Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 14, No. 3, 2024

Research on AIS Data Analysis and Its Integration into Discipline-specific Teaching

Yue Yan, Enze Wu, Taizhi Lv

College of Information Engineering, Jiangsu Maritime Institute, Nanjing 211170, China

Abstract: The processing of ship AIS (Automatic Identification System) data plays a crucial role in the digital transformation of the shipping industry. From data collection to data storage, data analysis, and data visualization, AIS data processing technology provides robust support for the shipping sector. In terms of data collection, the platform utilizes distributed message queues to aggregate AIS data. For data storage, the platform efficiently manages and stores massive amounts of AIS data using data warehouses. When it comes to data analysis, the platform mines and analyzes AIS data through Lambda, extracting valuable information such as ship trajectories and sailing speeds, thereby supporting areas such as ship management, navigation safety, and marine research. In data visualization, the platform intuitively displays ship navigation status through charts and maps, providing strong support for decision-making. As an essential component of big data applications, AIS data analysis technology is playing an increasingly important role in the shipping industry. The big data technology major of Jiangsu Maritime Institute integrates big data applications like AIS data analysis into its curriculum, taking the upgrading needs of the shipping industry as an entry point. It actively plans the cultivation of composite big data skilled talents under the backdrop of smart shipping, comprehensively serving the upgrading of the shipping industry and promoting the transformation from a major shipping country to a shipping powerhouse.

Keywords: Shipping industry; Digital transformation; AIS; Data analysis; Big data technology.

1. Introduction

The advent of Shipping 4.0 undoubtedly marks a new chapter in the development of the shipping industry, signifying its entry into a new era driven primarily by data [1]. This new trend is characterized by extensive data collection and in-depth application, where the shipping industry is accelerating its digital transformation by mining and utilizing valuable information from these data. The Automatic Identification System (AIS) data stands out as a crucial data source in the shipping field, and its importance cannot be overstated [2]. The massive AIS data, like information arteries, flow through every corner of the shipping industry, providing a primary channel for the penetration of big data technology. These data not only record basic information such as ships' navigation trajectories, speeds, and positions but also contain rich operational patterns and potential values, laying a solid data foundation for the intelligent development of the shipping industry. Research on AIS data processing has become a focal point in the shipping industry. By utilizing advanced data processing technologies and methods, the shipping industry can extract valuable information from massive AIS data, providing strong support for various fields such as intelligent maritime management, channel planning, abnormal ship monitoring, ecological protection, and port characteristic analysis [3]. These applications not only improve the operational efficiency and safety of the shipping industry but also promote harmonious development between the shipping industry and the ecological environment, pushing the shipping industry towards a more intelligent and green direction. Looking forward, with the in-depth development of the Shipping 4.0 era, AIS data processing technology will play a more important role in the shipping industry. Driven by data, the shipping industry will usher in a more prosperous, intelligent, and sustainable future.

Integrating AIS data analysis research into the teaching of Big Data Technology has a profound impact on cultivating students' data processing and analysis capabilities. Through in-depth study and practice of AIS data analysis, students not only hone their ability to handle massive amounts of data but also grasp the application of cutting-edge technologies such as data mining and machine learning in real-world problems. This process not only deepens students' understanding of big data processing workflows but also enables them to skillfully utilize advanced data analysis tools and techniques, thereby more effectively extracting valuable information from AIS data. This combination of theory and practice not only enhances students' professional skills but also lays a solid foundation for their future development in the big data field, enabling them to better tackle industry challenges and meet market demands.

2. Application of Ais In Smart Shipping

Big data is increasingly applied in the shipping industry, effectively driving innovation, efficiency, and service upgrades in ports and shipping, as well as promoting reform and innovation in government regulation and precise provision of public services. As a crucial data source for shipping, massive AIS data serves as an important pathway for applying big data technology in the shipping industry, providing a data foundation for the smart transformation of the shipping sector. Research based on massive AIS data has become a hotspot in the industry, with numerous applications in smart maritime management, channel calculations, abnormal ship detection, environmental protection, and port characteristic analysis, among others. AIS data analysis demonstrates extensive application value in maritime regulation and other fields.

2.1. Application of AIS data analysis in maritime regulation

AIS data analysis plays a vital role in maritime regulation. By collecting and processing ship AIS data in real-time, maritime regulatory authorities can comprehensively monitor the navigation status, position information, and navigation trajectories of ships, thereby achieving effective management of ship traffic. This not only helps to enhance navigation safety and prevent collision accidents but also provides timely and accurate information support for rescue and search operations. Simultaneously, it assists maritime management departments in marine environmental protection and ship emission monitoring, ensuring the sustainable development of the shipping industry. Ship trajectory analysis and ship collision avoidance are two key applications of AIS data analysis.

Ship trajectory analysis involves collecting and analyzing ship AIS data to understand the behavior and movement patterns of ships on the water. This analysis helps to comprehend ship routes, speeds, berthing positions, and potential navigation intentions. Scholars process AIS data through data mining, machine learning, statistics, and other methods to analyze ship trajectories. Lee et al. used the DBScan algorithm to perform cluster analysis on ship trajectories to discover ship behavior patterns and navigation laws [4]. Xiao et al. studied the problem of ship trajectory fusion and prediction based on multi-source information and proposed a method for ship trajectory fusion and prediction based on Kalman filtering and neural networks [5]. This method integrates information from various sensors such as AIS, radar, and satellites, improving the accuracy and reliability of trajectories and utilizing neural networks for future trajectory prediction. Yang et al. proposed a method for ship trajectory anomaly detection and prediction based on bi-LSTM [6]. This method uses bi-LSTM neural networks to model ship historical trajectories for trajectory prediction.

AIS data analysis technology enables ships to grasp real-time traffic situations in surrounding waters, thereby adjusting routes and speeds in a timely manner to avoid collision accidents. Rong et al. [7] proposed a method based on the sliding window algorithm to identify collision avoidance behaviors from ship trajectories, which is significant for deeply understanding the response measures of ships in potential collision scenarios. You et al. [8] selected ship AIS data from the waters near Zhoushan Island and performed preprocessing operations such as dynamic and static data matching, data cleaning, data sorting, and trajectory interpolation to analyze ship collision avoidance behaviors. Zhang et al. [9] collected, stored, and analyzed AIS data, obtaining ship historical trajectory information and motion parameters, and deeply mined ship navigation behaviors. Liu et al. [10] proposed a quantitative analysis method for quantifying collision risks, which combines static and dynamic information from AIS and incorporates a time-varying motion trajectory model, helping to identify high-risk hotspot areas for ship navigation.

2.2. Application of AIS data analysis in other fields

AIS data analysis demonstrates wide-ranging applications in fields beyond maritime regulation. In the logistics industry, AIS data analysis can track ship positions in real-time, optimize route planning, and improve transportation efficiency. In business analysis, AIS data is used for market trend prediction, risk management, and other purposes, providing strong support for corporate decision-making. n marine scientific research, AIS data serves as a valuable resource for ship trajectory analysis, marine environmental monitoring, and other studies. Furthermore, AIS data plays an important role in tourism planning, ecological protection, and other aspects, providing a scientific basis for decision-making and development in related fields by uncovering patterns and trends behind the data [11-13].

3. Key Technologies for Ais Data Processing

The technologies of data collection, data storage, data computing, and data visualization collectively constitute the core technical system of AIS (Automatic Identification System) stream data computing, providing strong support for realizing real-time monitoring and information services for ships. As shown in Figure 1, the system consists of a data collection layer, a data storage layer, a data computing layer, and a data visualization layer. The data collection layer is based on an AIS client program, which collects data from AIS base stations through a computer network and pushes the data to the data storage and data computing layers using Apache Kafka. The data computing layer includes data cleaning and data analysis. The data cleaning function performs operations such as AIS data deduplication, parsing, and erroneous data screening. Leveraging stream computing technology, Apache Flink is employed for real-time computation, with the results being pushed to the data analysis module and Redis cache. The data analysis function enables the statistics of relevant indicators. The data storage layer implements multi-level storage based on HBase, MySQL, and Redis. The data visualization layer adopts a front-end and back-end separation technology, where the back-end uses Spring Boot to implement business logic processing, and the front-end uses Vue to achieve visual representation.

3.1. Data collection based on Kafka

Kafka is a distributed message queue based on the publish/subscribe model, specifically designed for high-throughput distributed systems [14]. Kafka boasts better throughput, built-in partitioning, replication, and inherent fault tolerance, making it highly suitable for large-scale message processing applications. The platform adopts the Kafka framework, where Kafka producers convert AIS data from ship base stations into messages and send them to specified Kafka topics. Consumers, using Flink as the stream processing framework, read AIS data by subscribing to Kafka topics and receive and process the data sent by producers. Through this combination of Kafka and Flink, the platform achieves efficient AIS data transmission and processing.

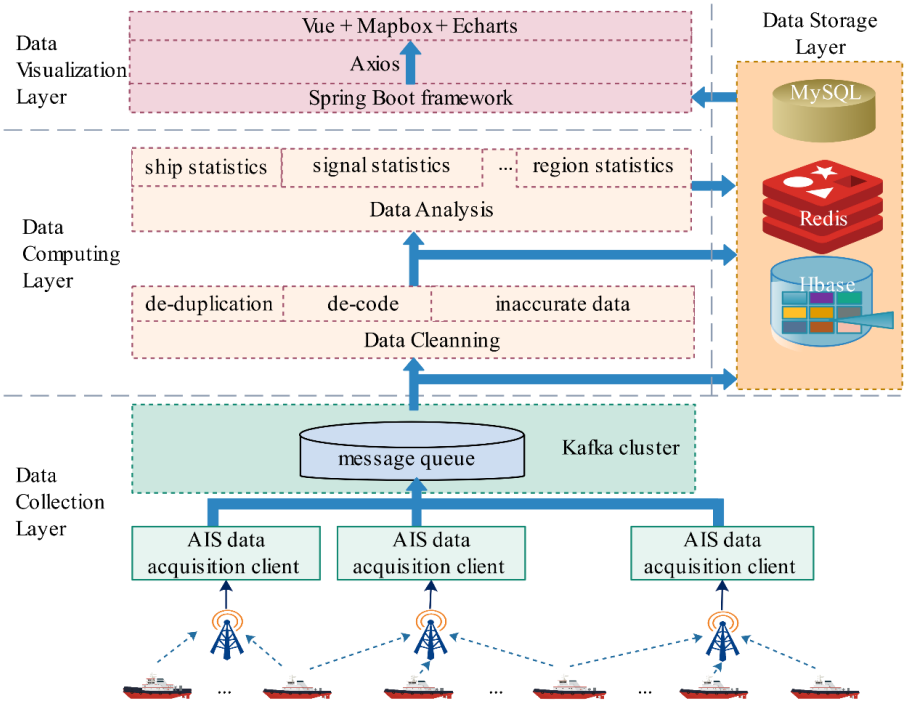

Fig.1 The AIS data processing architecture

3.2. Real-Time data computation based on Flink

Apache Flink is a powerful framework that supports the development and execution of various types of big data processing applications [15]. Flink's core features include unified batch and stream processing, fine-grained state management, event-time support, and guarantees of exactly-once state consistency. Leveraging these features, the AIS data processing platform fully utilizes Flink's real-time processing capabilities, achieving real-time reception of AIS data streams through Kafka connectors.

At the data computation layer, the platform parses and transforms the received data to meet the needs of subsequent processing. With Flink's powerful operator functions, the transformed data is deeply cleaned and filtered, effectively eliminating erroneous and abnormal data. The rigorously cleaned data is securely stored in the database, providing convenient and accurate data support for subsequent analysis and applications.

3.3. Data storage based on HBase and Redis

To facilitate the use of AIS data by various applications, AIS data is divided into two categories. The first category is real-time information, which stores real-time ship information and provides millisecond-level queries through Redis caching mechanisms. The second category is shipping historical trajectory information and historical region information, which provide second-level queries.

Redis is a high-performance distributed in-memory database based on key-value pairs, suitable for high-concurrency, low-latency application scenarios [16]. Applying Redis's distributed storage system to AIS real-time data helps improve system performance and meet real-time requirements. Real-time information includes two parts: (1) storing basic ship information with MMSI codes as keys; (2) storing ship MMSI codes with GeoHash-encoded location information as keys. GeoHash is an address encoding method that can encode two-dimensional spatial latitude and longitude data into a string, allowing for the quick retrieval of nearby target elements by comparing the similarity of GeoHash values. Redis's GEO, based on GeoHash encoding, uses sorted sets (Zset) to save geographic location coordinates, enabling rapid queries of ships within a certain area based on coordinates and radius.

HBase is a highly reliable, column-oriented, scalable NoSQL distributed database. HBase can query information from hundreds of millions of data within milliseconds [17]. Storing historical information in HBase, with reasonable settings for row keys and column families, ensures that massive AIS data is distributed in an orderly manner, meeting the demand for real-time querying of historical information. The data obtained from the message queue is stored in HBase as AIS historical trajectory data. Dynamic AIS includes 16 items of information, among which MMSI code, speed, course, longitude, latitude, and timestamp information are essential for maritime management and are grouped into the same column family in the design. In practical applications, there is a need to query the information of ships in a certain area at a specific time point in real time. This information can be queried directly from the historical trajectory table, but it requires scanning the entire table, which takes hours and cannot meet real-time requirements. Therefore, a historical region information table is designed, where region information and time information are stored as row keys, supporting spatio-temporal multidimensional data queries.

3.4. Data visualization based on front-end and back-end separation technology

As shown in Figure 2, a ship dynamic map based on AIS is constructed using technologies such as GeoServer, Spring Boot, and OpenLayers. It is divided into two modules: static map information and dynamic ship information. Adopting the traditional B/S architecture, it consists of four layers: data

layer, data processing layer, business logic layer, and presentation layer. The data layer includes TianDiTu (a Chinese geographic information service), vector tile-based channel charts, dynamic AIS data stored in Redis cache, and ship trajectory information stored in HBase. The data processing layer realizes the reading of different source data such as Redis, HBase, and MySQL. The business logic layer processes business logic and returns processed GeoJSON data to the front end. The presentation layer loads layers through OpenLayers and displays ship dynamic information based on GeoJSON data.

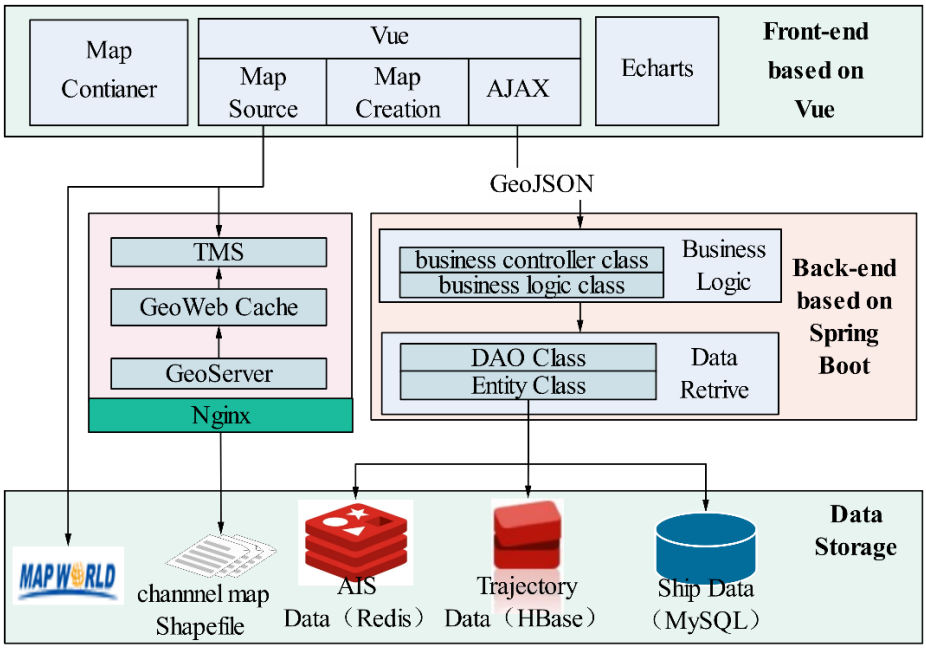

Fig.2 The data visualization architecture

4. Ais Data Analysis Applied To Teaching Practice

Smart shipping is a crucial component of building a powerful transportation nation, with data as its core driver. AIS data plays a significant role in this context. The emerging smart shipping industry in Jiangsu urgently requires high-quality, compound skilled talents who possess both shipping knowledge and big data technology skills. To better serve the digital transformation and upgrading of Jiangsu's shipping industry, Jiangsu Maritime Insitute has analyzed issues such as the basis, principles, and necessity for determining the distinctive direction of its big data major. Leveraging maritime characteristics, the college has constructed and practiced a big data technology curriculum system and talent cultivation model that integrates smart shipping features.

To reflect the shipping-specific big data talent cultivation research, the big data major incorporates AIS data analysis technology into the cultivation of big data talents. By enhancing practical teaching, scientific research innovation, and talent cultivation that integrate shipping features, the college aims to cultivate more high-quality talents with big data processing capabilities and innovative thinking, providing strong support for the development of both the shipping industry and the big data sector. The big data technology curriculum system revolves around teaching related to the AIS data analysis-centric ship-to-shore power data processing and query platform. All courses cover the entire range of knowledge and technology required for project development. The ship-to-shore power data processing and query platform is a smart shipping project that runs through the core courses of the professional group, serving as a common teaching carrier to organically link the professional group courses. Basic platform skills include foundations in information technology, programming, platform basics, and web basics, with the technical content used in the project broken down into corresponding courses. Centering on the ship-to-shore power data processing flow, from platform, acquisition, storage, analysis to visualization, this flow is integrated into the curriculum system. During the project completion process, students not only comprehensively apply professional knowledge and hone their operational skills but also develop abilities in autonomous learning, innovative thinking, teamwork, stress resilience, and responsibility-taking.

In accordance with the college's standard of "joining the mainstream while strengthening characteristics," without disrupting the existing curriculum system of big data technology, knowledge of smart shipping and AIS data is integrated into professional course teaching in the form of project background knowledge, forming distinctive teaching. For example, when displaying ship position information, maritime foundational knowledge such as the AIS ship reporting system and BeiDou Navigation Satellite System positioning and navigation is incorporated. Based on the OBE (Outcome-Based Education) concept, interdisciplinary majors and integrated science-education projects are established. Production and research projects combining "big data + shipping" are designed as project-based tasks that run through the entire professional group curriculum system, reconstructing project-based teaching content and continuously iterating the development of teaching content. The project is broken down into several sub-modules (courses), integrated with unified technical standards. Teachers collaborate and implement project-based, modular teaching. A "dual mentorship" system is adopted between schools and enterprises, using a "mentor team + project" co-creation model between teachers and students. This

transforms teacher-centered "instructive" and "declarative" teaching into student-centered, project-based, integrated learning and exploratory collaborative learning, stimulating students' intrinsic motivation to learn and cultivating their innovative thinking, scientific abilities, as well as their capacity to integrate theory with practice and discover and solve problems.

5. Conclusion

The Automatic Identification System (AIS) for ships is a crucial component of water transportation perception data. AIS data analysis plays an irreplaceable role in ship collision avoidance, route optimization, maritime regulation, and port management, and is of great significance for enhancing water transportation safety and promoting the construction of smart maritime systems. Technologies such as data cleaning, data storage, data fusion, and data visualization collectively form the core technical system for AIS stream data processing, providing strong support for real-time monitoring and information services for ships.

Big data technology represents a new generation of information technology, and the establishment of big data technology majors began in 2019. Research integrating big data majors with industry characteristics is just getting started. Shipping is currently undergoing its fourth technological revolution, with the core being the application of new-generation information technologies, including big data technology. Based on the shipping industry, conducting research on the cultivation of big data technology talents will play a leading role in the construction of big data technology majors in vocational colleges with industry characteristics.

Acknowledgements

This work was financially supported by the funding of the Philosophy and Social Science Research Project of the Jiangsu Higher Education Institutions of China (2023SJYB0755), the Fundamental Computer Education and Teaching Research Project of the Association of Fundamental Computing Education of Chinese Universities (2024-AFCEC-416), and the Excellent Teaching Team for QingLan Project of the Jiangsu Higher Education Institutions of China (Big Data Technology Teaching Team with Shipping Characteristic).

References

[1] Aiello, Giuseppe, Antonio Giallanza, and Giuseppe Mascarella. "Towards Shipping 4.0. A preliminary gap analysis." Procedia Manufacturing 42 (2020): 24-29.

[2] Yang, Dong, et al. "How big data enriches maritime research–a critical review of Automatic Identification System (AIS) data applications." Transport Reviews 39.6 (2019): 755-773.

[3] Ribeiro, Claudio V., Aline Paes, and Daniel de Oliveira. "AIS-based maritime anomaly traffic detection: A review." Expert Systems with Applications 231 (2023): 120561.

[4] Lee, Hyeong-Tak, et al. "An AIS data-driven approach to analyze the pattern of ship trajectories in ports using the DBSCAN algorithm." Applied sciences 11.2 (2021): 799.

[5] Xiao, Ye, et al. "Adaptive multi-source data fusion vessel trajectory prediction model for intelligent maritime traffic." Knowledge-Based Systems 277 (2023): 110799.

[6] Yang, Cheng-Hong, et al. "AIS-based intelligent vessel trajectory prediction using bi-LSTM." IEEE Access 10 (2022): 24302-24315.

[7] Rong, H., A. P. Teixeira, and C. Guedes Soares. "Ship collision avoidance behaviour recognition and analysis based on AIS data." Ocean Engineering 245 (2022): 110479.

[8] You, Bin, et al. "Ship collision avoidance behavior analysis based on AIS data." 2021 6th International Conference on Transportation Information and Safety (ICTIS). IEEE, 2021: 787-793.

[9] Zhang, Jinfen, et al. "An interpretable knowledge-based decision support method for ship collision avoidance using AIS data." Reliability Engineering & System Safety 230 (2023): 108919.

[10] Liu, Zhao, et al. "A quantitative method for the analysis of ship collision risk using AIS data." Ocean Engineering 272 (2023): 113906.

[11] Vicente-Cera, Isaías, et al. "Use of AIS data for the environmental characterization of world cruise ship traffic." International Journal of Sustainable Transportation 14.6 (2020): 465-474.

[12] Asborno, Magdalena I., et al. "Inland waterway network mapping of AIS data for freight transportation planning." The Journal of Navigation 75.2 (2022): 251-272.

[13] Serra-Sogas, Norma, et al. "Using aerial surveys to fill gaps in AIS vessel traffic data to inform threat assessments, vessel management and planning." Marine Policy 133 (2021): 104765.

[14] Peddireddy, Kiran. "Streamlining Enterprise Data Processing, Reporting and Realtime Alerting using Apache Kafka." 2023 11th International Symposium on Digital Forensics and Security (ISDFS). IEEE, 2023: 1-4.

[15] Zafeirakopoulos, Alexandros Nikolaos, and Euripides GM Petrakis. "HYAS: hybrid autoscaler agent for apache Flink." International Conference on Web Engineering. Cham: Springer Nature Switzerland, 2023: 34-48.

[16] Liao, Xuehua, et al. "Redis-based full-text search extensions for relational databases." International Journal of Machine Learning and Cybernetics (2024): 1-17.

[17] Chae, Sangeun, et al. "Optimizing Read Performance of HBase through Dynamic Control of Data Block Sizes and KVCache." Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing. 2024: 1495-1503.