability of occurrence and potential losses. Its value is the sum of the weights of all extreme events and is given by the following formula:

$$M = \sum_i \varpi_i \cdot L(A_i) \cdot P(A_i) \tag{2}$$

ω is the weight, reflecting the extent to which event A_i affects the overall risk. For example, if an area is more susceptible to earthquakes but flooding is infrequent, the weight of flooding is less than the weight of earthquakes on the risk to the area.

$TF - IDF$ is an algorithm commonly used in information retrieval and text mining. It is used to evaluate the importance of a word to a document in a collection of documents. TF refers to the word frequency, which indicates how often a word appears in a document, and the more times a word appears in a document, the more important it is. IDF refers to Inverse Document Frequency, which indicates the prevalence of a word in the entire document collection. The less frequently a word appears in the entire document collection, the greater its importance. the $TF - IDF$ is calculated as:

$$TF - IDF = TF \cdot IDF \tag{3}$$

$$TF = \frac{\text{Word Frequency}}{\text{Total words in document}} \tag{4}$$

$$IDF = \frac{\text{Total number of documents}}{\text{Number of documents containing the word}} \tag{5}$$

2.7 Classification Models

SVM is a commonly used machine learning algorithm for classification and regression tasks (Abdullah & Abdulazeez, 2021[10]). It separates samples of different categories by finding a hyperplane that maximizes the interval between classification boundaries.

In this paper, SVM maps the sample data x of U in the indicators affecting the development of insurance companies to the feature space, which can be expressed as:

$$z = \frac{P(t|C_k)\varphi(x)}{U} \tag{6}$$

$\varphi(x)$is the feature space. In the high-dimensional feature space, when considering the linear divisibility of the data, it is necessary to ensure that the weight vector w and bias b satisfy certain constraints:

$$(w, \varphi, (x_i)) + b \geq 1 \tag{7}$$

Then the SVM component hyperplane is obtained:

$$f(x) = \langle w, \varphi(x) \rangle + b \tag{8}$$

Introducing the relaxation factor:

$$\zeta_i = 1 - \langle w, \varphi(x) \rangle + b \tag{9}$$

For the classification of nonlinear data, the support vector machine algorithm needs to use the kernel function to map the data to solve the optimal hyperplane, the kernel function expression is:

$$K(x_i, x_j) = x_i \cdot x_j \tag{10}$$

In the problem of regional property risk assessment, there are usually nonlinear and complex features. When choosing the kernel function, the RBF kernel function requires only one parameter, so it is more appropriate to choose the RBF kernel function.

3 Results

3.1 Correlation Analysis

According to the figure 4, there is a weak correlation between the indicators, which indicates that the selected indicators have more comprehensive characteristics. Based on this, the adopted indicator system can be considered reasonable.

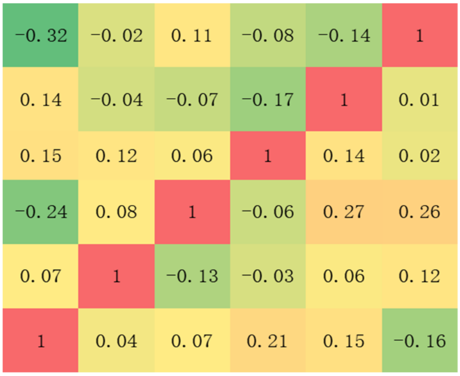

Fig. 4. Heat Map of Correlation Coefficients between Indicators

3.2 Historical Data on the Amount of Losses in the Two Regions of China and the United States

First, the data statistics are carried out. According to the collected historical data on the annual loss amount of the two regions of the United States and China, the visualization results are as figure 5:

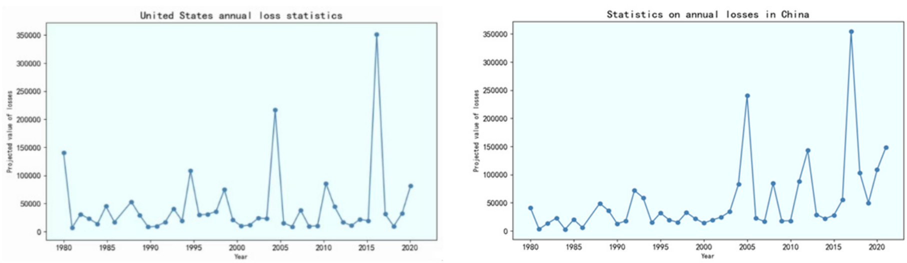

Fig. 5. Annual Loss Statistics for United States and China

Second, the dataset is divided into a training set and a test set. The model is trained using the training set with a specified number of iterations. Finally, the trained model is used to predict the test set, predicting the number of extreme weather occurrences in each region in the next five years, and back-normalizing the prediction results to obtain the true prediction value. Some of the prediction results are shown in table 1:

Table 1. Projected Losses for the Next Five Years in China and the United States Region

	2022	2023	2024	2025	2026
US	132458	84567	112465	113245	120874
China	121475	985413	104536	135167	131452

In the third step, evaluation metrics are adopted to measure the prediction accuracy. The results of LSTM model evaluation will be presented in Table 2:

Table 2. The Results of LSTM Model Valuation

Evaluation indicators	Numeric size
MAE	61.24
RMSE	83.94

From the above evaluation metrics, it can be concluded that the model is well-trained, performs well on the test set, which has a small error and a high prediction accuracy.

3.3 Risk Scores with Weights

By collecting articles on regional disaster factors in China and the United States to calculate the TF-IDF value for each word, the importance of the impact of extreme weather event A_ion a region can be determined, which can then be used to calculate the risk score with weights. Some of the keyword weights are listed in table 3:

Table 3. Partial Keyword Weights

Keywords	Weights in China	Weights in US
Flooding	0.07682	0.06673
Earthquake	0.06421	0.03047
Tornado	0.01293	0.07443

3.4 Regional Classification Results

By using the quantified indicators as model inputs, taking whether the policy is insured or not as the dependent variable, using SVM modeling and choosing the RBF kernel function, the evaluation results obtained are as table 4:

Table 4. Assessment Results

Accuracy	73.28%
Recall Rate	72.54%
Precision Rate	78.93%

From the table 4, it is clear that the final model obtained 73.28% accuracy, 72.54% recall and 78.93% precision on the test set, so the model is more effective.

Let 1 represent coverage for the region and 0 represent no coverage. The final partial classification results for whether or not to underwrite policies for some areas of the United States and China are as follows table 5:

Table 5. Regional Classification Results in China and U.S (Raihan, 2023).[2]

State	**Covered or not**	**Province**	**Covered or not**
ATLANTIC NORTH	1	Fujian	0
CALIFORNIA	0	Shaanxi	1
DELAWARE	0	Jiangxi	0

GEORGIA	0	Hunan	0
HAWAII	1	Guangdong	0
MAINE	0	Beijing	1
FLORIDA	1	Jiangsu	1
ILLINOIS	0	Yunnan	0
INDIANA	0	Chongqing	0
TENNESSEE	1	Shandong	0
TEXAS	0	Gansu	0
KANSAS	0	Heilongjiang	1
WISCONSIN	1		

4 Conclusions

By establishing risk assessment models, insurance companies can evaluate the risk level in specific regions, leading to more precise risk pricing and management. This, in turn, de-creases the likelihood of losses and enhances profitability. Moreover, through the model's forecasts and analyses, insurance firms can gain a deeper understanding of market trends and risk profiles, enabling them to optimize business strategies and product design, consequently improving the efficiency and accuracy of decision-making processes. Looking at the bigger picture, the model can incentive insurance companies to introduce more competitive and market-oriented insurance products based on the predictions of the risk model, ultimately boosting market share and customer satisfaction.

References

1. Newman, R., & Noy, I. (2023). The global costs of extreme weather that are attributable to climate change. Nature Communications, 14(1), 6103.
2. Raihan, A. (2023). A review of the global climate change impacts, adaptation strategies, and mitigation options in the socio-economic and environmental sectors. Journal of Environmental Science and Economics, 2(3), 36-58.
3. Geng, T., Jia, F., Cai, W., Wu, L., Gan, B., Jing, Z., ... & McPhaden, M. J. (2023). Increased occurrences of consecutive La Niña events under global warming. Nature, 619(7971), 774-781.
4. Zhong, L., Nie, J., Yue, X., & Jin, M. (2023). Optimal design of agricultural insurance subsidies under the risk of extreme weather. International Journal of Production Economics, 263, 108920.
5. Kaur, G. (2023). Environmental Risk and Sustaining the Insurance Industry. The Impact of Climate Change and Sustainability Standards on the Insurance Market, 429-438.
6. Zhou, X., Wang, Y., Zhang, Y., & Liu, F. (2023). Macro investigation on China's engineering insurance industry: based on industrial organization theories. Engineering, Construction and Architectural Management.
7. Osariemen, O. A., & Benedicta, O. (2023). Underwriting Capacity and Financial Performance of Listed Insurance Firms in Nigeria. Strategic Management Practices & Sustainable Development in a Global Economy, 191.

8. Ali Abd Al-Hameed, K. (2022). Spearman's correlation coefficient in statistical analysis. International Journal of Nonlinear Analysis and Applications, 13(1), 3249-3255.
9. Kumari, P., & Toshniwal, D. (2021). Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting. Applied Energy, 295, 117061.
10. Abdullah, D. M., & Abdulazeez, A. M. (2021). Machine learning applications based on SVM classification a review. Qubahan Academic Journal, 1(2), 81-90.