World Journal of Information Technology
ISSN: 2959-9911
DOI: https://doi.org/10.61784/wjit3014

DATA ANALYSIS MODEL OPTIMIZATION AND ARTIFICIAL POTENTIAL FIELD ALGORITHM FUSION APPLICATION IN THE INFORMATION SYSTEM

ZongYang Du
Kwangwoon University, South Korea.
Corresponding Email: duzy@kw.ac.kr

Abstract: This paper discusses the fusion application of data analysis model optimization and artificial potential field algorithm in information system. This paper expounds the importance of data analysis in information system, and introduces the principle and characteristics of artificial potential field algorithm in detail. Through the research on the optimization method of data analysis model, the innovative idea of combining artificial potential field algorithm and data analysis model is proposed, including the strategy, step and technical realization of algorithm fusion. Analyze the fusion application in data clustering, path planning, abnormal detection of the advantages and effect, through the experimental comparison verified the effectiveness and superiority of the fusion method, to improve the ability of data analysis and intelligent level provides a powerful technical support and theoretical basis, has important application value and research significance.
Keywords: Information system; Data analysis model; Optimization; Artificial potential field algorithm; Fusion application

1 INTRODUCTION

In today's digital age, massive data have been accumulated in information systems. How to extract valuable information and knowledge from these data has become the key for enterprises and organizations to improve their competitiveness. As the core component of the information system, the performance of the data analysis model directly affects the quality and efficiency of the data processing and analysis. However, with the continuous expansion of data scale and increasing data complexity, traditional data analysis models face many challenges, such as low computational efficiency, insufficient accuracy, and poor adaptability to complex data structures.

As an intelligent optimization algorithm based on physical principles, the artificial potential field algorithm has good path planning and dynamic environment adaptability, and has achieved remarkable application achievements in the fields of robot motion control and automatic driving. The integration of the artificial potential field algorithm and the data analysis model in the information system is expected to give full play to the advantages of both, overcome the limitations of the traditional data analysis model, and provide a new and efficient solution for the data analysis of the information system. Therefore, the study of the data analysis model optimization and the fusion application of the artificial potential field algorithm has important practical significance and theoretical value[1-2].

2 OVERVIEW OF THE DATA ANALYSIS MODEL AND THE ARTIFICIAL POTENTIAL FIELD ALGORITHM

2.1 Data Analysis Model

Cluster analysis model: such as K-Means clustering algorithm, widely used in customer segmentation, image segmentation, biological classification and other fields. By dividing the data objects into different clusters, the data objects within the same cluster have high similarity, while the data objects between different clusters have great differences, thus helping users to discover the potential patterns and structures in the data.

Classification model: such as decision tree, support vector machine (SVM), etc., often used in credit evaluation, disease diagnosis, text classification and other tasks. These models provide support for decisions by learning on training data of known categories, constructing classification rules or decision boundaries, and thus making classification predictions on data of unknown categories.

Regression analysis model: such as linear regression, multiple regression, which has important applications in economic prediction, market trend analysis, quality control and other aspects. By establishing the mathematical relationship between independent variables and dependent variables, it predicts and estimates the dependent variables, and helps enterprises and organizations in decision planning and risk assessment.

High computational complexity: With the increase of data volume, the computational time and spatial complexity of many data analysis models increase exponentially, resulting in low efficiency in processing large-scale data and failure to meet the real-time requirements.

Easy to fall into the local optimal: in the process of searching for the optimal solution, some optimization algorithms are easy to fall into the local optimal solution, but miss the global optimal solution, thus affecting the accuracy and

performance of the model.
Limit on data distribution: Some data analysis models have certain assumptions on the distribution of data, such as normal distribution, but in practical application, the data often does not meet these assumptions, leading to the decline of the applicability and reliability of the model[3-4].

2.2 Artificial Potential Field Algorithm

The artificial potential field algorithm originates from the concept of the potential field in physics, and regards the environment of the robot or the target object as a potential field, which includes the gravitational potential fields and the repulsive potential fields. The gravitational potential field is generated by the target point and attracts the object close to the target; the repulsion potential field is generated by the obstacle and prevents the object from collision with the obstacle. The object is subjected to the resultant force of gravity and repulsive forces in the potential field to plan a collisionless path from the starting point to the target point.
Good real-time: can quickly respond to environmental changes, real-time adjustment of path planning, suitable for decision-making and control in dynamic environment.
Flexible path planning: more flexible paths can be generated according to the distribution of different targets and obstacles to avoid collision with obstacles, with strong adaptability.
High computational efficiency: Compared with some traditional path planning algorithms, the computational complexity of the artificial potential field algorithm is relatively low, and it can complete the path planning task in a relatively short time[5-6].

3 FUSION STRATEGIES AND METHODS

3.1 Fusion of the Data Preprocessing Stage

During the data preprocessing stage, the data were cleaned and filtered using the artificial potential field algorithm. The outliers and noise in the data are regarded as "obstacles", and the initial cleaning of the data is achieved by pushing these abnormal data points away from the normal data area by constructing the repulsion potential field. At the same time, according to the characteristics and target distribution of the data, the gravitational potential field is constructed to guide the data to the potential valuable areas and improve the quality and availability of the data.
Based on the idea of the artificial potential field algorithm, the features of the data are selected and weight optimized. Each feature is considered as a variable with a certain "potential energy", giving the corresponding gravitational or repulsion weight according to the correlation and importance between the feature and the target variable. By adjusting these weights, the features with high correlation with the target variable play a greater role in the subsequent data analysis, while the influence of low correlation or noise features is weakened, thus improving the performance and accuracy of the data analysis model[7].

3.2 Fusion in the Model Construction Stage

In the clustering analysis, the artificial potential field algorithm is introduced into the clustering process. First, the initial cluster center is determined based on the distribution of the data, and each data point is treated as a particle with mass and charge. Then, by calculating the gravity and repulsion between the data points, the data points will gradually gather to the cluster center of the potential field, forming a stable cluster structure. This fusion method can effectively avoid the sensitivity of the traditional clustering algorithm to the initial clustering center, improve the accuracy and stability of clustering, and accelerate the convergence of clusters.
For the classification model, the classification boundary was optimized by using the artificial potential field algorithm. Considering different categories of data samples as regions with different potential energy, by constructing appropriate potential field functions, the classification boundary can better separate the data of different categories and improve the classification accuracy. For example, in the support vector machine (SVM), the support vector can be regarded as a point with large gravity. By adjusting the potential field parameters, the classification superplane better fits the data distribution and reduces the classification error under the gravity of these support vectors[8].

3.3 Fusion in the Model Optimization Stage

The parameters of the data analysis model were optimized by using the artificial potential field algorithm. The parameters of the model are considered as particles in a multidimensional space, and the range of values of each parameter corresponds to the activity space of the particles. By calculating the influence of the parameter value on the model performance, the corresponding potential field function is constructed so that the parameters move in the direction that maximizes the model performance under the action of the potential field. For example, in the neural network model, we can use the artificial potential field algorithm to be used to optimize the weights and thresholds of the network to improve the training speed and accuracy of the model.
During the model optimization, the model is dynamically adjusted using the artificial potential field algorithm. The evaluation index of the model is regarded as a potential field with target value, and the weight and threshold of the model by monitoring the performance changes in the training process are adjusted, so that the model can develop more

accurately in the direction of optimal performance. At the same time, according to the performance of the model in the potential field, the shortcomings of the model are found in time, and targeted improvement and optimization are made to improve the overall performance and stability of the model[9].

4 FUSION APPLICATION CASE ANALYSIS

4.1 Data Clustering Application

In the field of customer segmentation, the artificial potential field algorithm and K-Means clustering algorithm are integrated and applied. The traditional K-Means algorithm is susceptible to the initial clustering center when processing large-scale customer data, resulting in unstable clustering results and computationally inefficient. By integrating the artificial potential field algorithm, the artificial potential field is first used to preprocess the customer data, remove the noise and abnormal data, and the gravitational and repulsion potential field is constructed according to the customers consumption behavior, attribute characteristics and other factors, to guide the customer data to gather in a reasonable clustering area. During the clustering process, the clustering center is dynamically adjusted according to the force situation of the data points in the potential field, making the clustering results more accurate and stable. The experimental results show that the integrated algorithm improves the accuracy of customer segmentation by about 15%, and the calculation time is shortened by about 20%, which can better meet the needs of enterprises for customer market segmentation, and provide strong support for precision marketing and customer relationship management.

4.2 Path Planning and Application

In the logistics distribution path planning, combine the artificial potential field algorithm and the genetic algorithm for optimization. Many factors need to be considered in the process of logistics distribution, such as traffic conditions, customer demand time window, vehicle load limit, etc. Traditional path planning methods are often difficult to meet these complex constraints at the same time. The artificial potential field algorithm is used to build the potential field model of logistics distribution environment, and the customer point is regarded as gravitational source, and the traffic congestion area and prohibited area are regarded as repulsion force source. By calculating the force of the vehicle in the potential field, the feasible distribution path is preliminarily planned. Then, the genetic algorithm is used to optimize these preliminary paths and find the globally optimal distribution path through crossing, variation and other operations. The experimental results show that the fusion algorithm can effectively reduce the total mileage of logistics distribution by about 10%, improve the on-time delivery rate by about 8%, reduce the logistics cost, and improve the efficiency and service quality of logistics distribution.

4.3 Application of Abnormal Detection

In the network security anomaly detection, the artificial potential field algorithm is combined with the density-based local anomaly factor (LOF) algorithm. Network traffic data has the characteristics of high dimension, large-scale and dynamic change, and it is difficult for the traditional abnormal detection method to accurately identify the complex network attack behavior. By fusion artificial potential field algorithm, first of network flow data pretreatment, according to the characteristics of the flow and normal behavior pattern build potential field model, the normal flow data points as stable particles in the potential field, while the abnormal flow data points by abnormal gravity or repulsive force, deviate from the normal flow distribution area. Then, these data points that deviate from normal regions were further evaluated and detected using the LOF algorithm. The experimental results show that the integrated anomaly detection method can improve the accuracy of anomaly detection by about 12%, reduce the false positive rate by about 10%, effectively enhance the network security protection ability, and detect and prevent network attacks in time[10].

5 CONCLUSION

Through the fusion application of data analysis model optimization and artificial potential field algorithm in information system, we can draw the following conclusions:

The fusion of artificial potential field algorithm and data analysis model shows significant advantages in many aspects, and can effectively solve the problems faced by the traditional data analysis model, such as low computing efficiency, easy to fall into local optimum, and limitation on data distribution hypothesis, and improve the accuracy, real-time and adaptability of data analysis.

In the stages of data pre-processing, model building and model optimization, reasonable fusion strategies and methods can give full play to the advantages of the artificial potential field algorithm, optimize the performance of the data analysis model, so that it can better cope with the complex data environment and diversified analysis needs.

Through practical application case analysis, we verify the effectiveness and superiority of the fusion algorithm in the fields of data clustering, path planning, abnormality detection and other fields, and provide new ideas and methods for enterprises and organizations in the aspects of information system construction, data analysis and decision making, which has important practical application value.

In the future, with the continuous development of information technology and the continuous growth of data, should further study data analysis model and artificial field algorithm fusion mechanism, explore more efficient and intelligent

fusion methods and technology, expanding its application field, for the intelligent development of information system and data driven decision support to provide more powerful technical support, promote the digital transformation and innovation development of industries.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Wang Rui. Analysis on the application of data analysis model of Shaanxi Agricultural specialty products. Fortune Today, 2024, (31): 8-10.
[2] Hao Shijia. Analysis of the index data of industrial enterprises in Jiangsu Province based on the optimization of the maximum information coefficient model. Communication and Information Technology, 2024, (05): 99-102.
[3] Xu Xiaojuan, Li Zongchao, Deng Mingchun, et al. Optimization method for transient thermal analysis model of high pressure turbine disk based on transition state test data. Aero-engine, 2024, 50(01): 57-63. DOI: 10.13477/j.cnki.aeroengine.
[4] Dong Bo, Luo Forest. Optimization and application of Text Semantic similarity Analysis Model in small datasets. Information Security Research, 2023, 9(10): 980-985.
[5] Chen Shuaishuai. Dimensionality reduction techniques and model optimization methods in big data analysis. Shandong University, 2023. DOI: 10.27272/d.cnki.gshdu.
[6] Fang Gang. Design and optimization of the energy efficiency analysis model of cement grouting construction quality based on data mining. And Jianghan University, 2023. DOI: 10.27800/d.cnki.gjhdx. 2023.000437.
[7] Shi Huijun, Liu Xianjun, Wang Zhigang. Evaluation of hydraulic expansion pipe: application of measurement data analysis and model optimization design. Chemical Equipment of China, 2023, 25(03): 3-8.
[8] Chen Huazhou. Quantitative analysis method of near-infrared spectroscopy and its application of agricultural informatization. Jinan University Press, 2022.
[9] Chang Fu. LIBS spectroscopy correction and model optimization method for rapid detection of high-temperature samples. University of Science and Technology Beijing, 2022. DOI: 10.26945/d.cnki.gbjku.
[10] Meng Lingming. Optimization of the plate shape control model for the unsteady process of hot continuous rolling based on data analysis. And Northeastern University, 2020. DOI: 10.27007/d.cnki.gdbeu.