Research on Underwriting Decision-Making of Insurance Companies in Areas with Frequent Extreme Weather Based on Comprehensive Evaluation Method

Xinge Zhang [1,*,†], Jiaxin Teng [2,†], Yasi Zhu [3,†]

[1] College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 211106

[2] College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 211106

[3] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 211106

* Corresponding Author Email: 154195@nuaa.edu.cn

† These authors contributed equally.

Abstract. The purpose of this research is to develop an insurance company underwriting model by pre-processing and model construction of collected data in order to assess whether to underwrite insurance in extreme weather-prone areas and to analyse the factors that influence underwriting decisions. In this paper, data such as insurance company payout rate, number of insured, homeowners' income in selected areas, health insurance coverage, year of construction of historical landmarks, GDP per capita, and tourism income were processed, and minimum-maximum normalisation and zero-mean normalisation were used to eliminate differences in feature magnitude. In the model construction, the logistic regression model and BP neural network were used to derive the underwriting probability, and the three main factors of natural disaster risk assessment, payout rate, and homeowner influence were introduced, and the relationship between these factors and the underwriting probability was described by the logistic distribution function, which determined that the underwriting probability should be underwritten when the underwriting probability $P_i \geq 0.5$. The natural disaster risk assessment model selects disaster factors, disaster environment and disaster receptors as assessment indicators, constructs the CEDP model, and calculates the weights using the AHP model, the Topsis model and the CPITIC weighting method, and carries out fuzzy clustering analysis on the CEDP index of 25 regions to classify the risk into five levels. For insurer profits, solvency, premium income and claims ratios were analysed, ratios based on annual claims to total premiums were calculated, and the impact of homeowners' incomes, property and historic buildings on underwriting decisions was assessed. Insurance company underwriting decision factors in areas prone to extreme weather are ultimately determined.

Keywords: Comprehensive evaluation methods, extreme weather-prone areas, insurance company underwriting decisions.

1. Introduction

In recent years, increasingly frequent climate change has led to a growing number of extreme weather events including hurricanes, droughts, floods, etc., which have resulted in increased economic and insurance losses [1]. This phenomenon, coupled with the unpredictable nature of climate change, poses a significant challenge to the insurance industry. Due to the "non-linear" and "non-strictly cyclical" nature of climate change and extreme weather events, as well as the susceptibility of extreme weather to human influence, insurance companies do not have the means to accurately predict the frequency and severity of future catastrophes, thus causing serious losses to insurance companies underwriting catastrophe insurance [2]. By taking into account the influence of extreme weather, this paper enables insurance companies to precisely assess the frequency and intensity of future disasters and effectively minimize their significant economic losses in the realm of disaster insurance. In light of the continuous escalation of property insurance premiums and the increasingly prominent insurance coverage gap, this action is particularly crucial in providing robust support for insurance companies to optimize insurance models and formulate rational strategies.

2. The basic fundamental of BP neural network

2.1. The structure of BP neural network

BP neural network is a multi-layer network with error reverse propagation, which is composed of acceptable level of error to the network output, or to a predetermined number of learning times. The network structure is shown in Figure 1.

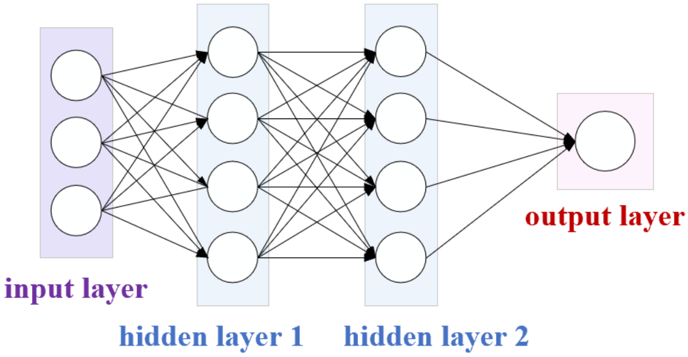

Figure 1. Neural network structure

The general model of artificial neural network consists of four basic elements, which are:

(1) The BP neural network is linked by different node coefficients. When connecting weights and weights are positive, it indicates that the current link is an exciting state. Conversely, if the link coefficient is negative, the link state is a state of suppression.

(2) The input signal and the linear signal are the combination of the signals for each input signal.

(3) The function of the nonlinear activation function: making the neuron output signal within a certain range.

BP neural network is back propagating, mainly composed of three parts: input layer, middle layer and output layer. The number of nodes in the input and output layers is relatively easy to determine, but the determination of the number of nodes in the hidden layer is a very important and complex problem[3].

2.2. The determination of the number of network layers

By analyzing the data collected in this paper, we have identified a number of problems and provide appropriate treatments. To solve the problem, we collect data such as insurance company payout rate and number of times insured, property owner income of the selected area, health insurance coverage, the year of construction of the area's historical landmarks, the area's per capita GDP and total tourism revenue, and soon. We first perform data cleaning, for missing values we mean-fill them and smooth the noisy data. Due to the different types of data collected, the magnitude of different data is different and the needs of different models, we use the min-max normalization for the data that need to be mapped between [0, 1], the formula is shown below.

$$x^* = \frac{x - \min x}{\max x - \min x} \tag{1}$$

In this paper, zero-mean normalization is also used to normalize data where the effect of magnitude between features needs to be eliminated and where regularization can be easily used. The formula is shown below.

$$x^* = \frac{x - \bar{x}}{\sigma} \tag{2}$$

We use logistics regression model and BP neural networks to derive the probability of whether they should underwrite or not, and introduce the three factors that influence insurance company underwriting: natural disaster risk evaluation, payout rate, and property owner impact.

Introduce the discrete variable y, when $y_i = 1$, insurance companies should underwrite; when $y_i = 0$, insurance companies shouldn't underwrite. Describe the relationship between $P(y_i = 1|x_i)$ and $\mathrm{x_i}$ in terms of the logistics distribution function. The formula is shown below.

$$P(y_i = 1|x_i) = \frac{1}{1+exp(-\beta_0-\beta_1x_1-\beta_2x_2-\beta_3x_3)} \tag{3}$$

Where, x_1, x_2 and x_3 are respectively the natural disaster risk evaluation factor, payout rate factor and property owner impact factor. The formula can be transformed into:

$$\frac{P_i}{1-P_i} = exp(\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3) \tag{4}$$

Taking the logarithm of the left and right sides yields the following equation [4].

$$ln\frac{P_i}{1-P_i} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 \tag{5}$$

In the above model, P_i is in the interval [0, 1]. The insurance company should underwrite when P_i≥0.5; when P_i<0.5, the insurance company should not underwrite.

In the underwriting process of insurance companies, the data of some factors cannot be obtained simply by numerical values, so they need to be processed in a certain way. This paper quantifies the factors based on the insurance company's internal underwriting manual. Meanwhile, the following three factors are mainly adopted.

2.2.1 Natural disaster risk evaluation Model

Figure 2. Indicator map for natural disaster risk evaluation

In This paper, a natural disaster risk evaluation model is designed based on the natural disaster risk evaluation index system and the corresponding weights, taking into account the risk of disaster causing factors, the sensitivity of the environment, the exposure of the disaster receptor [5-7], which can be seen in Figure 2. With the standardization of data due to the diferent scales of the indexes, and then the CEDP model is constructed by using the AHP model, the Topsis model, and the CPITIC weighting method. The formula is shown below.

$$\omega_{j,CEDP} = \frac{\sqrt{\omega_{AHP}\cdot\omega_{TOPSIS}\cdot\omega_{CPITIC}}}{\sum_{j=1}^{n}\sqrt{\omega_{AHP}\cdot\omega_{TOPSIS}\cdot\omega_{CPITIC}}} \tag{6}$$

Where j is C, E, D and P.C, Eand D separately indicate disaster causing factor conceived disaster environment and disaster receptor. We construct the formulae to calculate C, E, D and P.

$$\begin{cases} C = \omega_{C_1} * C_1 + \omega_{C_2} * C_2 + \omega_{C_3} * C_3 \\ E = \omega_{E_1} * E_1 + \omega_{E_2} * E_2 + \omega_{E_3} * E_3 \\ D = \omega_{D_1} * D_1 + \omega_{D_2} * D_2 + \omega_{D_3} * D_3 \end{cases} \tag{7}$$

Where C_1, C_2, C_3, E_1, E_2, E_3, D_1, D_2, D_3 are obtained by normalizing data in the previous para- graphs.

Accordingly, this paper reasonably and extremely creatively constructs the formula for calculating CEDP index [8].

$$CEDP = \omega_C C + \omega_E E + \omega_D D \tag{8}$$

We can get data from this URL https://www.gddat.cn/newGlobalWeb/#/home. The data from the 25 districts in the above paper were put into the CEDP model and the CEDP index was calculated for each location. Then, using fuzzy cluster analysis, all locations were categorized into five groups: low, relatively low, moderate, relatively high and high risk of natural disasters. The specific categorization is shown in the table 1.

Table 1. Natural disaster risk rating

Group	I	II	III	IV	V
Risk of natural disasters	Low	Relatively low	Moderate	Relatively high	High
CEDP	<0.2	0.2~0.4	0.4~0.6	0.6~0.8	>0.8

2.2 2 Insurance company profit

•Insurance company solvency

Solvency is the ability of an insurance company to meet its claims obligations to policy-holders and is assessed based on the insurance company's capital level, surplus position and market environment. The assessment of solvency is a dynamic process that needs to be carried out on a regular basis, taking into account changes in the market.

•Premium revenue

Premiums also known as insurance premiums, are payments made by the policyholder to the insurer as agreed in the insurance contract in order to obtain insurance coverage. Premiums are the main source of income for insurance companies. We have selected average individual premiums, market share, customer retention and customer growth to calculate premium revenue.

•Insurance Company Payout Rate

The payout rate is the percentage of claims expense to premium revenue for a given accounting period. A high payout rate for an insurance policy means that the cost of claims is driven up, and the insurance company can easily lose money by continuing to sell this policy. So basically, products with high payout rate will not be sold for very long, and an obviously high payout rate product is a customer acquisition product. Claims rate is too low, for the insurance company, it is not easy to lose money, which is conducive to the product stable business development. However, for the customer, the claim money becomes less, the willingness to buy insurance will become lower. The formula is shown below [9-10].

$$R = \frac{\sum_{i=1}^{N} X_i / N}{Y} \tag{9}$$

Where *R* is the payout rate, N is the number of years the insured has provided claims data, e.g., if I=5, the insured has provided claims for five years; Xi is the amount of claims paid in year I; and Y is the total premium for the year.

Using the same method as in Equation, the ICP is derived, I_1, I_2 and I_3 are the standardized version of the above three indicators. The formula is shown below.

$$ICP = \omega_1 I_1 \cdot \omega_2 I_2 \cdot (1 - \omega_3 I_3) \tag{10}$$

2.2.3 Property owner impact

•Level of property owner risk treatment

The level of property owner risk treatment refers to the property owner's response to the risk before the risk arrives, the addition of security doors and windows, the installation of fire alarms, and whether to purchase health insurance or not. Since direct assessment of cognitive ability is not possible, we use health insurance coverage in a region to assess risk perception [11].

•Level of property owner knowledge

The level of knowledge refers to the amount of knowledge and the amount of competence possessed. In the case of a particular discipline, it includes both theoretical and practical capabilities. In this paper, it refers to the knowledge of disaster prevention and mitigation and the ability of property owners to cope with natural disasters. In this paper, it is measured in terms of the education level of the property owner and the frequency of participation in disaster prevention and mitigation activities.

•Property owner Purchasing Power

Property owner purchasing power directly affects the ability to buy insurance. The higher the purchasing power, the more property owners are willing to spend to pay for risk. In this paper, purchasing power is measured using the property owner personal income, using a minimum-maximum normalized treatment of the data.

Using the same method as in Equation, the POI is derived, O_1, O_2 and O_3 are the standardized version of the above three indicators.

$$POI = \left(1 - \omega_{O_1} O_1\right) * \left(1 - \omega_{O_2} O_2\right) * \omega_{O_3} O_3 \tag{11}$$

3. Results

3.1. The establishment of simulation model

25 regions are selected, and the data of three factors are collected through the Global Natural Disaster Data Platform, the Global Natural Disaster Database and the official websites of insurance companies. And the data are normalized so that the range of values taken by the three indicators is specified between 0 and 1.

Before regression analysis, correlation test and multivariate linear covariance test are per- formed on the factors. In this paper, firstly, the Pearson correlation test (P=0.5) is carried out on many influential factors, and it is found that the Pearson correlation coefficients of the three factors before are less than 0.5, and all three factors pass the test. Secondly, the multivariate correlation existing between multiple factors is tested using the tolerance index, and the tolerance of all three factors is relatively high, so there is no serious multivariate linear relationship.

After that, the three testing indicators related to insurance formula underwriting in the training set, including natural disaster risk, insurance company profit, and property owner impact, are subjected to stepwise regression multifactor Logistic regression model construction, and the model is regressed using the reverse Wald stepwise selection method, and parameter estimation is performed. We can get the data from this website https://ourworldindata.org/natural-disasters# natural-disasters-data-explorer. The results are shown in the table 2.

Table 2. Results of the Reverse Wald Stepwise Selection Method

	coefficient	Standard deviation	Wald	Freedom	Significant level
constant	1.9404	0.2763	5.7326	1	0.025
Natural disaster risk	-2.9859	0.2387	6.5682	1	0.012
Payout Rates	-1.8062	0.0169	4.5671	1	0.034
Property owner impact	2.1119	0.2566	4.2385	1	0.031

$$ln\frac{P}{1-P} = 1.904 - 2.9859x_1 - 1.8062x_2 + 2.1119x_3 \quad (12)$$

3.2. Analysis of experimental results

According to the model, relevant data for the past 20 years are analyzed and predicted whether insurance companies should underwrite in the next few years. When the risk of natural disasters is high, the probability of underwriting is small, and the insurance company should not underwrite in the region; when the risk of natural disasters is low, the probability of underwriting is greater than 0.5, and the insurance company can underwrite in the region.

Then, we use a BP neural network to validate the accuracy of the model. The following figure shows a comparison of the logistics model and BP neural network fitted values [12-13].

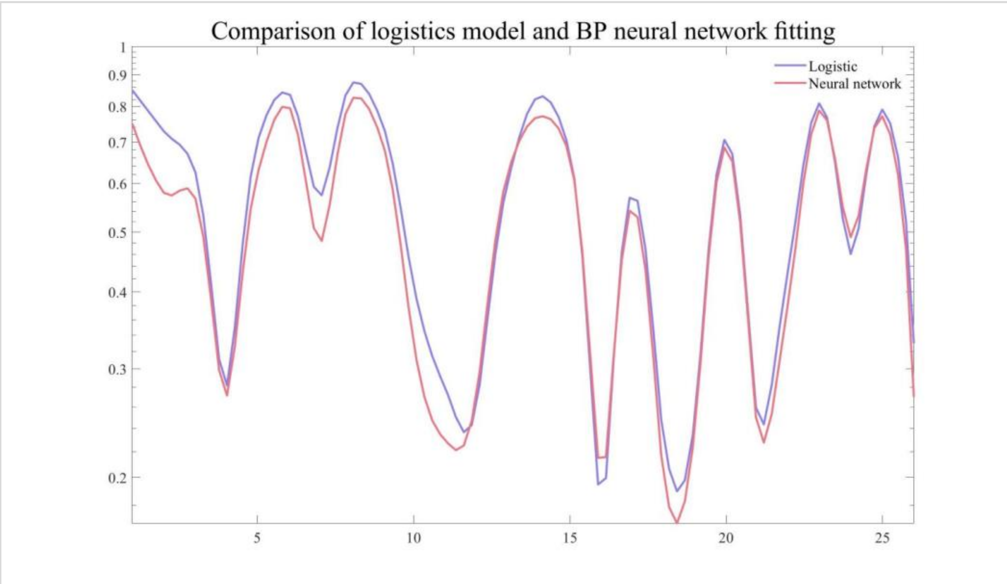

Figure 3. Comparison of logistics model and BP neural network fitting

From the Figure 3, We can get the conclusion that the accuracy of the model is 96.15% and the accuracy and reliability of the model is well.

4. Conclusion

The study presents a comprehensive Natural Disaster Risk Evaluation model that not only assesses risk based on traditional disaster-causing factors, environmental sensitivity, and disaster exposure but also integrates financial and socio-economic dimensions such as insurance company solvency, premium revenues, and payout rates. This holistic approach allows for a more nuanced classification of areas into different risk groups, which is crucial for informed decision-making in insurance underwriting.

The incorporation of a wide array of factors into the model enhances its feasibility in real-world applications. By considering not just the physical risk of disasters but also the economic and social contexts, the model provides a robust framework for insurers to evaluate risks more accurately. The use of data from 25 regions for simulation and regression analyses further validates the model's applicability across diverse geographical and socio-economic settings. This multi-dimensional approach ensures that the model can adapt to varying conditions and provide tailored risk assessments, making it a valuable tool for the insurance industry.

Looking ahead, the model's potential for growth is substantial. Future enhancements might encompass the incorporation of real-time data streams and advanced machine learning algorithms to refine risk prediction accuracy. This could enable more dynamic and responsive risk assessments. Moreover, the model's analytical insights could be instrumental for policymakers and urban planners in bolstering disaster preparedness and resilience strategies. By expanding the model's global footprint, it could catalyze international collaboration in disaster risk management, fostering the development of universally applicable assessment protocols. Finally, by factoring in socio-economic variables, the model can drive the creation of more inclusive insurance products, thereby advancing financial inclusion and strengthening community resilience in the face of natural disasters.

In summary, the Natural Disaster Risk Evaluation model is a forward-thinking approach to risk management in the insurance sector. Its comprehensive nature and potential for future development make it a promising tool for enhancing both industry practices and societal resilience against natural disasters.

References

[1] NiiArmah A O. Ratemaking in a changing environment [J]. ASTIN Bulletin, 2023, 53 (3): 596-618.

[2] Tian Zhenhua, WU Xuemin, QIU Jian. Problems and countermeasures of agricultural insurance operation under extreme weather [J]. The Banker, 2022, (02): 110-112+7.

[3] Zhang Yingbin, Xu Peiyi, Lin Jianfeng, et al. Based on BP neural network is the jiuzhaigou area of earthquake landslide hazard prediction research [J]. Journal of engineering geology, 2024, 32 (01): 133-145.

[4] Xing Zhen. Empirical Research on the Influencing Factors of insurance underwriting of multinational enterprises [D]. Zhejiang Gongshang University, 2014.

[5] Qijun Jiang. Research on flood disaster risk zoning and insurance protection countermeasures in the Yangtze River Basin [D]. Wuhan: Huazhong Normal University, 2023.

[6] Lu Jingbo. Research on safety risk assessment and prevention and control of Hunan Embroidery Intangible Cultural Heritage under the background of cultural and tourism integration [D]. Xiangtan university, 2020.

[7] Yating Yang. Research on evaluation system and optimization strategy of urban resilient com- munity under the perspective of earthquake resistance and disaster prevention [D]. Beijing University of Technology, 2017.

[8] Ye X, Zeng Q, Sun R, et al. Torque optimization strategy for outer rotor permanent magnet brushless DC motor based on metamodel and CRITIC–TOPSIS method [J]. Electrical Engineering, 2024, (prepubulish): 1-16.

[9] Hu Y. Research on vehicle risk analysis and insurance pricing model based on multi-source data [D]. Beijing: central university of finance and economics, 2017.

[10] C.DBC, Katrien A. Insurance pricing with hierarchically structured data an illustration with a workers' compensation insurance portfolio [J]. Scandinavian Actuarial Journal, 2023, 2023 (9): 853-884.

[11] V.SS, N.VG, V.EL, etal. Current Geomorphology: Natural Risk Assessment and Environmental and Anthropogenic Interaction [J]. Herald of the Russian Academy of Sciences, 2022, 92 (3): 361-369.

[12] Environmental Research; Studies from Sichuan University Yield New Data on Environmental Research (Bayesian Network of Risk Assessment for a Super-large Dam Exposed to Multiple Natural Risk Sources) [J]. Ecology Environment & Conservation, 2019.

[13] SHI Biao, LI Yu Xia, YU Xhua, YAN Wang. Short-term load forecasting based on modified particle swarm optimizer and fuzzy neural network model [J]. Systems Engineering-Theory and Practice, 2010, 30 (1): 158-160.