ISSN 2616-5902 Vol. 6, Issue 7: 167-174, DOI: 10.25236/AJBM.2024.060723

Multidimensional Analysis of Insurance Companies' Underwriting Choices: Case Studies of Yonghe Street and Cote de Nuits

Yu Zheng

Eurasia International School, Henan University, Zhengzhou, 475004, China

Abstract: *Under the background of globalization, natural disasters pose a challenge to the insurance industry, and an effective property underwriting strategy is urgently needed. Considering these issues, this study collected data on nine indicators related to disaster and property insurance in Turkey, France, India, Australia, and Gabon from websites such as EARTHDATE and predicted them backward by LSTM for five years. Using these data, this paper then built a Disaster Assessment model to determine whether insurance companies should choose to underwrite policies. DA model reflected the overall disaster degree from the three dimensions: population, construction, and economy. Firstly, the entropy weight method is used to calculate the weights of 9 secondary indicators. Then use the K-means clustering algorithm to divide the three-dimensional clustering into 8 boxes. Insurance companies will not underwrite policies in the hard-hit areas. Eventually, apply the DA model to Yonghe Street and Cote de Nuits.*

Keywords: *LSTM, DA Model, Entropy Weight Method, K-means Clustering Algorithm, Three-dimensional Cluster*

1. Introduction

With the growth of global industrialization, climate change has been dramatic, and extreme weather events have increased. Extreme weather is a meteorological event that causes significant damage beyond the normal range of regional climate, such as hurricanes, heat waves, heavy rains, and droughts. In areas where extreme events are severe, insurance companies are reluctant to underwrite policies in these areas because of the huge payouts. According to data, the frequency of hurricanes, wildfires, floods, and other disasters will skyrocket by 2040. The surge in losses from extreme weather events will result in even more dramatic increases in insurance costs. Therefore, it is essential to establish reasonable financial insurance models to ensure the long-term stable development of insurance companies.

In this paper, LSTM, entropy weight method, and K-means Clustering Algorithm are used to determine whether insurance companies should undertake policies considering disaster risk. In the selection of disaster assessment indicators, Zhang Peng et al. [1] used population, housing, economy, and other related indicators. Concerning the disaster assessment system of insurance companies, this study also adopted these three first-level indicators, which correspond to nine second-level indicators: the number of injured, the number of deaths, the number of emergency shelters, the number of houses collapsed, the number of houses damaged, the number of houses rebuilt, direct economic losses, production losses, and environmental losses. Liu Yihang et al. [2] used LSTM to predict the severity of traffic accidents in time and space. Therefore, we used the LSTM model to predict the data of 9 indicators related to disaster and property insurance backward. Zhang Xing et al. [3] used the entropy weight method to determine the weights in the comprehensive assessment of agrometeorological disasters, so we used EMW to determine the weights of nine secondary indicators. Xia Guangwei et al. [4] used the K-means Clustering Algorithm to classify the risk of financial customers, so we used this method to classify the degree of disaster.

This study first forecasts the data of 9 insurance-related indicators of disasters in five countries backward, and then calculates the weights of 9 secondary indicators to divide disaster risks into 9 categories, and insurance companies will not cover high-risk areas whose scores are in the top right corner of the disaster box. Finally, Yonghe Street and Cote de Nuits are selected to be applied into the model.

ISSN 2616-5902 Vol. 6, Issue 7: 167-174, DOI: 10.25236/AJBM.2024.060723

2. Disaster Risk Prediction Using LSTM

An increase in extreme weather events and natural disasters can have a dramatic impact on the insurance system. Many catastrophes disrupt people's lives and property, causing severe socio-economic losses, including housing, infrastructure, crops and casualties. Insurance companies play an important role in these events, but also bear a huge risk. For insurance companies and insurance systems, scientific forecasting of extreme weather and natural disasters is essential for assessing risks and formulating policies.

This study uses Long Short-Term Memory Network to forecast the relevant disaster indicator data for the later five years of the collected data. First of all, collect global disaster data, and then use SPSS and Python for data preprocessing. Then use the LSTM model to forecast the disaster data and finally show our prediction results.

2.1 Data Collection

The data is collected from the following websites, as shown in Table 1:

Table 1: Data collection sources

Data Set Names	Web Addresses
EARTHDATE	https://www.earthdata.nasa.gov/
International Disaster Database	https://www.emdat.be/
Our World Date	https://ourworldindata.org/
National Risk Index	https://hazards.fema.gov/nri/
World Bank	https://www.shihang.org/
European Centre for Medium-Range Weather Forecasts	https://www.ecmwf.int/

This study picked five countries with varying degrees of disaster: Turkey, France, India, Australia and Gabon. Nine disaster-related indicators were then collected for these countries: the number of injured, the number of deaths, the number of emergency shelters, the number of houses collapsed, the number of houses damaged, the number of houses rebuilt, direct economic losses, production losses, and environmental losses.

2.2 Data Pre-processing

Data cleansing: Since many websites had data missing or abnormal for a certain number of years. We used SPSS and Python to clean the data and remove the outliers in the data. These outliers deviate from the mean by more than twice the standard deviation, so we discard these data.

Data Filling: For missing value data, we solve this problem in the following way.

- If an indicator for a country has few missing values and a weak correlation with the year indicator, we use the mean interpolation method to complete.

- If there is rather little missing data for an indicator in a country and there is a strong correlation with the year indicator, we use the regression interpolation method to complete.

- If a country is missing all data for an indicator, we complete using the mean of all other countries. And regression interpolation method to fill in the missing values.

2.3 The Process of LSTM

Long Short Term Memory Network [5] is a deep learning model that is commonly used to predict time series data and build models. We collected data for nine indicators in five countries because global climate change is dramatic and unusual, and predicting natural factors such as weather is not a priority for insurers. Therefore, we choose nine indicators that have strong correlation with natural disaster risk in the field of property insurance for prediction.

The basic principle of LSTM is to design a unique memory unit structure, and its working process is as follows:

Step1: The forgetting gate determines which information can pass through the cell state through the output of the previous moment, and the algorithm formula is as follows:

ISSN 2616-5902 Vol. 6, Issue 7: 167-174, DOI: 10.25236/AJBM.2024.060723

$$f_t = \sigma(W_t \bullet [h_{t-1}, x_t] + b_f) \tag{1}$$

Where h_{t-1} represents the hidden state of the last moment, x_t represents the input characteristic data, f_t is the activation value of the forgetting gate, W_f is the weight of the input gate, b_f is the bias of the input gate, and σ is the activation function of the input gate.

Step2: Input can control the new input information and update the memory cell through the sig, and then add new candidate values through the tanh layer. These two steps work together to eliminate useless data and retain new data. The algorithm formula is as follows:

$$C_t = f_t \times C_{t-1} + i_t \times C_t \tag{2}$$

i_t denotes the input gate activation value, C_t is the memory cell activation value, C_{t-1} is the memory cell activation value at the previous time, and is the activation value of the candidate memory cell.

tanh is the activation function of the candidate memory unit, and its value ranges from - 1 to 1. When the input is less than zero, the output approaches - 1. When the input is greater than zero, the output approaches 1. When the input is close to 0 and the output is close to 0, the tanh function can constrain the input data between - 1 and 1, and maintain nonlinear characteristics, which can enhance the model expression.

Step3: The output data is first passed through sig, then the tanh layer is used to constrain the data between - 1 and 1, and the final output of the model is obtained by multiplying with s. The algorithm formula is as follows:

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{3}$$

$$h_t = o_t \times tanh\left(C_t\right) \tag{4}$$

Where o_t is the output gate activation value, b_o is the output gate bias, h_t is the hidden state of the current time step, and C_t is the memory unit of the current time step.

2.4 Prediction Results

Plug the data into the LSTM model to get the prediction result of 9 indicators from 5 countries. Here present India's emergency relocation of population, as shown in Figure 1.

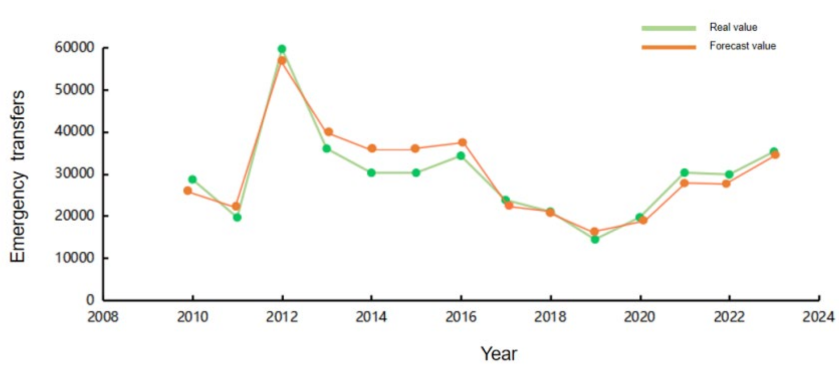

Figure 1: Emergency transposition population projections for India

As we can see, emergency disaster-related migration in India reached its maximum in 2012, mainly due to the large number of people displaced by the 2012 floods in India. Towards 2020, the number of

ISSN 2616-5902 Vol. 6, Issue 7: 167-174, DOI: 10.25236/AJBM.2024.060723

emergency migrants is on a downward trend, mainly due to the significant reduction in natural disasters during the year, the strengthening of measures to deal with the risk of natural disasters in India, and the economic growth is also an important factor. After 2020, it can be seen that the LSTM prediction curve of the data has a high degree of fit with the displayed data, and the prediction accuracy is better.

3. Identification of High-Risk Areas and Insurance Underwriting Decisions

This study will build DA Model to make decisions about insurance companies' coverage in areas that are catastrophically affected by extreme weather events.

First, DA Model will be divided into three dimensions, namely population, buildings, and economy. The nine indexes corresponding to them are normalized, and the indexes under the unified dimension are obtained. Then use the entropy weight method to analyze the weights of these indicators, and finally use the K-means method to get the final disaster risk classification.

3.1 Index System

After summarizing the different disaster assessment index systems all over the world, the disaster index that Zhang Peng et al. [1] used is more suitable for the assessment of disaster level. The new disaster assessment index is divided into three dimensions: population (P), housing (H), and economy (E).

Population: There is a strong correlation between population and disaster index. A catastrophe in a densely populated area will cause more casualties and missing, and the population is not easy to transfer in time, so the number of affected people increases and the disaster index increases. Therefore, we divide the population dimension into disaster-affected population, dead and missing population and emergency relocation population.

Buildings: The link between buildings and the disaster index is mainly reflected in the extent to which houses are affected. Disasters can cause the collapse and damage of buildings, thus directly affecting the population and economy, resulting in higher disaster conditions. Therefore, it is an important dimension to measure the disaster index. We divide this dimension into collapsed buildings, damaged buildings, and dangerous buildings.

Economy: The relationship between the economy and the risk index is strong. Natural disasters have a great impact on the economy, and the economy of a region also has a great impact on the extent of the disaster and the ability to recover after the disaster. So this study uses direct economic losses, economic loss of production and environmental economic losses to represent this dimension.

The three dimensions include many related indicators. After modifying the indicators, the secondary index is adjusted to better suit insurance companies' risk assessments and the new indicator system is created. In the DA model, three dimensions and 9 indicators are considered, as shown in Table 2.

Table 2: The disaster assessment index

Level 1	Level 2
Population	Disaster-affected population (P1)
	Dead missing persons (P2)
	Emergency relocation population (P3)
Buildings	Collapsed buildings (B1)
	Damaged building (B2)
	Reconstruction of buildings (B3)
Economy	Direct economic loss (E1)
	Economic loss of production (E2)
	Environmental economic losses (E3)

3.2 Entropy Weight Method

Among the nine indexes obtained above, the entropy weight method [6] is used to determine the weights. Since the disaster data is a multivariate statistic, it is necessary to normalize all indicators to ensure that all indicators are calculated in a unified dimension.

(1) Normalize the indicators: For the disaster population index, the greater the difference between

ISSN 2616-5902 Vol. 6, Issue 7: 167-174, DOI: 10.25236/AJBM.2024.060723

the index values, the greater the role of the index in the comprehensive evaluation, and the smaller the vice versa. Therefore, we used Min-Max Scaling to normalize data for 9 indicators of 5 countries. If we assume that there are n indicators and m evaluation objects, we can write it as a judgment matrix: $X = \left(x_{ij}\right)_{m \times n}$, where $i = 1, 2\ldots m, j = 1, 2\ldots n$ .Then the normalized judgment matrix Y is:

$$y_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}} \tag{5}$$

Where x_{ij} is the value of the jth index of the ith evaluation object, and x_{max} and x_{min} are the maximum and minimum values of the jth index, respectively.

(2) Calculate information entropy: Calculate the proportion of the i evaluation object in the j index, which is

$$P_{ij} = \frac{y_{ij}}{\sum_{i}^{n} y_{ij}} \tag{6}$$

The information entropy e_j of the jth index is calculated.

$$e_j = -\frac{1}{\ln n}\sum_{i=1}^{n} p_{ij} \ln(p_{ij}) \tag{7}$$

Where $k = -\frac{1}{\ln n}$, and the range of information entropy is between [0, 1].

(3) Calculate the information utility value: d_j is the difference value of the information entropy of the jth index, which is

$$d_j = 1 - e_j \tag{8}$$

The larger the d_j, the smaller the information entropy of the jth index and the more information provided.

(4) Calculate the weight: Calculate the weight of the indicator according to its information utility value.

$$W_j = \frac{d_j}{\sum_{j=1}^{m} d_j} \tag{9}$$

(5) Calculate the index score: According to the weight vector of each indicator, the comprehensive score of the three dimensions can be calculated.

$$s_i = \sum_{j=1}^{m} w_j \bullet p_{ij} \tag{10}$$

The P, B, E weights calculated by EMW are shown below.

$$WP = (0.23, 0.57, 0.20)^T \tag{11}$$

$$WB = (0.63, 0.19, 0.18)^T \tag{12}$$

ISSN 2616-5902 Vol. 6, Issue 7: 167-174, DOI: 10.25236/AJBM.2024.060723

$$WE = (0.47, 0.26, 0.27)^T \tag{13}$$

The final integral weights are shown in Figure 2.

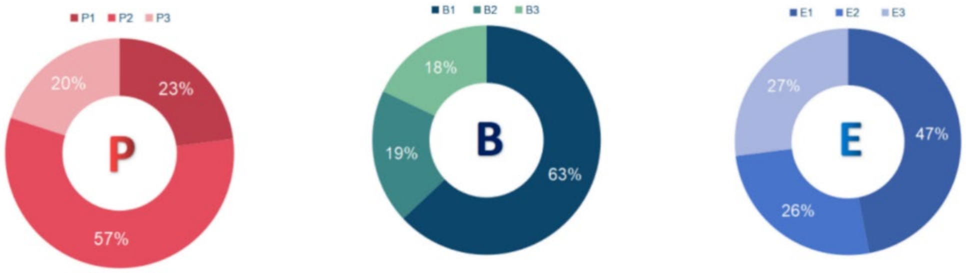

Figure 2: Weight for 3 dimensions

From Figure 2, it can be seen that when we assess catastrophe risk, P2 accounts for a large part of the population dimension, up to 57%. In other words, when a disaster occurs, the death indicator is more important to measure the disaster score of the population, while the number of injuries is very close to the number of emergency shelters. In terms of building dimensions, B1 accounts for the largest proportion of 63%. The index of building collapse is very important for buildings, and the importance of damaged houses and rebuilt houses is the same. In the economic dimension, direct economic loss is more im- important, accounting for 47%.

3.3 K-means Clustering Algorithm

After calculating the weight of each index, this study uses the K-means clustering algorithm [7] to calculate the evaluation threshold value of the obtained weight index. The threshold table is shown in Table 3:

Table 3: Three dimensions of cluster analysis results

SP	SB	SE
0-0.67	0-0.79	0-0.52
0.67- 1	0.79- 1	0.52- 1

According to the table, we construct a three-dimensional coordinate system, in which the X-axis represents the population score index, the Y-axis represents the construction score index, and the Z-axis represents the economic score index, and draw a three-dimensional cluster graph according to the three thresholds, as shown in Figure 3.

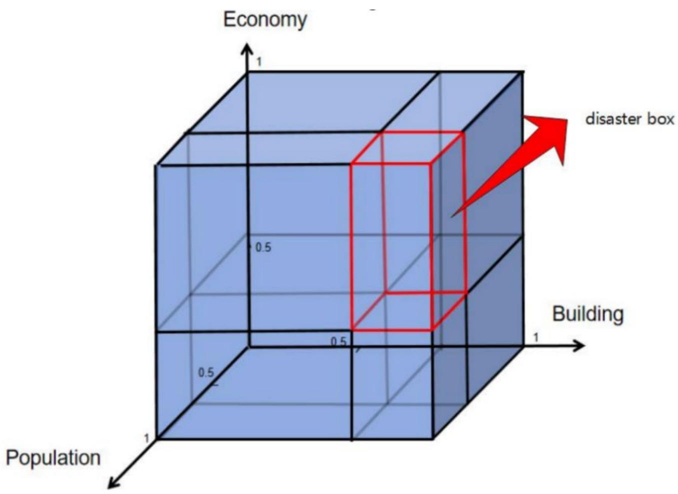

Figure 3: Three-dimensional cluster

If a region's score is located in the top right corner of the disaster box, the insurance company will not underwrite in that region. Otherwise, the insurance company will underwrite policies.

ISSN 2616-5902 Vol. 6, Issue 7: 167-174, DOI: 10.25236/AJBM.2024.060723

4. Application

To better display the model, two regions located on different continents and experiencing extreme weather events are selected to test and display the model. The locations we selected were Yongning Street in Zengcheng District, Guangzhou, China, and the Cote de Nuits wine-producing area of Burgundy, France.

China's Guangzhou city is close to the South China Sea, hot and humid, heavy rainfall, vulnerable to typhoons, rapid urbanization, and prone to extreme weather events. In 2023, the rainfall intensity in Guangzhou set a new record, which caused serious flooding and flooding in Yongning Street, Zengcheng District. While in 2022, the famous Burgundy Cote de Nuits wine region in eastern France was threatened by fires in the surrounding woodlands due to a heatwave and drought caused by extreme weather. In addition to drought, frost, hail and other disasters seriously threaten French wineries. The extreme weather events in these two areas are representative and relevant, so they are selected as typical examples to be applied into DA model.

By plugging the collected data into our DA model, the disaster assessment scores of Yongning Street, Zengcheng District, Guangzhou City, China were obtained as (0.33, 0.45, 0.21). From the results, it can be seen that the two dimensions of population and economy were not too serious in the natural disasters in this region, but the loss of buildings was slightly higher, mainly due to floods which may destroy a lot of buildings. As can be seen from the population indicators, the emergency relocation of the population is small, indicating that the protection measures in the area are relatively good. Therefore, the insurance company can still choose to underwrite policies in Yongning Street.

The disaster assessment scores for the Burgundy Night Hills region in eastern France were (0.35, 0.67, 0.55). From the perspective of population, the population is not seriously affected, mainly because the area is mostly wine-producing areas, relatively empty, and people are more convenient to move emergency devices. However, the construction and economic damage was severe. Therefore, from the evaluation score, underwriting policies in the area is not a good choice.

5. Conclusions

In all, we collected data on 9 indicators related to catastrophe and property insurance in five countries, Turkey, France, India, Australia, and Gabon, and made predictions with the LSTM model. After that, we built a DA model to identify the conditions under which insurance companies choose to undertake insurance. In this model, we reflect the overall disaster situation from the three dimensions of population, construction, and economy. We first used the Entropy Weight Method to calculate the weight of 9 secondary indicators and then used the K-means Clustering Algorithm to divide the three-dimensional cluster into 8 boxes. Insurance companies will not cover areas that fall into the disaster box. We apply the DA model to Yonghe Street and Cote de Nuits and conclude that the insurance company will choose to underwrite in Yonghe Street but not in Cote de Nuits.

In today's era of frequent heavy disasters, when a disaster occurs, the own- era's personal safety and economic property losses will affect the insurance company's claims. By using the LSTM forecast and DA model, this study has established a system for evaluating the disaster risk of a certain region which is convenient for insurance companies to make underwriting decisions. In addition, after a disaster, the owner's protective decisions on the house will also greatly affect the assessment and underwriting decisions. Therefore, the owners can also take some measures to lower the disaster risk level of the buildings, such as improving the solidity and firmness of the house and building, and enhancing the solidity of the wall.

References

[1] Zhang Peng, Zhang Yunxia, Sun Zhou, et al. Comprehensive disaster index: a quantitative evaluation method of natural disaster loss[J]. Catastrophology, 2015, 30(04):74-78.
[2] Liu Yihang, Shen Hangxian. Spatiotemporal prediction of Traffic Accident Severity based on Ensemble Learning Model [J]. Science and Technology Innovation and Application, 2024, 14(08):28-35.
[3] Zhang Xing, Zhang Chungui, Wu Juxin. Research on weight determination method in comprehensive assessment of agro-meteorological disasters[J]. Chinese Agricultural Science Bulletin, 2008(11):448-452.

ISSN 2616-5902 Vol. 6, Issue 7: 167-174, DOI: 10.25236/AJBM.2024.060723

[4] Xia Guangwei, He Qingquan, Li Yushan. Research on Financial customer risk based on K-means data analysis [J]. Journal of Insurance Vocational College, 2022, 36(05):52-57.

[5] Wang Chunjing, Liu Dan, Zhang Wenkang, et al. Based on the grade of flotation LSTM neural network prediction model [J]. Journal of chemical industry automation and instrumentation, 2024 ploidy (02):325-332.

[6] Xiao-Ting W , Ai-Hong L , Pei-Long S .Research on Regional Financial Risk Measurement and Evaluation Based on Macro Balance Sheet: Taking Shanxi Province as an Example[J].On Economic Problems, 2019.

[7] Huang Zhimin, Liang Chengdong. Grade Evaluation data Analysis based on K-means clustering algorithm [J]. Electronic Quality, 2023(12):40-44.