PREDICTING THE NATURE OF TERRORIST ATTACKS IN NIGERIA USING BAYESIAN NEURAL NETWORK MODEL

Tayo P. Ogundunmade
Department of Statistics
University of Ibadan
ogundunmadetayo@yahoo.com

Adedayo A. Adepoju
Department of Statistics
University of Ibadan
pojuday@yahoo.com

Abstract

Terrorist attacks is one of the major problems Nigeria is facing presently. Seeking for techniques to understand the different factors involved in terrorism and how to deal with those factors in order to completely eradicate or reduce terrorist activities is the topmost priority of the government in any country. This study focuses on using a Bayesian Neural Network (BNN) model for predicting the nature of attacks of terrorists in Nigeria . The developed model was considered under different activation functions and training sets. The results showed that hyperbolic tangent activation function outperformed the other activation functions better in predicting the important variables in terrorist attacks in Nigeria.

Keyword: Terrorism, Weapons, Bayesian Neural Network, Activation function, Accuracy

1 Introduction

Common responses to terrorist events are constant sense of fear, feeling of helplessness/hopelessness, anger and intolerance or aggression towards certain ethnicity or religious groups. It is important to note that the emotional reactions of the affected population are understood in order to be able to effectively manage the situation so that it does not degenerate to a reprisal attacks. Terrorism has been studied for decades to understand the major factors responsible for the act or to understand how to institute anti-terrorism policies to alleviate its effects on the social and economic lives of the citizens.. However, the complex nature of the menace has made it difficult for the government to find a lasting solution to the problem. Identification of terrorist ideologies and prediction of future terrorist attacks have been proven to be of great importance and time-consuming process.

Studies which have applied machine learning and deep learning techniques to make AI-based model for terrorism are Olufemi et. al (2022), Ogundunmade and Adepoju (2021), Udin etal(2020), and Ogundunmade et. al. (2022a, 2022b). Current state-of-the-art research papers are based on understanding the pattern of terrorism and have proposed different solutions to analyzing factors of terrorism. However, research works on making predictions of future occurrences of terrorist activities as regards different factors such as suicide, weapon type, attack type, and region are not common. Clearly, there is a research gap for modeling and predicting future terrorist activities using deep learning. This paper seeks to develop a model for predicting nature of terrorist attacks in Nigeria.The rest of the paper is structured as follows: the next section contains the materials and method, while section 3 discusses the results and their interpretations, section 4 concludes the paper.

2 Materials and Method

2.1 Data for the study

The dataset used in this work is a primary data collected via an electronic medium designed in Google form (which is accessible through the link: www.https://forms.gle/gQ3idgnUT5kk8gx47) The form was designed to keep the confidentiality of the respondents by making sure that respondent's personal information cannot be traced to their responses. The distribution of the form was done randomly to different platforms in the different states in Nigeria.

2.2 Model variables for the Terrorism data The target or dependent variable is: The nature of the terrorist attack: armed assault, assassination, bombing/explosion, facility/infrastructure attack, hijacking, unarmed assault and Kidnapping.

The variables used as feature variables in the study are:

- Weapon Type: Fake Weapons, Firearms, Melee like knife, Nuclear/Explosives, Sabotage Equipment and so on.
- Number of perpetrators

- Targets/Victim type : educational institution, government, media, military, NGO, police, private business organization, private citizen and property, religious institutions, and political parties).
- State of the incidence.

2.3 Bayesian Neural Network Model

Consider random variables $y_1, \dots, y_N$ such that conditional on $p_1, \dots, p_N$, these are independent binary variables with success probabilities p_i, $i = 1, \dots, N$. We model
$\theta_i = logit(p_i)$ as

$$\theta_i = \sum_{j=1}^{M} \beta_j \psi(\boldsymbol{x}_i^T \boldsymbol{\gamma}_j) + e_i = \boldsymbol{\beta}^T \boldsymbol{\eta}_i(\boldsymbol{\gamma}) + e_i, \qquad (1)$$

Where $\boldsymbol{\beta} = (\beta_i, \dots, \beta_M)^T$, $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})^T$ $\boldsymbol{\eta}_i(\boldsymbol{\gamma}) = \left(\psi(\boldsymbol{x}_i^T \boldsymbol{\gamma}_1)\right)^T$, $i = 1, \dots, N$. Also, the errors e_i are assumed to be iid $N(0, \sigma^2)$. The regression part of the model can be identified with the feed-forward neural network. The $\boldsymbol{x}_i$ are the inputs, M is the number of hidden nodes (here assumed to be known), and β_j are the weights attached to the hidden notes $j (j = 1, \dots, M)$. The $\boldsymbol{\gamma}_j's$ are the weights attached to the inputs for node $j (j = 1, \dots, M)$, and ψ is the activation function. We write $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)^T$.

At the first stage of the hierarchical prior, $\boldsymbol{\beta}, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_M$ are mutually independent with $\boldsymbol{\beta} \sim N(\mu_\beta \mathbf{1}_M, \sigma_\beta^2 \mathbf{I}_M)$ and $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_M \overset{iid}{\sim} N(\boldsymbol{\mu}_\gamma, \mathbf{S}_\gamma)$. At the second stage, the prior parameters $\mu_\beta, \boldsymbol{\mu}_\gamma, \sigma^2, \boldsymbol{\sigma}_\beta^2$ and $\mathbf{S}_\gamma$are mutually independent with $\mu_\beta \sim N(a_\beta, A_\beta), \boldsymbol{\mu}_\gamma \sim N_p(\boldsymbol{a}_\gamma, \boldsymbol{A}_\gamma)$, $\sigma^2 \sim IG(\, c_\sigma/2, c_\sigma C_\sigma/2$,
$\sigma_\beta^2 \sim IG(\, c_\beta/2, c_\beta C_\beta/2\ ))$ and $\boldsymbol{S}_\gamma \sim IW(c_\gamma, c_\gamma^{-1} \mathbf{C}_\gamma^{-1})$.

Here IG and IW denote the inverse gamma and inverse Wishart distributions. Specifically, σ_β^2 has p.d.f

$$f(\sigma_\beta^2)\ \alpha\ exp\left(-\frac{c_\beta C_\beta}{2\sigma_\beta^2}\right)(\sigma_\beta^2)^{-c_\beta/2-1}, \qquad (2)$$

and $\mathbf{S}_\gamma$ has p.d.f

$$f(\mathbf{S}_\gamma)\alpha\ |\mathbf{S}_\gamma|^{-(c_\gamma+p+1)/2} \exp\left[-\tfrac{1}{2} tr(\boldsymbol{S}_\gamma^{-1} c_\gamma \boldsymbol{C}_\gamma)\right]. \qquad (3)$$

The selection of priors is usually problem-specific. Ideally, one would like to elicit these priors from past history. For this work, the above priors were chosen for $\mu_\beta, \boldsymbol{\mu}_\gamma, \sigma^2, \sigma_\beta^2$, and $\mathbf{S}_\gamma$.

The joint posterior of $\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma_\beta^2$, and $\mathbf{S}_\gamma$,given **y**, is $f(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\beta}, \mu_\beta, \boldsymbol{\mu}_\gamma, \sigma^2, \sigma_\beta^2, \mathbf{S}_\gamma | \mathbf{y})$

$$\alpha \prod_{i=1}^{N} [\exp\{\phi_i^{-1}(\theta_i y_i - k(\theta))\}](\sigma^2)^{-N/2} \times \exp\left[-(2\sigma^2)^{-1} \sum_{i=1}^{N} (\theta_i - \boldsymbol{\eta}_i^T(\boldsymbol{\gamma})\boldsymbol{\beta})^2\right] \times (\sigma_\beta^2)^{-M/2} \exp [$$

$$- \|\boldsymbol{\beta} - \mu_\beta \mathbf{1}_M\|^2 / 2\sigma_\beta^2)] \times |\mathbf{S}_\gamma \quad|^{-M/2} \exp [-\frac{1}{2} \sum_{j=1}^{M} (\boldsymbol{\gamma}_j - \boldsymbol{\mu}_\gamma)^T \mathbf{S}_\gamma^{-1} (\boldsymbol{\gamma}_j - \boldsymbol{\mu}_\gamma)] \times \exp [$$

$$- (2A_\beta)^{-1} (\mu_\beta - a_\beta)^2 - \frac{1}{2} (\boldsymbol{\mu}_\gamma - \boldsymbol{a}_\gamma)^T \boldsymbol{A}_\gamma^{-1} (\boldsymbol{\mu}_\gamma - \boldsymbol{a}_\gamma)]$$

$$\times \exp[-(c_\sigma C_\sigma)/(2\sigma^2)]\,(\sigma^2)^{-c_\sigma/2-1}$$

$$\times \exp[-(c_\beta C_\beta)/(2\sigma_\beta^2)]\,(\sigma_\beta^2)^{\times \exp}[-(c_\sigma C_\sigma)/(2\sigma^2)]\,(\sigma_\beta^2)^{\times \exp}[-(c_\sigma C_\sigma)/(2\sigma^2)]\,(\sigma^2)^{c_\beta/2-1}$$

$$\times |\boldsymbol{S}_\gamma|^{-(c_\gamma+p+1)/2} \exp\left[-\frac{1}{2} tr(\boldsymbol{S}_\gamma^{-1} c_\gamma \boldsymbol{C}_\gamma)\right].$$

These full conditionals are as follows:

(a) $\boldsymbol{\beta} | \boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma_\beta^2, \boldsymbol{S}_\gamma, \boldsymbol{y}$

$$\sim N\left[\left(\sigma^{-2} \sum_{i=1}^{N} \boldsymbol{\eta}_i(\boldsymbol{\gamma}) \boldsymbol{\eta}_i^T(\boldsymbol{\gamma}) + \sigma_\beta^{-2} \mathbf{I}_M\right)^{-1} \times \left(\sigma^{-2} \sum_{i=1}^{N} \theta_i \boldsymbol{\eta}_i(\boldsymbol{\gamma}) + \sigma_\beta^{-2} \mu\beta \mathbf{1}_M\right).\right.$$

$$\left. \times \left(\sigma^{-2} \sum_{i=1}^{N} \boldsymbol{\eta}_i(\boldsymbol{\gamma}) \boldsymbol{\eta}_i^T(\boldsymbol{\gamma}) + \sigma_\beta^{-2} \mathbf{I}_M\right)^{-1}\right];$$

(b) $\sigma_\beta^2 | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{S}_\gamma, \boldsymbol{y} \quad \sim IG\left(\frac{M+c_\beta}{2}, \frac{\|\boldsymbol{\beta} - \mu_\beta \mathbf{1}_M\|^2 + c_\beta C_\beta}{2}\right)$;

The pdfs (a) and (b) are the full conditionals for the posterior distribution.

2.3 Activation functions

2.3.1 Sigmoid Activation function

The sigmoid function is a mathematical function that gives a sigmoidal curve; a characteristic curve for its S shape. This compresses the input to any value between 0 and 1 and makes the model logistic. This function is known as a special case of logistic function defined by the following formula:

$$f(x) = \frac{1}{1 + e^{-x}}$$

2.3.2 Hyperbolic Tangent Activation function(TANH)

Another commonly and mostly utilized activation function is the *tanh* function. This is a nonlinear function, characterized in the scale of values (-1, 1). The function is specified by the formula:

$$f(x) = \tanh x$$

2.3.3 Rectified Linear Unit Activation function (ReLU)

Rectified Linear Unit (ReLU) is a predominantly utilized activation function. The function is defined by the following
formula:

$$f(x) = 0, \text{when } x < 0$$
$$x, \text{when } x > 0$$

The scale of the result is between 0 and infinity. RELU finds usage in computer vision and speech identification using deep neural networks.

2.4 Predictive Measures

The following predictive measures are used to compare the performance of BNN under different activation functions(RELU, Sigmoid and Hyperbolic Tangent) and training sets(70, 80 and 90).

(1) Accuracy = (TN + TP)/ (TN + TP + FN + FP)
(2) Precision = (TP)/(TP + FP),
(3) Recall = TP /(TP + FN),
(4) F1 - Score = 2 × (precision × recall)/(precision + recall)

where TN, TP, FN and FP True Negative, True Positive, False Negative and False Positive respectively. (1)-(4) above were obtained in the confusion matrix and then used to compute the performance criteria. The confusion matrix is a performance measurement in machine learning classification problems.

3 Results and Interpretations

Fig. 1 gives the frequency and percentage distribution of the responses to the questions in the Google form in the different regions where the incidents occurred. It shows from the plot that majority of the terrorists activities captured in this study occurred in the Southwest with a frequency of 608(68%) and South South the lowest with the frequency of 20(2%). North Central has a frequency of 87(9%), North East has 88(10%), North West 55(6%) and South East 67(7%).

The nature of the terrorist attack was categorized into different levels of attack, namely: armed assault, assassination, bombing/explosion, facility/infrastructure attack, hijacking, unarmed assault and unknown type. It shows that Armed Assault has the highest occurrence of 271(30%) and the minimum as Barricade Incident with frequency of 12(1%). Assassination has 37(4%), Bombing/Explosion has 26(3%), Facility/Infrastructure Attack has 66(7%), Hijacking has 91(10%), Kidnapping has 162(18%), Unarmed Assault has 83(9%) and unknown occurrence has 169(18%) frequencies.

3.1 Model Performance results

This section discusses the statistical analysis of the terrorism data using the Bayesian Neural Network with three activation functions namely: Hyperbolic Tangent (Tanh), Rectilinear (RELU) and Sigmoid activation functions. The results were analyzed using the training sets of 70, 80 and 90. Tables 1-3 below show the performance of BNN in predicting the occurrence of the attack type by terrorism under different activation functions and training sets. The Tables compared using the accuracy, precision, Recall and the F1 score performance of the models. The results also categorize these performance measures for training and test sets data. This is to check for the performance of BNN in ameliorating over-fitting in Neural Network models.

Measures	**Activation functions**					
	ReLU		Sigmoid		TANH	
	Train	Test	Train	Test	Train	Test
Accuracy	0.766	0.851	0.779	0.865	0.791	0.879

Precision	0.272	0.302	0.325	0.361	0.373	0.414
Recall	0.295	0.328	0.353	0.392	0.41	0.456
F1 Score	0.253	0.281	0.325	0.361	0.369	0.41

Table 2: Performance of BNN using different activation functions at 80% training set

Measures	**Activation functions**					
	ReLU		Sigmoid		TANH	
	Train	Test	Train	Test	Train	Test
Accuracy	0.843	0.936	0.857	0.952	0.87	0.967
Precision	0.299	0.332	0.358	0.397	0.41	0.455
Recall	0.325	0.361	0.388	0.431	0.451	0.502
F1 Score	0.278	0.309	0.358	0.397	0.406	0.451

Table 3: Performance of BNN using different activation functions at 90% training set

Measures	**Activation functions**					
	ReLU		Sigmoid		TANH	
	Train	Test	Train	Test	Train	Test
Accuracy	0.899	0.959	0.914	0.970	0.928	0.986
Precision	0.305	0.339	0.365	0.405	0.418	0.464
Recall	0.332	0.368	0.396	0.44	0.46	0.512
F1 Score	0.284	0.315	0.365	0.405	0.414	0.46

Tables 1-3 show the comparison of activation functions in Bayesian Neural Network model. It can be deduced that Hyperbolic-Tangent activation function has better prediction compared to other activation functions as it produces highest accuracy values for all the training sets and using both the training data and the test data.

3.2 Relative Importance

The degree of importance and contribution of the explanatory variables to the target variable is reported here.

Table 6: Relative Importance

Variable	**Relative Importance**
State of the incident	52.303
Number of Perpetrators	46.885
Weapon Type	0.812
Target/Victim Type	0.000

4.0 Conclusion

The performance of Bayesian Neural Network under different activation functions, namely; Sigmoid, Hyperbolic-Tangent and ReLU activation functions was carried out. The result shows that Hyperbolic-Tangent activation function predicts the attack type of terrorist activities well. The most contributing variable in determining the attack type of terrorists activity is the state in which the attack occurs, followed by the number of terrorists intended to carry out the assignment and so on. This study will be useful for government and security agents to plan for future terrorist attacks and to prevent such from happening in Nigeria.

References

(1) Olufemi A. Odeniyi, Mabel E. Adeosun and Tayo P. Ogundunmade, "Prediction of terrorist activities in Nigeria using machine learning models" *Innovations*, Number 71, pp 87-96, December 2022.

(2) M. Irfan Uddin, Nazir Zada, Furqan Aziz, Yousaf Saeed, Asim Zeb, Syed Atif Ali Shah, Mahmoud Ahmad Al-Khasawneh, Marwan Mahmoud, "Prediction of Future Terrorist Activities Using Deep Neural Networks", *Complexity*, vol. 2020,Article ID 1373087, 16 pages, 2020. https://doi.org/10.1155/2020/1373087.

(3) Ogundunmade, T. P. and Adepoju, A. A. (2021). The Performance of Artificial Neural Network Using Heterogeneous Transfer Functions. International Journal of Data Science. 2: 92-103. https://doi.org/10.18517/ijods.2.2.92-103.2021.

(4)Ogundunmade TP, Adepoju AA, Allam A. (2022a) Stock price forecasting: Machine learning models with K-fold and repeated cross validation approaches. *Mod Econ Manag,* 1: 1. DOI: 10.53964/mem.202200.

(5) Ogundunmade TP, Adepoju AA, Allam A. (2022b) Predicting crude oil price in Nigeria with machine learning models. Mod Econ Manag; 1:4. DOI: 10.53964/mem.2022004.