ISSN 2616-5775 Vol. 7, Issue 6: 32-37, DOI: 10.25236/AJCIS.2024.070605

Artificial neural network modelling in GIS spatial analysis

Xingxin Yang[1], Yongkang Wang[2], Ke Lu[1], Youquan Wu[1], Di Zhao[1]

[1]*Guilin University of Technology at Nanning, Chongzuo, Guangxi, China*
[2]*Guangxi Vocational and Technical College of Communications, Nanning, Guangxi, China*

Abstract: *This study aims to explore the application of Artificial Neural Networks (ANN) in spatial analysis of Geographic Information Systems (GIS) to improve the accuracy and efficiency of spatial data analysis. By combining ANN with GIS, the analysis focuses on land use and environmental monitoring data, especially the prediction of air quality index (AQI). Traditional GIS methods have limitations in dealing with complex nonlinear and high-dimensional data, while ANN can effectively solve these problems through its self-learning and self-adaptive capabilities. The study performed detailed pre-processing of raw data, including data cleaning, standardisation and normalisation to ensure data quality and consistency. These data, including land use type, surface temperature, AQI and population density, were converted into numerical vectors suitable for ANN model processing, laying the foundation for model training. For model construction, the BP neural network model was chosen, and the network structure and parameters were optimised through experiments. The number of nodes in the input layer is consistent with the feature dimension, the number of nodes in the hidden layer is determined by experiment, and the number of nodes in the output layer is set according to the task requirements. During the training process, a back-propagation algorithm is used to continuously adjust the weights and biases to minimise the mean square error (MSE). To prevent overfitting, the study introduces L2 regularisation and cross-validation methods to improve the generalisation ability of the model. After the training was completed, the predictive performance of the model was evaluated using the test set data, and the results showed that the model exhibited high accuracy in the AQI prediction task, and the predicted values were highly consistent with the actual values, which verified the effectiveness of ANN in GIS spatial analysis. The study also mapped the prediction results onto geospatial space through spatial visualisation techniques to generate AQI distribution maps for each region, which visually demonstrated the spatial distribution of air quality and provided an important reference for environmental management and decision-making. The AQIs of industrial areas and high population density areas are higher, while those of woodlands and waters are lower, which is consistent with the actual geographic features and human activity patterns. Despite the remarkable results of the study, there are still some limitations. Data quality and data quantity have a direct impact on model performance, and future research should consider collecting more diverse and larger data sets. In addition, the 'black box' nature of the ANN model makes it difficult to explain its decision-making process, and future research should explore methods that incorporate explanatory techniques to improve the transparency and interpretability of the model.*

Keywords: *Geographic Information Systems (GIS); Artificial Neural Networks (ANN); Spatial Analysis; Backpropagation Neural Networks (BP Neural Networks); Air Quality Index (AQI)*

1. Introduction

Geographic Information Systems (GIS), as a powerful tool for capturing, storing, analysing, managing and presenting spatial data, has achieved wide application in a variety of fields. The core of GIS lies in its ability to analyse spatial data, which has enabled it to excel in areas such as urban planning, environmental monitoring, resource management and disaster response. With the increasing amount of data and the diversification of data types, traditional spatial analysis methods have gradually exposed their limitations in dealing with these complex data [1]. Therefore, seeking more efficient and accurate analysis methods has become the focus of current research.

Artificial neural network (ANN) is a computational model that imitates the biological nervous system with strong self-learning and self-adaptive ability, and it has excellent performance in dealing with complex nonlinear problems, pattern recognition and prediction tasks, and its application has been

ISSN 2616-5775 Vol. 7, Issue 6: 32-37, DOI: 10.25236/AJCIS.2024.070605

extended from image recognition to speech recognition, natural language processing and other fields. Introducing ANN into GIS spatial analysis can make up for the shortcomings of traditional methods and provide new ideas and methods for dealing with complex spatial data.

Currently, spatial analysis in GIS mainly relies on traditional statistical methods and spatial interpolation techniques, which are more effective in dealing with linear relationships and smaller data sets. However, with the expansion of the data scale and the increase of complexity, these methods are not capable of capturing the nonlinear relationships and high-dimensional patterns in the data. In contrast, ANN can better handle high-dimensional and non-linear data through multi-layer network structure and complex weight adjustment mechanism, thus improving the accuracy and efficiency of analysis.

This study aims to explore the feasibility and effectiveness of applying ANN to GIS spatial analysis. Specifically, the study chooses a BP neural network model to focus on analysing and predicting the air quality index (AQI) in conjunction with the land use and environmental monitoring data of an area [2]. Through this application case, the performance and advantages of ANN in spatial analysis are verified. The study not only theoretically proposes a new approach to combine ANN with GIS, but also experimentally proves its effectiveness in practical applications.

In addition, this study focuses on the impact of data quality and processing on model performance. Data cleaning, standardised processing, and reasonable training and validation methods are all key steps to ensure the accuracy of the model. Through detailed data preprocessing and scientific model construction, the study seeks to provide a systematic approach for future related research.

The significance of the study lies not only in verifying the effectiveness of ANN application in GIS, but also in exploring new spatial analyses methods to enhance the responsiveness of GIS in the face of big data and complex problems [3]. With the improvement of computational power and the continuous progress of data science, the application of ANN in GIS will have a broader prospect, and can provide intelligent solutions for all kinds of spatial problems.

2. Theoretical foundation

A geographic information system (GIS) is a technology and tool for capturing, storing, analysing, managing and presenting spatial data. Spatial data refers to geographic location data, which can be discrete points, lines, surfaces, or continuous surface features. GIS spatial analysis refers to the processing and parsing of spatial data using GIS technology to discover the spatial patterns and relationships of geographic phenomena. Spatial analysis has a wide range of applications in the fields of environmental monitoring, urban planning, disaster management, resource assessment, etc., and can help decision makers make more scientific and reasonable judgements.

Artificial neural network (ANN) is a computational model that simulates the structure and function of biological nervous system, widely used in pattern recognition, classification, prediction and other tasks. ANN consists of a large number of nodes (neurons) and connections (weights), and the nodes are connected to each other through the weights to form a complex network structure. Each node receives input signals from other nodes, processes them through an activation function, and then passes the results to the nodes in the next layer [4]. ANN has the ability of self-learning and self-adaptation, and is able to optimise the performance of the model by adjusting the weights.

In GIS spatial analysis, traditional methods such as spatial interpolation and regression analysis have achieved certain results, but these methods show certain limitations when dealing with complex nonlinear problems, and the introduction of ANN provides a new solution for GIS spatial analysis, which is excellent in dealing with high-dimensional data, nonlinear relationships, and complex pattern recognition, and can effectively make up for the shortcomings of traditional methods.

When applying ANN to GIS spatial analysis, spatial data need to be converted into a format suitable for ANN processing. This usually involves converting geographic coordinates, attribute data, etc. into numerical vectors and normalising them to improve the training effect of the model. Common ANN models such as BP Neural Networks (Back Propagation Neural Networks) and Convolutional Neural Networks (CNNs) can be selected and designed according to the specific spatial analysis task [5]. BP Neural Networks are suitable for general regression and classification tasks, whereas CNNs perform particularly well when dealing with spatially characterised image data.

In the training process of ANN models, the labelling and division of data is crucial. The dataset is

ISSN 2616-5775 Vol. 7, Issue 6: 32-37, DOI: 10.25236/AJCIS.2024.070605

usually divided into a training set, a validation set and a test set, and the model parameters are optimised by repeated iterative training. In order to improve the generalisation ability of the model, techniques such as cross-validation and regularisation are often used. After the training is completed, the ANN model can be used for various spatial analysis tasks, such as the prediction of geographic phenomena, the classification of land use changes, and the identification of spatial distribution patterns of environmental pollution.

3. Research methodology

This study applies artificial neural network (ANN) models in geographic information system (GIS) spatial analysis, and the research methodology includes data preparation, model construction and optimisation, and result analysis [6]. Data preparation is the foundation of the study, and the selected dataset is derived from land use survey and environmental monitoring data of an area, which contains a variety of geospatial features and environmental attributes.

In the data preparation stage, the raw data are first cleaned and processed. The raw data contains land use type, surface temperature, air quality index (AQI), population density, etc. The accuracy and completeness of the data were ensured by processing missing values, outliers and noisy data. Subsequently, data standardisation was carried out to convert the values of each attribute to the same scale in order to facilitate the training of the neural network model. The data are shown in Table 1 below:

Table 1 Land use survey and environmental monitoring data of a region

Parcel Number	Land use type	Surface Temperature (°C)	AQI	Population density (person/km²)
1	Agricultural land	25	42	500
2	Industrial area	30	80	1500
3	Residential area	27	55	2000
4	Forest land	22	35	300
5	Water	20	30	100

In the model construction phase, a BP neural network was chosen as the main model for the study. The BP neural network consists of an input layer, a hidden layer, and an output layer. The number of nodes in the input layer is 5, which corresponds to the five features in the data table above. The number of nodes in the hidden layer was determined experimentally to obtain the best model performance. The number of nodes in the output layer depends on the specific analysis task, such as classification of land use types or prediction of environmental indicators [7]. The ReLU function was chosen for the activation function of the model and the mean square error (MSE) was chosen for the loss function to optimise the regression of the model.

The model is trained using a supervised learning method, where the dataset is divided into a training set and a test set with a ratio of 8:2. During the training process, the weights and biases of the network are continuously adjusted by the back-propagation algorithm to minimise the loss function. In order to prevent overfitting, regularisation techniques and cross-validation methods are used [8]. Regularisation suppresses the complexity of the model by adding weight penalty terms to the loss function. Cross-validation, on the other hand, improves the generalisation ability of the model by dividing the training and validation sets multiple times. During the training process, the loss function values and validation set errors are monitored to determine the convergence of the model.

4. Experiments and analysis of results

In the result analysis phase, the performance of the model on the test set is first evaluated to verify the predictive ability of the model by calculating metrics such as mean square error (MSE) and accuracy. Table 2 below shows some of the prediction results of the model on the test set compared with the actual values:

Table 2 Partial prediction results on the test set versus actual values

Plot No.	Actual AQI	Forecast AQI
1	42	40
2	80	78
3	55	57
4	35	33
5	30	32

From the results, it can be seen that the predicted values of the model are closer to the actual values, indicating that the constructed BP neural network performs well in dealing with GIS spatial analysis tasks. In addition, through the spatial visualisation technique, the prediction results are mapped onto the geospatial space to visualise the distribution of environmental indicators in each area, which helps further analysis and decision-making. Through the spatial distribution map, it can be observed that the AQIs of industrial areas and high population density areas are high, while the AQIs of woodlands and waters are relatively low, which is in line with the actual geographic features and human activity patterns [9]. Further analyses can also incorporate time-series data to study the temporal dynamics of AQI, which can support environmental monitoring and early warning on long time scales.

In this study, the effectiveness and advantages of BP neural network in GIS spatial analysis are verified through detailed data preparation, scientific model construction and optimisation, rigorous experimental design and result analysis. The experimental results not only proved the high prediction accuracy of the model, but also demonstrated the analysis results through spatial visualisation technology, which provided an important scientific basis for environmental management and decision-making. Future research can further optimise the model parameters and extend the dataset to improve the applicability and accuracy of the model.

5. Discussions

In this study, the combination of Geographic Information System (GIS) and Artificial Neural Network (ANN) demonstrated its unique advantages and wide application prospects in spatial analysis. Through detailed analyses of land use and environmental monitoring data, the BP neural network model performs well in predicting the air quality index (AQI) with high accuracy and reliability. The experimental results show that the model can effectively capture the complex nonlinear relationships in spatial data and provide accurate predictions, which opens up a new path for spatial analysis in GIS [10].

Firstly, the high accuracy of the model indicates that ANN has significant advantages in dealing with complex geospatial data. Traditional spatial analysis methods such as spatial interpolation and regression analysis, although effective in some scenarios, often appear to be inadequate when facing high-dimensional and nonlinear data. ANN is able to better adapt to and capture complex patterns in the data through its multilayered network structure and powerful self-learning capability. This is fully verified in this study, where the ANN model successfully predicts the AQI values of each region through the comprehensive analysis of multiple variables such as surface temperature and population density, which are highly consistent with the actual values.

However, there are some limitations and challenges in this study. Firstly, data quality and data quantity have a direct impact on the performance of the model. Although the data were adequately preprocessed in this study, there may still be noise and outliers in the raw data that are not completely eliminated, and these factors may have some impact on the model training and prediction results. In addition, the time span and spatial coverage of the data also limit the generalisation ability of the model [11]. In order to enhance the applicability of the model, future research should consider collecting more diverse and larger datasets, especially those that cover different time periods and a wider geographic range.

Another noteworthy issue is the interpretability of the model. Although ANN models are excellent in terms of prediction accuracy, their ‘black box’ nature makes it difficult to understand the internal mechanism and decision-making process of the models. This is especially important for geospatial analyses, because decision makers not only need accurate predictions, but also need to understand the logic and rationale behind the predictions. For this reason, future research should explore models that are more interpretable or incorporate other interpretive techniques (e.g., feature importance analysis) to enhance model transparency.

In addition, this study mainly focused on the prediction of AQI, but the potential applications of ANN models in GIS go far beyond that. For example, the model can be used to predict land use changes, assess the risk of natural disasters, and monitor ecological changes. These applications can not only improve the efficiency of environmental management, but also provide a scientific basis for urban planning and resource management. Therefore, future research can further expand the application scenarios of ANN in GIS and explore its potential in more complex spatial analysis tasks.

Finally, this study demonstrated the environmental conditions and prediction results of each region through the spatial visualisation of the experimental results. This intuitive presentation not only enhances the comprehensibility of the results, but also provides strong support for decision makers. However, the application of spatial visualisation techniques also needs to be continuously optimised by combining higher resolution data and more advanced graphic processing techniques to enhance the accuracy and visual effect of the analysis results.

6. Conclusion

Through this study, we can conclude that the combination of Geographic Information System (GIS) and Artificial Neural Network (ANN) shows great potential and application value in the field of spatial analysis. The study employed a BP neural network model to analyse land use and environmental monitoring data, especially for the prediction of air quality index (AQI), and achieved remarkable results. This research not only innovates in methodology, but also provides scientific basis for practical application.

Firstly, the study shows that the ANN model is able to effectively process and analyse complex spatial data, especially excelling in the face of non-linear and high-dimensional data. Compared with traditional spatial analysis methods, ANN models are able to capture complex patterns and relationships in data and provide more accurate prediction results through their multilayer structure and self-learning ability. The experimental results show that the predicted values of the BP neural network in the AQI prediction task are highly compatible with the actual values, proving its effectiveness in spatial analysis.

Secondly, this study emphasises the importance of data quality and data processing in model training and prediction. By cleaning and standardising the raw data, the quality of the input data is ensured, which improves the training effect and prediction accuracy of the model. Meanwhile, L2 regularisation and cross-validation methods are used to effectively prevent overfitting and improve the generalisation ability of the model. The application of these methods and techniques provides valuable experience and reference for similar studies in the future.

In addition, the study points out the prospect of widespread application of ANN models in geospatial analysis. Although this study mainly focuses on AQI prediction, the applicability of the model goes far beyond that. the ANN can be applied to a variety of fields such as predicting land use changes, assessing natural disaster risks, and monitoring ecological changes. These applications can not only provide a scientific basis for environmental management and decision-making, but also enhance the efficiency and accuracy of urban planning and resource management.

However, the study also identified some problems that need to be further explored and solved. For example, the ‘black box’ nature of the model makes it difficult to explain the decision-making process, which is particularly important for geospatial analysis and decision support. Future research should endeavour to improve the interpretability of the models and explore ways to incorporate other explanatory techniques so that the prediction results are not only accurate but also transparent and easy to understand. In addition, the time span and spatial coverage of the data limit the generalisation ability of the model, and future consideration should be given to expanding the size and diversity of the dataset to improve the applicability and accuracy of the model.

Finally, the study provides strong support for decision makers by visualising the prediction results through spatial visualisation techniques. This visualisation not only enhances the understandability of the results, but also provides intuitive data support for environmental management and planning. In the future, the application of spatial visualisation techniques should continue to be optimised and improved, combining higher resolution data and more advanced graphic processing techniques to further enhance the accuracy and visibility of the analysis results.

References

[1] Mollalo, A., Mao, L., Rashidi, P., & Glass, G. E. (2019). A GIS-based artificial neural network model for spatial distribution of tuberculosis across the continental United States. International journal of environmental research and public health, 16(1), 157.

[2] Lee, S., Ryu, J. H., Min, K., & Won, J. S. (2003). Landslide susceptibility analysis using GIS and artificial neural network. Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group, 28(12), 1361-1376.

ISSN 2616-5775 Vol. 7, Issue 6: 32-37, DOI: 10.25236/AJCIS.2024.070605

[3] Biswajeet, P., & Saro, L. (2007). Utilization of optical remote sensing data and GIS tools for regional landslide hazard analysis using an artificial neural network model. Earth Science Frontiers, 14(6), 143-151.
[4] Noack, S., Knobloch, A., Etzold, S. H., Barth, A., & Kallmeier, E. (2014). Spatial predictive mapping using artificial neural networks. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 40, 79-86.
[5] Lee, S., Song, K. Y., Kim, Y., & Park, I. (2012). Regional groundwater productivity potential mapping using a geographic information system (GIS) based artificial neural network model. Hydrogeology Journal, 20(8), 1511.
[6] Maithani, S., Jain, R. K., & Arora, M. K. (2007). An artificial neural network based approach for modelling urban spatial growth. ITPI Journal, 4(2), 43-51.
[7] Pradhan, B., & Lee, S. (2009). Landslide risk analysis using artificial neural network model focusing on different training sites. Int J Phys Sci, 3(11), 1-15.
[8] Pradhan, B., Lee, S., & Buchroithner, M. F. (2010). A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses. Computers, Environment and Urban Systems, 34(3), 216-235.
[9] Khoirunisa, N., Ku, C. Y., & Liu, C. Y. (2021). A GIS-based artificial neural network model for flood susceptibility assessment. International Journal of Environmental Research and Public Health, 18(3), 1072.
[10] Mokarram, M., Taripanah, F., & Pham, T. M. (2024). Spatial-temporal analysis of atmospheric environment in urban areas using remote sensing and neural networks. Sustainable Computing: Informatics and Systems, 42, 100987.
[11] Krivoguz, D. (2024). Geo-spatial analysis of urbanization and environmental changes with deep neural networks: Insights from a three-decade study in Kerch peninsula. Ecological Informatics, 102513.