

# Research on Intelligent Data Mining Technology Based on Geographic Information System

Junyu Chen

Alexa Music, Amazon, Seattle, WA 98121, US

---

**Abstract:** Geographic Information Systems (GIS), as a key means of managing, storing, and analyzing spatial information, have attracted much attention in the field of GIS due to the rapid development of big data and intelligent technology. This article elaborates on the intelligent data mining process based on information collection, data storage, data processing, and data analysis, supplemented by practical application cases, including urban traffic flow research, intelligent transportation system design, and commercial network layout optimization, to demonstrate its broad prospects in practical applications. This article summarizes how intelligent data mining technology can improve the efficiency and accuracy of GIS work, and predicts its future application prospects in smart cities, environmental monitoring, and spatial layout planning, providing a theoretical basis for the deep integration of GIS and data mining technology.

**Keywords:** Geographic Information System; Intelligent data mining; Data mining algorithms; application area.

---

## 1. Introduction

Geographic information system is a spatial data processing technology that integrates data collection, storage, analysis, and visualization. It is often applied in multiple industries such as urban planning, natural ecological protection, and resource management. With the rise of intelligent data mining technology, its excellent data processing and decision assistance capabilities have injected new vitality into the development of GIS. By filtering massive amounts of data, mining techniques can identify hidden patterns and knowledge, enhancing the practicality of GIS in various industry scenarios. Based on the intelligent data mining technology of geographic information system, this paper analyzes its application in practical cases such as traffic management and commercial layout, providing an innovative perspective for building intelligent cities.

## 2. Research Background and Current Situation

### 2.1. Development background and application prospects of intelligent data mining

In the current era of rapid development of information technology, intelligent data mining, as a combination of artificial intelligence and massive data processing, has demonstrated excellent capabilities in data processing and information mining. With the rapid upgrading of computer hardware and the deepening of algorithm research, this technology has penetrated into many industries such as finance, healthcare, logistics, and commerce, and plays a key role in urban intelligent management, social management optimization, and scientific research exploration. In recent years, with the emergence of cutting-edge technologies such as machine learning and deep learning, intelligent data mining technology has moved towards a higher level of intelligence. Its efficiency and accuracy in mining complex data patterns and predicting future trends have been improved[1].

Intelligent data analysis technology has opened up a new path for large-scale spatial data processing in the field of geographic information processing, providing powerful

assistance for spatial management and planning in dynamic environments. Looking ahead to the future, intelligent data analysis technology is expected to penetrate into multiple industries such as autonomous driving, precision agricultural production, and ecological environment monitoring, helping to address various challenges brought about by globalization and urbanization processes. The development potential of this technology is enormous, which can enhance the efficiency of data utilization, promote interdisciplinary integration, innovation, and cooperation, and lay a solid foundation for building an intelligent society.

### 2.2. The Importance of Data Mining Techniques in Geographic Information Systems

After the introduction of data mining technology, the functions of geographic information systems have been greatly enriched and improved, and their application scope and influence have also been enhanced. This is intuitively reflected in Figure 1. In the field of urban traffic governance, data mining technology can dynamically analyze traffic flow information, formulate more reasonable traffic planning schemes, and reduce traffic pressure. In the monitoring of ecological environment, data mining technology helps to filter out core environmental information from complex multi-source data, providing scientific support for environmental governance and resource management[2]. In the field of business location analysis, data mining technology assists enterprises in making more accurate site selection decisions and enhancing operational efficiency through in-depth analysis of consumer behavior and geographical location. In the field of disaster emergency management, data mining technology also plays an indispensable role. Through in-depth analysis of historical data on natural disasters such as earthquakes and floods, a disaster warning model has been constructed, providing a solid basis for the formulation of emergency response plans. At the level of spatial pattern analysis and trend prediction, the application of data mining enhances the insight ability of GIS in dealing with complex spatial problems.

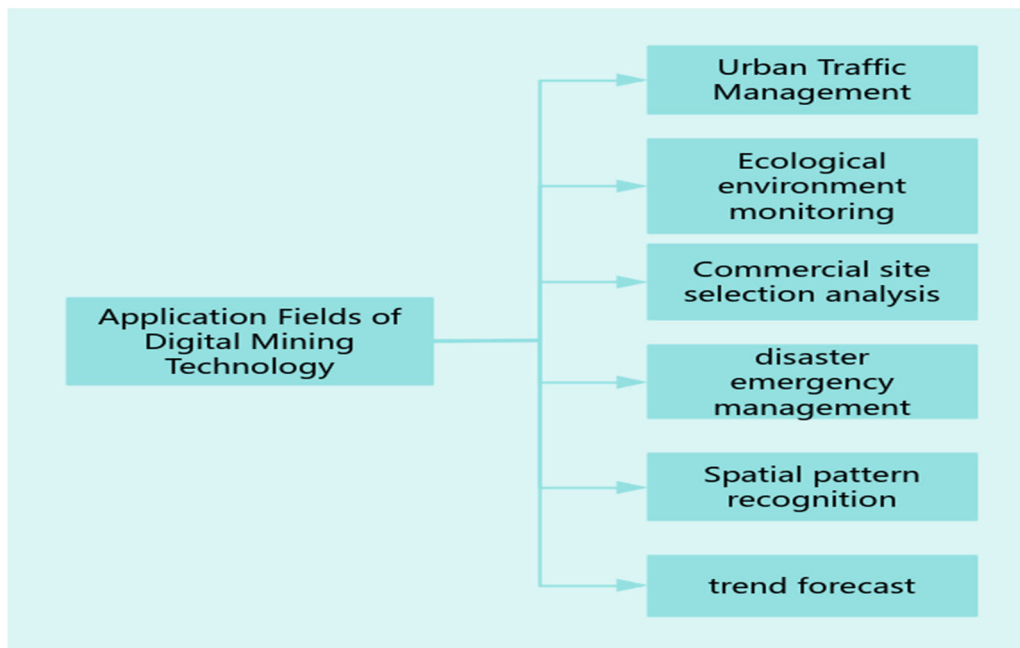


Figure 1. Application of Digital Mining Technology.

### 3. Advantages of Intelligent Data Mining Technology Based on Geographic Information Systems

The intelligent data mining technology that integrates geographic information systems combines the specificity of spatial information with the excellent processing efficiency of intelligent algorithms, occupying a leading position in spatial data analysis and decision-making. This technology can efficiently integrate and process heterogeneous data from different sources, making it possible to quickly screen valuable spatial models and attributes from large-scale datasets, improving the efficiency of data application and the depth of analysis. Compared with traditional GIS analysis methods, this intelligent data mining technology can perform real-time processing and prediction of dynamic data, enhancing the response speed to emergency events or complex situations. This technology automates the process of knowledge discovery, reducing the need for manual intervention. Its advantages are particularly prominent when dealing with multidimensional spatial analysis. With strong scalability and applicability, this technology can be flexibly applied to multiple industries such as transportation, environmental protection, and commerce, promoting the transformation of GIS from a simple auxiliary tool to an intelligent and comprehensive decision support system[3].

### 4. Intelligent Data Mining Technology Process and Application Based on Geographic Information System

#### 4.1. Data Collection and Storage

The data collection work covers rich sources of spatial and non spatial information data, such as remote sensing satellite images, surface sensor information, aerial photography data from drones, and statistical data related to the economy and society. These data sources are diverse and structurally different, and must rely on a standardized data collection architecture to implement unified and standardized

processing of data, ensuring that data of different types, accuracies, and time nodes can be integrated. Especially in real-time data collection, the comprehensive application of sensor network technology and Internet of Things technology provides strong support for the rapid collection of dynamic spatial data[4].

In the field of data storage, traditional storage solutions such as file management systems and relational databases are somewhat inadequate when faced with large and complex data storage tasks. This has led to the gradual rise of distributed storage technology and cloud storage services as industry standards. Taking Hadoop's HDFS, Google's Earth Engine, and Amazon's AWS as examples, they can provide efficient distributed data storage and computing capabilities. In order to optimize the speed of data queries, geographic information databases must establish effective spatial indexing mechanisms, such as R-trees, Quad trees, etc., to facilitate rapid retrieval of large-scale spatial data.

#### 4.2. Data Processing and Analysis

In multiple key steps of data processing, data cleaning is responsible for addressing various issues that may arise during the data collection phase, such as data noise, duplicate information, white spaces, and outliers. For blank values in the data, interpolation methods, average imputation, or prediction using the characteristics of neighboring data can be implemented to fill them. The identification of outliers can be achieved through the use of statistical methods, data density based detection techniques, or machine learning algorithms. After the data cleaning work is completed, it is data integration, which involves complex data from different channels with different structures. Through semantic unification and formatting adjustment, consensus is reached to eliminate semantic and structural inconsistencies caused by inconsistent sources. During the process, it is necessary to attach great importance to the diverse attributes of spatiotemporal data, and ensure high-quality data processing through sequence correspondence on the timeline and spatial dimension reset sampling. Before data analysis, data transformation is also necessary, which involves multiple

steps such as data normalization, dimensionality reduction, and feature processing. Standardization processing can unify data with different dimensions and prevent errors in the analysis process due to differences in magnitude[5].

The purpose of data analysis is to explore the deep patterns and laws of data, which is the cornerstone of data mining. In practice, techniques such as basic statistical description, spatial layout characteristic analysis, and trend exploration are usually used. With the help of hotspot analysis techniques, regions with abnormal geographical distribution can be identified, and time series research can provide insights into the evolution of data over time. Visualization tools play an important role in this process, such as geographic distribution heat maps, contour line drawing, and dynamic 3D demonstrations, which can vividly reflect the distribution characteristics and changing trends of data, helping researchers to quickly grasp the essence of data. Data organization and analysis are fundamental to data mining work and a key link in improving the efficiency of intelligent analysis and decision-making.

### 4.3. Research on Data Mining Algorithms

The data mining algorithm determines the depth of spatial data mining operations and the accuracy of the results. When conducting research on data mining algorithms in GIS environments, it is necessary to consider the unique properties of spatial data, such as data heterogeneity, temporal and spatial relationships, and multidimensional features[6]. Common mining algorithms can be divided into multiple aspects such as classification, clustering, association rule mining, regression analysis, and prediction models, as shown in Figure 2. The main focus of the research is how to improve and innovate algorithms based on the characteristics of geographic spatial data. In the field of data mining in geographic information systems, classification algorithms are a fundamental tool often applied in various aspects such as land resource classification and environmental hazard assessment. Common traditional classification techniques, such as decision trees and support vector machines, are known for their high computational efficiency and excellent

adaptability. However, when dealing with high-dimensional and complex pattern geological data, they may encounter performance bottlenecks. With the continuous development of deep learning technology, its performance in classification problems has improved. Typical breakthroughs in remote sensing image classification, such as convolutional neural networks, can effectively extract spatial features and perform efficient data classification. Using clustering analysis techniques to explore regularities and pattern features in spatial data.

The density based clustering method DBSCAN can detect highly dispersed abnormal blocks in geographic information, and is commonly used to locate areas with frequent traffic accidents and disaster warnings. In response to the complexity of geographic information distribution, scholars have developed improved clustering algorithms based on weighted distance and spatiotemporal constraints, making the results of clustering analysis more closely related to the actual geographical environment. Spatial correlation mining technology, as an important means in the intelligent process of geographic information systems, provides support for decision-making by analyzing the interrelationships between spatial data. Using Apriori algorithm to explore the potential relationship between land types and climate factors, and to assist in planning work in the agricultural field. By leveraging the advantages of geographic information systems in spatial analysis, researchers have innovatively introduced association rule mining techniques that integrate time and spatial constraints, improving the algorithm's adaptability to changing environments. In the field of prediction algorithms, the main focus is on the temporal variation characteristics of geographic location information. The traditional ARIMA model, relying on time series analysis, has played an important role in predicting natural disasters such as earthquakes and floods. With the development of technology, long short-term memory networks in the field of deep learning have been widely used in traffic flow analysis and climate simulation prediction due to their excellent time series data parsing ability.

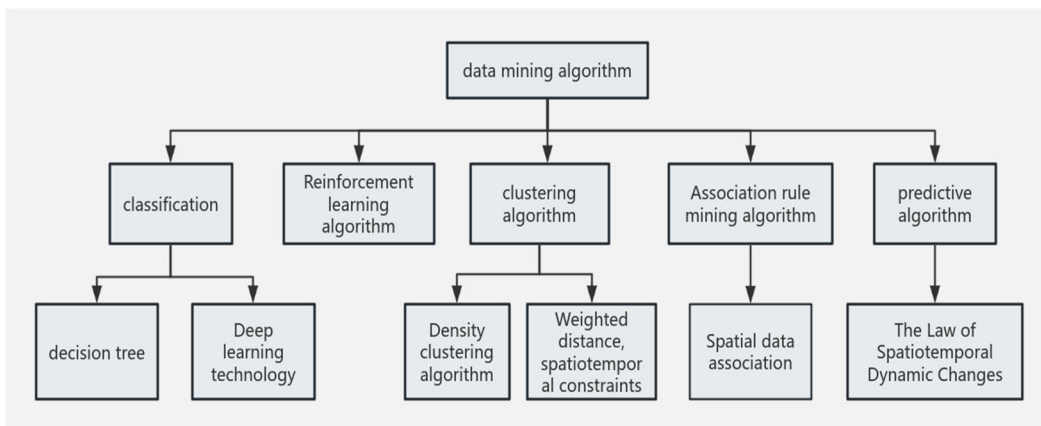


Figure 2. Data Mining Algorithm.

### 4.4. Analysis of Urban Traffic Flow and Intelligent Transportation Planning

In urban traffic management and layout optimization, intelligent data mining technology relying on geographic information systems plays an important role[7]. With the

rapid advancement of urbanization and the increasing complexity of transportation networks, real-time monitoring and rational planning of traffic flow have become the core strategies for improving urban operational efficiency and reducing traffic pressure. Geographic information system, as an efficient spatial information management platform,

integrates intelligent data mining methods and opens up new technological paths for traffic flow prediction and evaluation. The data collection channels involved in traffic flow analysis are diverse, covering various information sources such as traffic sensors, street monitoring equipment, satellite remote sensing technology, and mobile device GPS positioning. On the geographic information system platform, various data are fused and processed, relying on spatial correction and time series synchronization technology to comprehensively monitor traffic flow. By using intelligent data mining techniques and clustering analysis methods, we can effectively identify peak traffic flow periods and congested areas, providing data reference for optimizing traffic signal control and adjusting road network load. With the help of time series prediction algorithms, such as long short-term memory networks, it is possible to accurately predict changes in traffic flow during specific time periods, providing scientific references for scheduling and command in emergency situations.

The combination of geographic information systems and data mining technology has brought the possibility of dynamic adjustment and reasonable layout to the field of transportation planning. This technology provides comprehensive data support for planners by conducting in-depth research on road network architecture, vehicle distribution, and residents' travel habits. For example, using association rule mining techniques can reveal the inherent relationship between road layout and traffic flow, guiding the optimization design of roads. With the spatial analysis capability of GIS, path optimization methods such as Dijkstra's algorithm can calculate the most efficient driving route, providing scientific basis for fields such as cargo transportation and public transportation route design. By utilizing intelligent data analysis technology, real-time optimization of traffic management strategies can be achieved. For example, using simulation prediction systems to evaluate the effectiveness of different traffic control measures provides rigorous data support for formulating measures such as road congestion fees and vehicle restrictions. The advancement of integrated autonomous vehicle technology, geographic information systems, and data analysis technology provide core assistance for route optimization and real-time avoidance of autonomous navigation vehicles.

#### 4.5. Business Site Selection Based on Geographic Information Systems

The integration of geographic information systems and advanced data mining methods provides more rigorous and accurate decision-making assistance for enterprises in selecting commercial locations. When selecting a commercial location, it is important to comprehensively evaluate various variables such as regional location, potential consumer groups, transportation network conditions, market competition trends, and resident distribution. The integrated application of GIS and data mining technology has contributed efficient technical solutions for comprehensively processing and interpreting these complex spatial and non spatial data. In the process of selecting commercial locations, the primary task is to use GIS technology to integrate various information resources (including satellite image data, geographic information data, transportation network information, and market research results), and create a comprehensive spatial information warehousing system. By utilizing intelligent data mining techniques, this information can be analyzed to reveal

the distribution, consumption level, and behavioral habits of the population within a specific region. By using clustering analysis techniques, classify the distribution characteristics of the population and accurately target key market areas. Clustering model based on K-Means algorithm:

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (1)$$

In formula (1), J is the objective function of clustering,  $C_i$  represents the i-th cluster,  $\mu_i$  is the cluster center, and  $x - \mu_i$  represents the Euclidean distance between the data point and the center. By minimizing the J value, the population distribution has been classified into various geographic regions, accurately targeting the concentrated areas of the target customer group. In the process of selecting a commercial address, transportation conditions are a decisive factor. GIS technology is used to conduct in-depth analysis of traffic flow and road network layout, in order to evaluate the transportation convenience and accessibility of the designated area. Dijkstra's path optimization algorithm is used to calculate the shortest path:

$$d(u, v) = \min \{d(u, w) + c(w, v)\} \quad (2)$$

In formula (2),  $d(u, v)$  is the shortest path distance from node u to node v, and  $c(w, v)$  is the weight (such as travel time or distance). With the help of this algorithm, the connectivity between potential locations and key commercial areas can be evaluated. With the help of regression analysis methods in the field of machine learning, it is possible to predict the future market development prospects of a specific region, using a linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (3)$$

In formula (3), Y represents market growth potential,  $X_1, X_2, \dots, X_n$  are influencing factors (such as population growth, income level),  $\beta_0, \beta_1, \dots, \beta_n$  are regression coefficients, and  $\epsilon$  is the error term. The use of regression models can estimate the future profitability of the target area. Integrating this algorithm into the GIS analysis system makes the selection of commercial locations more systematic and accurate, endowing enterprises with solid data support and market competitiveness.

## 5. Conclusion

Intelligent data mining technology has become a core trend in the field of GIS, enhancing the ability to process, store, and analyze large-scale spatial data. The article reviews the research progress and application examples of intelligent data mining in the field of geographic information systems, and explores the significant advantages of this technology in improving operational efficiency, improving resource

allocation, and assisting scientific decision-making. By using specific application examples, such as monitoring urban traffic congestion and developing intelligent evacuation strategies, and optimizing commercial layout based on geographic information systems, the important role of it in practical applications has been demonstrated. Looking ahead, with the continuous innovation of big data and artificial intelligence technology, intelligent data analysis technology is expected to further expand the application scope of GIS, involving many fields such as environmental protection, precision agriculture, and emergency management.

## References

- [1] Pankaj Adhikari."Temporal Floodplain Change Assessment Using GIS: A Case of Tinau River, Butwal."International Journal of Environment 12.2(2023):103-120.
- [2] Lawmchullova Imanuel,and Rao Ch. Udaya Bhaskara."Estimation of siltation in Tuirial dam: a spatio-temporal analysis using GIS technique and bathymetry survey."Journal of Sedimentary Environments 9.1(2023):81-97.
- [3] K. M. Gojiya, et al."Remote Sensing and GIS Applications in Soil Salinity Analysis: A Comprehensive Review."International Journal of Environment and Climate Change 13.11 (2023):2149-2161.
- [4] Adarsha Gopalakrishna Bhat, et al."Geospatial Analysis of Kurumanpuzha Sub Watershed in the Chaliyar River Basin: A Remote Sensing and GIS Approach for Geomorphological Assessment."Journal of Experimental Agriculture International 45.11(2023):9-21.
- [5] Benaissi Lakhelifa, et al."Reply: Comment on “Geological mapping and mining prospecting in the Aouli inlier (Eastern Meseta, Morocco) based on remote sensing and geographic information systems (GIS)” by Benaissi et al. (China Geology, 5 (2022))."China Geology 6.4(2023):762-766.
- [6] Vipul Kumar, et al."Mapping Asian Elephant Habitat Suitability Index in Alur Taluk, Central Western Ghats: A GIS, Remote Sensing, and Analytical Hierarchy Process (AHP) Approach for Conservation Assessment."International Journal of Environment and Climate Change 13.10(2023):2052-2062.
- [7] Piniarski Witold."Challenges of a GIS-based physical-geographical regionalization of Poland."Environmental monitoring and assessment 195.9(2023):1125-1125.