Journal of Electronic Research and Application, 2024, Volume 8, Issue 4
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Application of Bayesian Analysis Based on Neural Network and Deep Learning in Data Visualization

Jiying Yang*, Qi Long, Xiaoyun Zhu, Yuan Yang

Yunnan Open University, Kunming 650224, Yunnan Province, China

Corresponding author: Jiying Yang, 19446964@qq.com

Abstract: This study aims to explore the application of Bayesian analysis based on neural networks and deep learning in data visualization. The research background is that with the increasing amount and complexity of data, traditional data analysis methods have been unable to meet the needs. Research methods include building neural networks and deep learning models, optimizing and improving them through Bayesian analysis, and applying them to the visualization of large-scale data sets. The results show that the neural network combined with Bayesian analysis and deep learning method can effectively improve the accuracy and efficiency of data visualization, and enhance the intuitiveness and depth of data interpretation. The significance of the research is that it provides a new solution for data visualization in the big data environment and helps to further promote the development and application of data science.

Keywords: Neural network; Deep learning; Bayesian analysis; Data visualization; Big data environment

Online publication: August 12, 2024

1. Introduction

In the data-driven era, the volume and complexity of data have exploded, challenging traditional data analysis and visualization methods. Traditional approaches struggle with large, high-dimensional, and unstructured data. Therefore, we explore new approaches that combine neural networks, deep learning, and Bayesian analysis. Neural networks and deep learning are good at processing large amounts of data and recognizing patterns, while Bayesian analysis can accurately quantify the uncertainty of data. By constructing and optimizing this method, we successfully realized the efficient visualization of large-scale data sets, demonstrating the great potential of this method in the field of data visualization. The results show that this method improves the efficiency and accuracy of data visualization. This not only promotes the development of data science but also provides technical support for research and application in related fields. To sum up, our study not only discusses the combination of these three methods in theory but also validates its effect in practice, which is of great significance.

2. Theoretical background and technical basis

2.1. Development of data visualization

As an important means of information presentation and data analysis, data visualization has experienced a long development process [1]. Early data visualizations relied on simple charts and hand drawings, and examples of this period include the 18th-century statistician John Snow's cholera map of London, and a 19th-century French engineer, Charles Joseph Minard, drawing the march of Napoleon's army. Simple as these early visualizations were, they demonstrated the power of data and graphics to convey information [2].

After entering the 20th century, with the rise and development of computer technology, data visualization began to change to the digital direction. In the 1960s, computer scientist John Tukey proposed the concept of Exploratory Data Analysis (EDA), which emphasizes the use of data visualization to discover patterns and anomalies in data [3]. During this period, basic statistical charts such as histograms and box plots were gradually popularized, providing intuitive tools for data analysis.

In the 1980s, data visualization entered a period of rapid development, and the progress of computer graphics technology promoted the emergence of interactive data visualization tools. In 1987, John Chambers, a scientist at Bell Labs, developed the famous statistical software S to graphically present the results of data analysis. During this period, more interactive visualization software such as Tableau and D3.js came out one after another, making data visualization not limited to static charts, and users could explore data through interactive operations.

In the 21st century, with the advent of the era of big data, the rapid growth of data volume and data dimension has put forward new challenges and requirements for data visualization [4]. Traditional 2D charts have been unable to cope with the demand for the presentation of complex data sets. High-dimensional data visualization and multivariate data visualization technologies have emerged. The development of machine learning and artificial intelligence has brought new opportunities for data visualization, and the accuracy and intelligence level of data visualization has been significantly improved through neural networks and deep learning models.

In recent years, visualization technology based on deep learning has gradually become a research hotspot [5]. Deep learning models can automatically extract features from massive amounts of data to produce more complex and precise visualizations. The neural network model combined with Bayesian analysis further improves the credibility and interpretation of the visualization results and provides a new idea and method for the research and application of data science.

The development of data visualization, from simple manual charts to complex interactive tools to intelligent visualization based on deep learning, fully reflects the profound impact of technological advances on data analysis and presentation. As technology continues to evolve, data visualization will continue to play an important role in the field of big data and artificial intelligence.

1.2. Overview of neural networks and deep learning

Neural networks and deep learning are two key technologies in the field of artificial intelligence in recent years [6]. Neural network simulates the structure of the human brain and realize the approximation of complex nonlinear functions by transmitting information layer by layer. A typical neural network consists of an input layer, a hidden layer, and an output layer. Each layer is connected by weighting, and the activation function is used to introduce nonlinear characteristics. Deep learning, a subfield of neural networks, enables the capture of more complex patterns and features by increasing the depth of the network, i.e. the number of hidden layers. With the improvement of computing power and the accumulation of massive data, deep learning models have demonstrated excellent performance in areas such as image recognition, speech processing, and natural language processing. Deep learning methods mainly include convolutional neural network (CNN), recurrent neural network (RNN), and generative adversarial network (GAN) [7]. CNN is usually used to process two-dimensional data such as images and extract multi-level image features through convolution and pooling operations. RNNs are suitable for processing sequence data, such as time series and text, in which a long short-term memory network (LSTM) solves the gradient disappearance problem of traditional RNNs. GAN generates realistic data samples through the antagonistic training of the generator and discriminator.

2.3. Basis of Bayesian analysis

Bayesian analysis is a statistical method that deduces the distribution of unknown parameters by calculating the posterior probability of data. The underlying theory is derived from Bayes' theorem, which describes the process of updating a probability distribution with prior knowledge and new data. Bayesian analysis excels in dealing with uncertainty and incorporating prior knowledge, making it particularly suitable for complex data environments. Bayesian networks and Bayesian inference are its main tools, the former for modeling dependencies between variables and the latter for parameter estimation. Combined with neural networks and deep learning, Bayesian analysis can further optimize model performance and improve the accuracy and intuitiveness of data visualization [8].

3. Bayesian analysis method based on neural network and deep learning

3.1. Combination of deep learning and Bayesian analysis

Deep learning and Bayesian analysis respectively show significant advantages in data processing and uncertainty quantification. Their combination not only enables in-depth mining of features in the data but also provides uncertainty assessment for the model, thus improving the robustness and interpretability of the model. Deep learning usually realizes efficient information processing and feature extraction through large amounts of data and complex models [9]. Bayesian analysis, on the other hand, describes the probabilistic nature of the model through statistical theory, enabling the model to have the functions of interpretability and uncertainty assessment. The combination of the two realizes the effective integration of data-driven and uncertainty-driven, providing comprehensive and reliable support for the visualization of complex data sets.

The superior performance of deep learning models in large amounts of data and complex tasks is widely recognized. Through multi-layer nonlinear transformation, deep learning models can automatically learn effective features from data, which greatly improves the efficiency of data processing and the accuracy of results [10]. Because of its powerful presentation capabilities, deep learning is increasingly becoming an important tool in data science. Deep learning models are often regarded as "black boxes", which are difficult to explain the model output and its decision-making process [11]. Bayesian analysis can effectively solve this problem by providing a probabilistic framework that can provide a clear explanation for the model output and its uncertainty.

By integrating the Bayesian statistical framework into the training process of neural networks, the deep learning model can not only process the data efficiently but also quantify the uncertainty. A common method is to model the weights in the neural network through the Bayesian model, to obtain the posterior distribution of the weights [12]. This method can not only capture the uncertainty of the model but also effectively avoid overfitting. In the concrete implementation process, variational inference or Markov Chain Monte Carlo (MCMC) method is usually used to approximate the posterior distribution, to realize the efficient training of the Bayesian deep learning model [13,14].

The quantification of model uncertainty through Bayesian analysis is helpful to improve the accuracy and efficiency of data visualization. In data visualization, it is not only necessary to show the basic characteristics of the data but also to clarify the uncertainty of the data and the model prediction results. Bayesian deep learning model can provide the confidence interval or probability distribution of the predicted results so that the data visualization chart can not only show the predicted results but also reflect its reliability, providing more comprehensive information support for data analysis and decision-making.

On the application side, Bayesian deep learning has made great strides in fields as diverse as medical image analysis, finance, and environmental science. In medical image analysis, the deep learning model combined with Bayesian analysis can provide disease detection results and uncertainty assessment, which significantly improves the reliability and accuracy of diagnosis and treatment. At the same time, in the financial field, the method is widely used in market risk analysis and prediction, which provides a probabilistic basis for investment decisions. In addition, in the field of environmental science, Bayesian deep learning models have also been applied to the prediction of climate change to better reflect the reliability and uncertainty of the prediction results [20].

3.2. Strategies for data visualization

In the process of data visualization based on the Bayesian analysis method based on neural networks and deep learning, the key is to transform the complex data processing process into intuitive and easy-to-understand visual information. To achieve this goal, we need to design the strategy from three levels data input, processing, and output.

The pre-processing stage of data is crucial [15,16]. It is important to ensure that the input data is sufficiently cleaned and normalized so that the model can be trained most efficiently. Data preprocessing not only includes traditional processing methods such as missing value filling, outlier processing, and feature scaling but also requires special processing of time series data and high-dimensional data to adapt to the needs of neural networks and deep learning models [17,18].

In the data processing stage, by constructing a suitable data flow graph, the neural network and deep learning model are used for multi-layer feature extraction and analysis. Specific strategies include choosing appropriate network architectures, such as convolutional neural networks (CNN) for image data analysis and recurrent neural networks (RNN) for time series data analysis. Bayesian analysis can provide uncertainty quantification for parameter estimation of deep learning models, which can improve the robustness and generalization ability of the models [19].

In the data output stage, the analysis results are presented in the form of graphs and charts through visualization techniques. Use graphical tools such as Matplotlib, Seaborn, or D3.js to generate rich visualizations, including but not limited to scatter, line, heat maps, and 3D maps [20]. Interactive visualization platforms, such as Plotly and Tableau, provide a deeper user interaction experience where users can dynamically adjust visual parameters to explore deeper layers of data.

4. Closing remarks

By constructing a neural network and deep learning model, combined with Bayesian analysis, this study has conducted in-depth discussion and application verification of data visualization. The study found that this approach can significantly improve the accuracy and efficiency of data visualization when processing large-scale data, making the data easier to understand and analyze. This method provides a new solution for data visualization in big data environments and shows a wide range of application prospects. However, there are some problems with the research. Firstly, the process of building and optimizing the model requires a lot of computational resources and may be difficult to apply in resource-limited environments. Secondly, the data set used in the study has its particularity, and the effect of the model on other data sets needs to be further verified. Finally, the complexity of Bayesian analysis increases the difficulty of model interpretation and requires further simplification and optimization. Future research can be deepened in the following aspects: The first is more extensive testing of different types of data sets to evaluate the versatility of the method. The second is to study more efficient computing methods to reduce the consumption of computing resources. The third is to explore the application of Bayesian analysis to other data science problems and expand its scope. The fourth is to optimize the interpretability of the model and develop more intuitive tools to help users better understand and utilize the data. In conclusion, this study provides a useful exploration and lays a foundation for the application of Bayesian analysis in data visualization. It is hoped that future research will overcome existing limitations, promote the development of data science, and provide more effective tools and methods.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Zhang W, Migiti A, Zheng F, et al., 2021, Uyghur Speech Keyword Retrieval Based on Deep Neural Network. Computer Age, 2021(11): 21–24.

[2] Tang JY, Zhou M, 2019, Case Analysis of Network Data Visualization Based on Big Data. Integrated Circuit Applications, 38(11): 80–81.

[3] Shi Y, Pu X, Shen L, et al., 2021, Object Detection System Based on Convolutional Neural Networks and Keywords. Computer Knowledge and Technology: Academic Edition, 17(08): 162–164.

[4] Ruan L, Wen SS, Niu YM, et al., 2021, Deep Neural Network Visualization Based on Explainable Basis Disassembly and Knowledge Graph. Chinese Journal of Computers, 44(09): 1786–1805.

[5] Lou YS, Guo W, 2022, Prediction of Nuclear Mass by Bayesian Deep Neural Networks. Acta Physica Sinica, 71(10): 32–41.

[6] Zhang W, 2021, Visual Research on Keywords of Master’s Thesis in Fine Arts. The Classic of Shanhaijing, 2021(13): 0337–0337.

[7] Liu S, Chang J, 2019, The Concept and Application of Big Data Visualization. Information Recording Materials, 22(09): 42–44.

[8] Cao R, Huang R, Huang X, et al., 2019, A Review of Deep Learning Research in China: Visual Analysis Based on CNKI Literature Keywords. Electronic Education in Primary and Secondary Schools, 2019(Z2): 91–94.

[9] Cao Q, 2020, Research on Keyword Distribution Visualization Based on Long Tail Theory. Science and Technology Achievements Management and Research, 2020(06): 32–37.

[10] Peng T, Qian H, Zhou J, 2022, Keywords and Visualizing the “Twin Engines” of Major Report Communication. Urban Party Newspaper Research, 2022(04): 61–64.

[11] Wang K, Yu H, 2021, Patent Keyword Analysis Method Based on Bayesian Network and Factor Analysis. China Science and Technology Information, 2021(18): 23–26.

[12] Zhang Y, 2019, Visual Analysis of Literature in Environmental Journals. Environmental Protection and Circular Economy, 41(05): 97–101.

[13] Langenberg B, Helm LJ, Mayer A,2024, Bayesian Analysis of Multi-Factorial Experimental Designs Using SEM. Multivariate Behavioral Research, 2024: 21–22.

[14] Huang Z, Zhou Y, Shu X, et al., 2019, Speech Enhancement Method by Combining Bayesian Estimation and Deep Neural Network. Minicomputer Systems, 40(01): 40–44.

[15] Liu Z, Liu D, 2024, Dynamic Modulus Analysis of Resonance Fracture Layer of Old Cement Concrete Slab Based on BP Neural Network and Deep Learning Model. Highway, 2024(06): 44–51.

[16] Zhang L, Liu X, 2024, Research on Two-Dimensional Rectangular Layout Optimization Method Based on Graph Neural Network and Deep Reinforcement Learning. Forging Equipment and Manufacturing Technology, 59(02): 117–122.

[17] Zhao J, Zhang Y, Shi Q, et al., 2024, Automatic Pedicle Screw Planning Based on Deep Learning Neural Network Technology. China Digital Medicine, 19(04): 84–91.

[18] Han XY, Xie MX, Yu K, et al., 2024, A Computational Power Network Resource Allocation Method Combining Graph Neural Networks and Deep Reinforcement Learning. Frontiers of Information Technology and Electronic Engineering, 25(05): 701–713.

[19] Chen Y, Huang X, 2024, Solution of Time Disturbance Wave Equation Based on Physical Information Neural Network. Oil and Gas Geophysics Committee of the Chinese Geophysical Society, Proceedings of the 6th Annual Petroleum Geophysical Conference, 2024: 4.

[20] Zhang W, Lin A, Yang X, et al., 2024, Fusion deep learning Bayesian filtering review. Acta Automatica Sinica, 50: 1–16.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.