Prediction of terrorist attacks based on GA-BP neural network

To cite this article: Qinghao Li et al 2019 IOP Conf. Ser.: Mater. Sci. Eng. 490 062081

View the article online for updates and enhancements.
Prediction of terrorist attacks based on GA-BP neural network

Qinghao Li¹, Zonghua Zhang², Zhen Shen²
¹Foshan University of Foshan Guangdong, China
²Foshan University of Foshan Guangdong, China

Corresponding author e-mail: 136367198@qq.com

Abstract: Terrorism is a common threat to mankind. Combating terrorism is also the responsibility that every country should assume. In the face of the development trend of terrorism, in-depth quantitative analysis of data related to terrorist attacks will help deepen people's understanding of terrorism and provide valuable information support for counter-terrorism, and effectively improve the pertinence and efficiency of the anti-terrorism struggle. This paper designs a risk assessment and prediction system for terrorist attacks. The mathematical model calculates the relative risk index of each type of target by factor analysis. The evaluation indicators include three factors: "threat", "vulnerability" and "consequence". Mathematical model uses neural network to evaluate and predict risk index. However, because the BP neural network is easy to fall into the local best, it is difficult to jump out. Therefore, the genetic algorithm (GA) optimizes the initial weight threshold of the BP neural network to enhance the prediction accuracy. Finally, the simulation experiments of 21 main targets are carried out to verify the effectiveness of the model, so as to carry out accurate strategy analysis.

1. Introduction
A terrorist attack is an aggression committed by an extremist that is not in conformity with international morality and is directed against, but not limited to, civilians and civilian installations.¹² With the rapid development of the world economy, terrorism has also grown rapidly in line with the changes of the times. At present, many countries are facing the threat of terrorism. As a cancer of human society, terrorism not only causes incalculable casualties and property losses to human society, but also has great lethality and destructive power. Terrorism has also brought tremendous psychological pressure to people, causing social unrest to a certain extent. In addition, terrorism can cause people to panic, hinder people's normal work and life order, and thus greatly hinder economic development. This paper makes use of the principles of operations research, and also predicts and evaluates the specific problems of quantitative analysis of terrorist attacks. This demonstrates the potential of quantitative research methods in the quantitative analysis of terrorist attack records.

2. Risk assessment indicator selection
Regarding the selection of risk assessment indicators for terrorist attacks, RAND believes that it should be composed of three parts: "threat", "vulnerability" and "consequence" [³], which respectively represent the "probability of the attack", "the premise of the attack" and Expected
loss value under the premise of loss, The product of these three is the specific risk value, and in later research, the evaluation index has been widely used\cite{4-5}, indicating that it is feasible and accurate.

3. Risk assessment model

Step 1: Determine the evaluation object and obtain the data of each indicator.

The GTD database is categorized according to different targets, including 12 different regions, 220 different countries, provinces and cities of various countries, 8 types of attacks, 21 major targets, 110 subdivision targets, 12 major assault weapons and 28 subdivision attack weapons. Any attack target is selected for evaluation and data is collected based on six assessment indicators.

Step 2: Construct an initial matrix.

After the data is collected, a matrix is obtained. To eliminate the influence of different dimensions of each index, standardization processing is performed to obtain an initial matrix $X = (X_1, ..., X_p)$ and $X_p$ is a data vector about the index P.

Step 3: Construct a factor analysis model. The model is as follows:

$$X = AF + \varepsilon$$ (1)

In the equation, $F = (F_1, F_2, ..., F_m)$ is a common factor matrix and $A$ is a factor load matrix. The initial variable matrix $X$ is represented by a linear combination of $(m < p)$ common factor vectors, $F_1, F_2, ..., F_m$ and the dimensionality reduction is achieved by omitting the special factor matrix $\varepsilon$ outside the common factor. $m$ and $A$ are usually calculated using principal component analysis methods.

Step 4: Factor rotation and calculate the common factor score.

The rotation process is as follows:

$$F_j = \beta_{j1}X_1 + ... + \beta_{jp}X_p, (j = 1, ..., m)$$ (2)

In the equation, $F_j$ is the score vector of the common factor $j$, $\beta_{ij}$ represents the degree of correlation of $i$ to the common factor $j$, which is usually calculated by the orthogonal rotation maximum variance method.

Step 5: Calculate the composite score.

The calculation process is as follows:

$$Score = \sum_{j=1}^{m} W_j F_j$$ (3)

In the equation, Score is the composite score vector containing each sample, $W_j$ is the weight vector of the common factor vector $F_j$, calculated by the proportion of the variance contribution rate of the common factor $j$ to the total variance contribution rate.

Step 6: Normalize and obtain a comprehensive risk index. Normalize the Score to $[0, 1]$, as follows:

$$S'_k = (S_k - S_{\min}) / (S_{\max} - S_{\min})$$ (4)

$S'_k$ and $S_k$ are the risk values of the sample $k$ before normalization and normalization, respectively. And $S_{\max}$ and $S_{\min}$ are the maximum and minimum risk values of all samples before normalization, respectively.

4. GA-BP based risk prediction model

With regard to BP neural networks, Kosmogorov’s theorem states that a three-layer feedforward network can approximate arbitrary nonlinear continuous functions under reasonable conditions and appropriate weights\cite{6-7}, but the theorem does not give the method of reasonable structure and appropriate weight. The initial weight and threshold can be optimized by genetic algorithm, which improves the prediction accuracy of BP neural network.
First, assess the risk of attack on different targets. The evaluation index data of 7 periods of 2001-2003, 2003-2005, 2005-2007, 2007-2009, 2009-2011, 2011-2013, 2013-2015 were collected and analyzed by SPSS to obtain the risk index of 7 periods. The risk index is shown in Table 1.

| Sample number | target         | 2001-2003 | 2003-2005 | 2005-2007 | 2007-2009 | 2009-2011 | 2011-2013 | 2013-2015 |
|---------------|----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1             | business       | 0.4124    | 0.4525    | 0.2885    | 0.3203    | 0.3760    | 0.3544    | 0.2765    |
| 2             | government     | 0.2740    | 0.4361    | 0.2692    | 0.3105    | 0.3683    | 0.3354    | 0.2147    |
| 3             | Policemen      | 0.3475    | 0.7672    | 0.5742    | 0.4132    | 0.3223    | 0.6804    | 0.4618    |
| 4             | army           | 0.4407    | 0.7869    | 0.3791    | 0.2885    | 0.2327    | 0.6835    | 0.6206    |
| 5             | Hospital       | 0.1102    | 0.1803    | 0.0000    | 0.0000    | 0.1407    | 0.0000    | 0.0000    |
| 6             | Aircraft       | 0.0141    | 0.0000    | 0.0549    | 0.0954    | 0.0026    | 0.1044    | 0.1235    |
| 7             | embassy        | 0.0593    | 0.0295    | 0.0000    | 0.0000    | 0.1407    | 0.0000    | 0.0000    |
| 8             | education      | 0.0847    | 0.1934    | 0.1044    | 0.1614    | 0.1483    | 0.1551    | 0.1324    |
| 9             | food           | 0.0000    | 0.1869    | 0.1071    | 0.1222    | 0.0716    | 0.0918    | 0.0941    |
| 10            | media          | 0.0876    | 0.1607    | 0.0962    | 0.1247    | 0.1125    | 0.0949    | 0.1059    |
| 11            | Maritime       | 0.1186    | 0.2033    | 0.1071    | 0.1125    | 0.0179    | 0.0475    | 0.1206    |
| 12            | NGO            | 0.0678    | 0.1180    | 0.1126    | 0.154     | 0.1023    | 0.1076    | 0.1324    |
| 13            | other          | 0.113     | 0.2000    | 0.0852    | 0.154     | 0.1483    | 0.1044    | 0.1265    |
| 14            | civilian       | 1.0000    | 1.0000    | 1.0000    | 1.0000    | 1.0000    | 1.0000    | 1.0000    |
| 15            | religion       | 0.1977    | 0.341     | 0.2088    | 0.2249    | 0.2302    | 0.2437    | 0.1912    |
| 16            | communication  | 0.1073    | 0.0951    | 0.0742    | 0.1027    | 0.1100    | 0.1171    | 0.1147    |
| 17            | Terrorist      | 0.0819    | 0.1934    | 0.1236    | 0.1638    | 0.0921    | 0.1772    | 0.1794    |
| 18            | Tourist        | 0.0763    | 0.1639    | 0.0824    | 0.1296    | 0.1458    | 0.0886    | 0.0971    |
| 19            | traffic        | 0.3107    | 0.2689    | 0.1896    | 0.1980    | 0.1560    | 0.1551    | 0.1471    |
| 20            | public Utilities | 0.0847 | 0.1344    | 0.1099    | 0.1198    | 0.0946    | 0.1139    | 0.1206    |
| 21            | Violent party  | 0.1045    | 0.0787    | 0.0659    | 0.1443    | 0.1509    | 0.0728    | 0.1000    |
| 22            | FA             | 0.0000    | 0.0000    | 0.0000    | 0.0000    | 0.0000    | 0.0000    | 0.0000    |

The BP neural network optimized by genetic algorithm is used to test the prediction performance of the system. Compile GA-BP programs with Matlab2014 and test the predictions. As shown in Table 4, according to the principles of 70% training data and 30% test data, the forecasting index data of each sample in 2001, 2003, 2005, 2007, and 2009 was collected as the input part of the training data. The forecasted indicator data for each of the 2011 and 2013 samples was collected as an input part of the test data. The risk index of each sample in 2001-2003, 2003-2005, 2005-2007, 2007-2009, and 2009-2011 was used as the output part of the training data. The risk index of each sample in 2011-2013 and 2013-2015 was used as the output part of the test data. A total of 147 samples, 105 training samples and 42 test samples. Figure 4 to Figure 6 show the prediction results before the optimization of the genetic algorithm for the test data, the prediction results after the optimization of the genetic algorithm, and the error comparison of the samples before and after the optimization.
Figure 1. Test results of 42 test samples before genetic algorithm optimization

Figure 2. Test results of 42 test samples after optimization by genetic algorithm

Figure 3. Error comparison of samples before and after optimization

The case study shows that: From the overall indicators, using BP neural network for prediction, the mean square error and the total error of the test samples were 0.008 and 2.279, respectively. The BP neural network optimized by genetic algorithm was used for prediction. The mean square error and the total error of the test samples were 0.005 and 1.853 respectively, indicating that the prediction accuracy after optimization was greatly improved. According to Fig. 4 and Fig. 5, both prediction methods can reflect the overall distribution of the real risk index. In the prediction of the two global best advantages of the maximum risk value and the minimum risk value, the BP neural network is easy to converge to the local best advantage due to the use of the gradient descent algorithm, and the prediction bias is large; Although the optimized BP neural network still cannot reach the global best, but the prediction effect is improved.
Table 2. Comparison of forecasted overall indicators before and after optimization

| Forecast performance indicator | BP Neural Networks | GA-BP neural network | Performance optimization value |
|--------------------------------|--------------------|----------------------|--------------------------------|
| Mean square error              | 0.008              | 0.005                | 0.003                          |
| Total error                    | 2.279              | 1.853                | 0.426                          |

From the error of a single sample, it can be seen from Fig. 6 that except for sample 2 (government), the optimized prediction error is less than or equal to the prediction error before optimization, indicating that the improvement of prediction accuracy after genetic algorithm optimization is relatively balanced. In the prediction results, the police (sample 3, sample 24) and the military (sample 4) have larger prediction errors. From Table 3, it can be found that the fluctuation range is large. Therefore, it is necessary to combine other data or qualitative indicators in the prediction analysis.

References

[1] Feng Weiguo, Jia Yu, Shang Jin. The Analysis of Disagreement of Terrorism Definition[J]. Legal Forum, 2018(05): 112-118.
[2] Zeng Xianghong, Zou Jinjian. Counter-Terrorism and Recognition: Global Governance Processes and the Norm Fragmentation[J]. Contemporary Asia-Pacific Studies, 2018(04): 113-155.
[3] Willis H H. Guiding resource allocations based on terrorism risk [J]. Risk Analysis, 2007, 27(3): 597 - 606.
[4] McGill W L, Ayyub B M, Kaminskiy M. Risk analysis for critical asset protection [J]. Risk Analysis, 2007, 27(5): 1265 - 81.
[5] Chatterjee S, Abkowitz D. A methodology for modeling regional terrorism risk [J]. Risk Analysis, 2011, 31(7): 1133 - 1140.
[6] HE Y, LI X L, DENG X F. Discrimination of varieties of tea using near infrared spectroscopy by principal component analysis and BP model[J]. Journal of Food Engineering, 2007, 79 (4): 1238-1242.
[7] WANG Z L, LI Y C, SHEN R F. Correction of soil parameters in calculation of embankment settlement using a BP network back-analysis model[J]. Engineering Geology, 2007, 91 (2/3/4): 168-177.