Journal of Computers Vol. 36 No. 2, April 2025, pp. 327-333
https://doi.org/10.63367/199115992025043602022

A BPNN Model Based on Genetic Algorithms for Predicting Policyholders' Midway Surrender

Yi-Ching Hsieh[1], Hsiao-Hui Ho[2], Fang-Kun Kuo[3*], Chao-Chun Shen[1], Shun-Nian Luo[4], and Lipeng Cheng[4]

[1] School of Art Design and Media, Sanda University, Shanghai, China

[2] School of Business, Sanda University, Shanghai, China

[3] School of Management, Sanda University, Shanghai, China
fkguo@sandau.edu.cn

[4] School of Information Science and Technology, Sanda University, Shanghai, China

Received 25 September 2024; Revised 1 October 2024; Accepted 1 October 2024

Abstract. Due to the booming development of the internet, people can easily purchase a life insurance product on their mobile devices. However, it is easy for people to overlook whether their own situation is suitable for their long-term payment ability. Midway surrender will inevitably damage the policyholder's economic and credit situation. For insurance companies, predicting the credit of each policyholder is an important step in risk management to avoid encountering the worst-case scenario of policyholders withdrawing their insurance midway. This study applied a genetic algorithm-based BP neural network to build a model for predicting policyholder behavior. It integrates the optimal implementation results into a policyholder behavior prediction system. This system could be utilized by insurance company employees for internal purposes and to assist users in making decisions on whether to underwrite while ensuring data security. This research can improve insurance companies' risk management, promote data security, and serve as a case study for the academic community on the application of genetic algorithms and neural networks in the insurance industry. The contribution of this research is beneficial to assist insurance companies in better risk management, promote data security, and serve as a case study for the academic community on the applications of the genetic algorithms and the neural networks.

Keywords: BP neural network, prediction, genetic algorithm, midway surrender

1 Introduction

With the increasing level of social aging in various countries and the increasing interest in healthy lifestyles, the life insurance industry has great development prospects. Taking China as an example, it is estimated that the number of people aged 65 and above in China is expected to exceed 300 million by 2035 [1, 2]. At the 2022 Work Conference of China Life Insurance, its leader Bai Tao proposed the following: "Continuously promote product service upgrading, improve product development iteration mechanism, optimize customer service investment mechanism, improve consumer rights protection mechanism, and build a high-quality product service supply system are centered on customers and market-oriented" [3]. Meanwhile, it is necessary to optimize customer experience and make life insurance products a guarantee for the interests of the people, enterprises, and society [4].

Owing to the burgeoning expansion of the life insurance industry and the remarkable progress of the internet, an increasing number of individuals can readily assume the role of "policyholders" via mobile devices, such as smartphones. However, this carries inherent risks [5, 7]. Policyholders often find analyzing their own assets and credit situation challenging, especially when considering insurance terms extending up to ten or twenty years. During this period, they could encounter various situations that prevent them from completing renewal or even withdrawing the insurance and causing losses to their own assets [6]. For insurance companies aiming to broaden

* Corresponding Author

their business endeavors, relaxation of policyholder scrutiny has the potential to not only harm their own reputation, but also put themselves at risk of more policyholders maliciously filing false insurance claims [8].

The significance of this study is to establish a supplementary decision support system designed to assist insurance companies in predicting the financial credibility of policyholders. and determine the risk that policyholders may not be able to fully renew their insurance due to a range of potential factors throughout the insurance period. Customer credit risk assessment is a vital component of the financial operations of commercial enterprises.

2 Theory and Technology

The schematic diagram of this project's BP neural network with one output in four layers is shown as Fig. 1. The construction and training of BP neural network models [9] comprise two distinct steps. The first step is forward propagation, which follows the information transmission process of a basic neural network and weights the eigenvalues of the samples. The second step is back propagation. In my opinion, according to the neural network model, self-learning is carried out in the opposite direction of the gradient to adjust parameters such as weights and bias values.

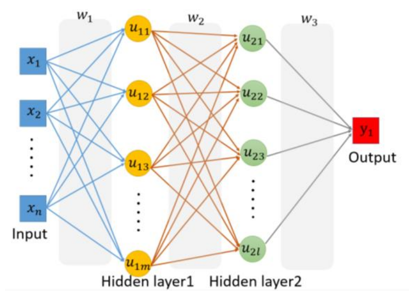

Fig. 1. Schematic diagram of a BP neural network with one output in four layers

Genetic algorithms [10] use a process of meta heuristic natural selection, and belong to the general category of evolutionary algorithms. Genetic algorithms typically utilize biological heuristic operators such as mutation, crossover, and selection to generate high-quality optimization and search solutions for problems. There are six steps of Genetic Algorithm. Step1: Initialize the parameters of the genetic algorithm, shown as Fig. 2. Step 2: Calculate fitness. Step 3: Select the replicating chromosome. Step 4: Intersection shown as Fig. 3. Step 5: Mutation. shown as Fig. 4. Step 6: End of judgement

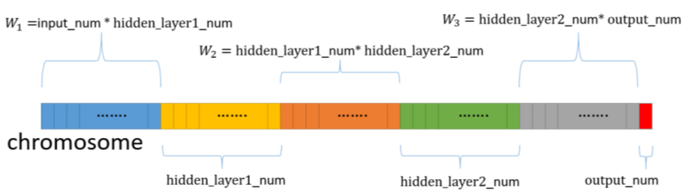

Fig. 2. Schematic diagram of the gene and chromosomes formed by the weight parameters

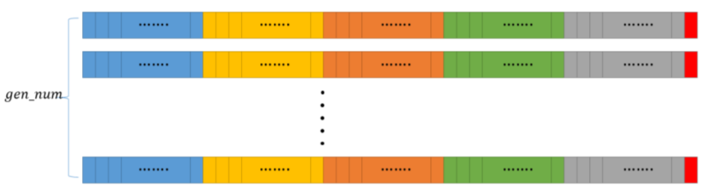

Fig. 3. The gene and chromosomes of the BPNN

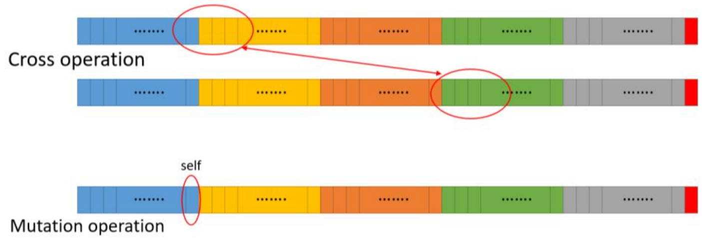

Fig. 4. Exchange chromosomal or genetic mutations

3 Model Establishment

The policyholder behavior prediction model in this investigation was constructed with a BP neural network model as its core, and it was realized through the utilization of the TensorFlow [11] and Keras [12] frameworks. Shown as Fig. 5, the implemented model incorporates a policyholder behavior prediction function, which could find utility in the subsequent scenarios: staff within the insurance company responsible for overseeing customer insurance transactions, managing customer's personal information, and predicting whether customers will choose to renew their policies in full or withdraw mid-term. Through an examination of the application problem, it can be determined that the model should address a binary classification challenge. It should take customer information characteristics as input samples and produce outcomes consistent with binary classification scenarios to facilitate predictive capabilities and support insurance company personnel in decision-making.

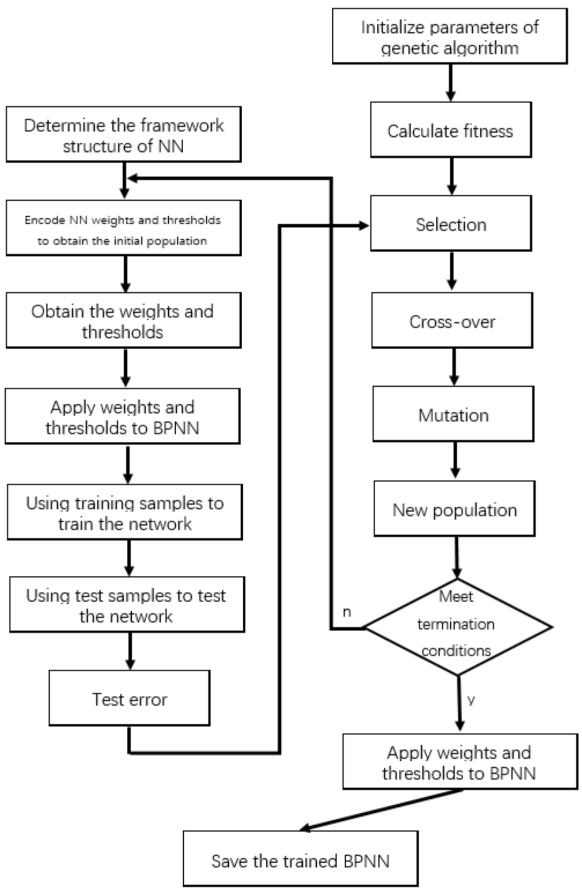

Fig. 5. The process of building the model

In determining the credit evaluation indicators for insured customers from the individual, economic, and risk levels, mentioned that "it is necessary to focus on factors that can highlight the common characteristics of insurance industry users". The assessment criteria for this model are likewise devised based on the aforementioned considerations. After conducting these analytical steps, we were able to derive 13 credit assessment criteria pertaining to the policyholder's "willingness to renew" and "ability to renew". We devised a questionnaire comprising 13 input questions and 1 output question for the sample. These encompass a total of 14 credit evaluation indicators, which are listed in Table 1.

Table 1. Questionnaire items

No.	Item	Description	Value
1	**houses**	Customer house ownership	0 to 3
2	*cars*	Customer vehicle ownership	0 to 3
3	*smoke*	Whether the customer smokes	Y or N
4	*drink*	Whether the customer drinks	Y or N
5	*chronic illness*	Whether the customer has chronic diseases	Y or N
6	*dangerous sports*	Whether the customer often participates in dangerous sports	Y or N
7	*disability*	Whether the customer is disabled	Y or N
8	*advance consumption*	Customer's advance consumption	"none", "occasionally", or "frequently"
9	*delay tel. bill*	Delay in the customer paying phone bills	"none", "occasionally", or "frequently"
10	*debt default*	Delay in customer's debt default	"none", "occasionally", or "frequently"
11	*overdue insurance*	Overdue insurance of customer during the insurance period	"none", "occasionally", or "frequently"
12	*stabilize income*	Whether the customer has a stable income	Y or N
13	*annual incomes*	Level of annual income of customer	"less than USD 100,000", "between USD 100,000 and USD 250,000" and "over USD 250,000"
14	*canceling of insurance*	Discontinued insurance payments in the past three years	Y or N

4 Experimental Results and Discussion

The experimental test was run on a Python environment to analyze the gap between the BP neural network given to the genetic algorithm and the randomly initialized BP neural network in predicting policyholder behavior. Table 2 is the example of training set samples.

Table 2. Example of training set samples

Sample	Sample characteristic values												Renewal
64	0.5	0.85	0.5	0.5	1	0.5	0.5	1	1	1	1	1	1	1
65	0.7	0.7	0.5	0.5	0.5	0.5	0.5	1	0.75	1	0.5	0.75	1	0.5
66	0.7	0.7	1	0.5	1	1	0.5	1	0.75	1	1	0.5	1	1

In the experiment conducted in this study, the number of neurons in the hidden layer was 2 and the crossover probability was p_c=0.8, with a probability of variation of p_ m=0.05, with a crossing position of 15L to 20L. The number of iterations was set to 10, 50, 100, and 200 and results were compared in Fig. 6. And the experimental results are shown as Table 3.

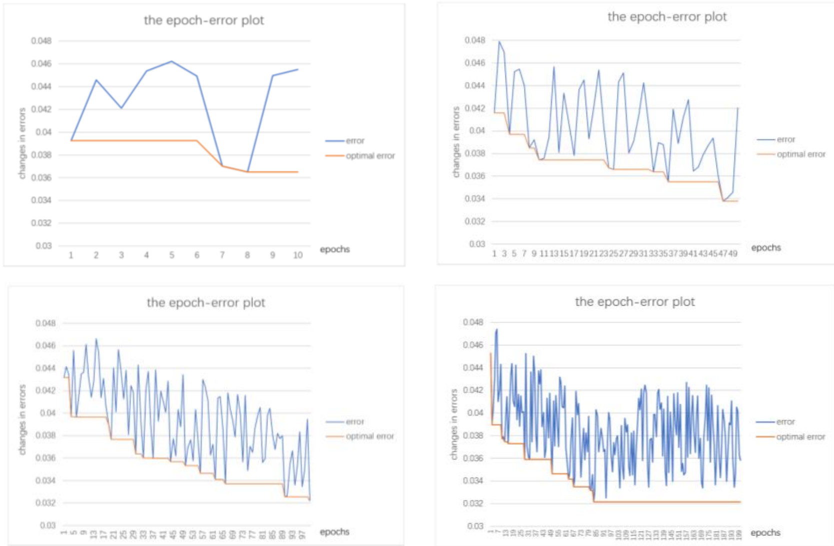

Fig. 6. The epoch–error plot after training four times

Table 3. Experimental results

Iterations	Last generation minimum error	Optimal error
10	0.045500	0.036511
50	0.042007	0.033803
100	0.032185	0.032185
200	0.035824	0.032131

From the graph, it can be seen that the best prediction effect was from the BP neural network optimized by genetic algorithms after 200 iterations. In fact, it already converged after 84 iterations. So in the end, the model in this study retained the BP neural network after 100 iterations, avoiding the possibility of overtraining.

Table 4 shows the example of comparison between test set and prediction results. Based on the above experiments, it can be concluded that under the optimization of genetic algorithms, the predicted values of the BP neural network can perform accurately and stably within a certain number of training iterations. Further optimization of genetic algorithm was carried out, with the key points being the selection, crossover, and mutation operations of the genetic algorithms. After optimization, the accuracy of prediction was also further improved.

Table 4. Example of comparison between test set and prediction results

Sample	Sample characteristic values													Predictive output	True output
341	0.7	0.5	1	0.5	0.5	0.5	0.5	1	1	1	1	1	1	0.9105	1
342	0.7	0.85	1	1	1	0.5	0.5	1	1	1	1	1	1	0.8861	1
343	0.7	1	0.5	0.5	0.5	0.5	0.5	1	1	0.5	1	0.5	1	0.6347	0.5

The final accuracy of the model reached 92%, and the predicted results have a certain auxiliary decision-making effect on users. Which is shown as Table 5.

Table 5. The accuracy of the model

Neural network	Training epoch	p_c	p_m	Accuracy
BP Neural Network Model Based on Genetic Algorithm	100	0.8	0.05	92%

5 Conclusion

This study mainly outlines the entire process of constructing a predictive model for policyholders' non-renewal behavior. Evaluation indicators for the samples are selected using the expert-system method applied to insurance policy indicators. Through a questionnaire survey, credibility and authenticity of the samples are enhanced. Sampling data is thereafter standardized using the Max-min method as part of the data preprocessing steps. In the construction of the model structure, two hidden layers are set with 24 and 16 neural nodes, respectively, and the output layer consists of a single output neural node. The activation function used the Relu function, while the output layer uses the sigmoid function. After incorporating genetic algorithms, the model fit-ting and training are completed, and an appropriate method for saving and deploying the model is established. Furthermore, myModel class interface is examined and can be called by the policyholder behavior of prediction system and eventually resulted in a final accuracy of 92%.

For the logit model which utilized the more kernelized variants as a linear classifier, in which the standard model i.e. the original linear-regression model, [13] mainly dealing with the dataset in terms of more or less manner on the linear separability. On the contrary, the neural networks only utilized one or two hidden layers, although the cost function is non-convex, which can be improved by the error backpropagation algorithm to an output of a local minimum convergence to reach the improved result. Accordingly, four different epoch-error works are validated in this work and reached a promising result. Generally, tradeoff between the model's quality of the fitting data and the predictive ability are commonly occurred during the modeling. Typically, gains of an objective generally treated as the losses of another objective, further, the large models often have a higher likelihood yet given the lower predictive quality i.e. over-fitting problem, and vice versa. Therefore, to enhance the predictive quality required more prediction research works that are the priority than the explanation of the predicted results, our observation herein in terms of the neural network is the same as the investigation discussed.

Acknowledgments

The research is conducted under the Humanities and Social Sciences Planning Fund Project of the Ministry under the title of "Research on the Social Media Communication of Mainland China's Image in Taiwan", Project No. 24YJAZH122, with special thanks.

References

[1] C.M. Zhu, Aging of Population in the Process of Modernization in China and Its Countermeasures, [dissertation] Jilin: Jilin University, 2021.

[2] X. Bai, Alignment or struggle? Exploring socio-demographic correlates of individual modernity in Chinese older people, Ageing and Society 36(1)(2016) 133–159. https://doi.org/10.1017/S0144686X14001020

[3] Y.X. Zhu, Bai has set the tone for the work of National Life Insurance, and 2022 should maintain stable growth and position, China's insurance quote, January 28, 2022 (4).

[4] Monetary Policy Analysis Group of the People's Bank of China, China Monetary Policy Report Q2 2021. (August 9, 2021). http://www.pbc.gov.cn/en/3688229/3688353/3688356/4271098/4344444/2021113016272453935.pdf

[5] S. Deng, Research on Financial Risk Assessment of Listed Accommodation and Catering Companies - Based on the Z-Scored Model, Science and Technology Economic Market/KEJI JINGJI SHICHANG 2021(7)(2021) 139–140. https://d.wanfangdata.com.cn/periodical/kjjjsc202107065

[6] H.J. Jiang, M. Li, Analysis of the financial situation of Yurun Food based on Z-score model and F-score model, Times Finance 27(2017) 170–172

[7] G.L. Xue, An empirical research on early warning model of financial risks in Lvjing Holding, Journal of Tianjin College of Commerce 8(2)(2020) 69–73. https://cstj.cqvip.com/Qikan/Article/Detail?id=7101711009

[8] Y. Bu, H. Li, X.Q. Wu, Effective regulations of FinTech innovations: the case of China, Economics of Innovation and New Technology 31(8)(2022) 751–769. https://doi.org/10.1080/10438599.2020.1868069

[9] H. Peng, Research on Credit Evaluation of Financial Enterprises Based on the Genetic Backpropagation Neural Network, Scientific Programming 2021(1)(2021) 7745920. https://doi.org/10.1155/2021/7745920

[10] H. Baek, A CNN-LSTM Stock Prediction Model Based on Genetic Algorithm Optimization, Asia-Pacific Financial Markets 31(2)(2024) 205-220. https://doi.org/10.1007/s10690-023-09412-z

[11] V. Joshi, M.L. Gallo, S. Haefeli, I. Boybat, S.R. Nandakumar, C. Piveteau, M. Dazzi, B. Rajendran, A. Sebastian, E. Eleftheriou, Accurate deep neural network inference using computational phase-change memory, nature communications 11(2020) 2473. https://doi.org/10.1038/s41467-020-16108-9
[12] A.K. Palanivinayagam, S.S. Gopal, S. Bhattacharya, N. Anumbe, E. Ibeke, C. Biamba. An Optimized Machine Learning and Big Data Approach to Crime Detection, Wireless Communications and Mobile Computing 2021(1)(2021) 5291528. https://doi.org/10.1155/2021/5291528
[13] S.R. Mohandes, X.Q. Zhang, A. Mahdiyar, A comprehensive review on the application of artificial neural networks in building energy analysis, Neurocomputing 340(2019) 55–75. https://doi.org/10.1016/j.neucom.2019.02.040