ISSN: 2093-5374 / E-ISSN: 2093-5382

Application of Big Data Analysis in Chinese Art Song Market Research

Meng Deng[1*]

[1*]Ph.D, Art Department, International College, Krirk University, Bangkok, Thailand.
dengmeng202208@163.com, https://orcid.org/0009-0000-8623-0699

Received: July 05, 2024; Revised: August 20, 2024; Accepted: September 16, 2024; Published: December 30, 2024

Abstract

Data reliant on fisheries are essential for successful managing fisheries. A lack of fisheries data results in clarity regarding stock status, thereby jeopardizing the financial and food safety of the consumers reliant on that stock and heightening the risk of exploitation. Recent advancements in technology for gathering, managing, and evaluating fishery-related data offer a range of options to enhance and modernize fisheries' information systems, significantly increasing data collecting and analytical capabilities. Notwithstanding the widespread availability of pertinent consumer electronics, incorporating sophisticated information systems into fisheries oversight is still atypical rather than the norm. The article delineates the present condition, obstacles, and prospective trajectories of advanced information systems in the fishing industry to elucidate the factors hindering their acceptance. Our examination of the utilization of fishery-dependent information technologies across several worldwide fisheries sectors reveals that technological progress is stalling due to a deficiency of trust and collaboration between fishers and management. The research advocates for a solution grounded in a cross-disciplinary approach to fisheries management that underscores the necessity of working together to solve problems among participants. In our suggested structure, data suggestions are essential for efficient fishing information systems, ensuring fishermen and managers acquire, access, and profit from data related to fishing while striving towards a shared objective. An innovative strategy for fisheries information systems will enhance data protection, precision, and settlement while minimizing costs and facilitating adaptive, adaptable, near real-time managing decisions to optimize fisheries performance.

Keywords: Big Data, Market Analysis, System Algorithms, Music Recommendation Systems.

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 15, number: 4 (December), pp. 1-10. DOI: *10.58346/JOWUA.2024.I4.001*

*Corresponding author: Ph.D, Art Department, International College, Krirk University, Bangkok, Thailand.

1 Introduction

Observing maritime areas is essential for safeguarding the marine environment, overseeing vessel activities, and enhancing supply networks within the marine industry (Alqurashi et al., 2022). Marine traffic information is necessary for monitoring maritime areas. Examples encompass Vessel Management Services (VMS) (Affandi & Sumpeno, 2020) and Automatic Identifying Service (AIS) data (Chen & Wu, 2022). Either is extensively utilized in measurement controlling and supervision (MCS) systems locally and worldwide (Asiaei et al., 2020). VMS is an online fishing platform that enables bodies to follow and oversee the operations of vessels. The objective is to enhance the oversight and long-term viability of the sea ecosystem by implementing appropriate learning and curbing illicit fishing activities. Additional inputs of relevant data include meteorology (e.g., rainfall and velocity), ecological (e.g., ocean temperatures and greenary concentrations), and topographical (e.g., shoreline and bathymetric) information.

Efficient fisheries management is becoming increasingly crucial as overfishing endangers global fish populations, diminishes biodiversity, disrupts ecosystem functionality, and affects millions' food safety and livelihoods (Elisha & Felix, 2020). Fisheries administration is a multifaceted socio-political endeavor, and there is no singular solution to enhance global fishing. Access to precise and consistent information regarding the status of a fishery and the quantity, location, and methods of species capture is an essential element for implementing operational fishery administration, irrespective of the fishing industry or management framework (Jaya et al., 2022).

Accelerating alterations in oceanic conditions resulting from human-induced global warming and climate variability necessitate acquiring more accurate time and space data on fishing over reduced timescales to mitigate increasing uncertainty regarding the state of stocks and enable managers to modify points of reference in response to environmental changes (Brodie et al., 2021). This necessitates enhanced data collection, expedited and sophisticated reporting, preparation, evaluation, and better techniques for disseminating results to provide near real-time reactions. The ecological traits of fished shares complicate data collection, rendering it arduous and expensive due to their near disappearance in the seas, extensive shipping, accessibility across jurisdiction borders, and complicated relationships between aquatic ecosystems and their physical surroundings.

2 Related Works

Recent advancements and emerging technologies—frequently utilizing the omnipresence of mobile devices and the increasing availability of cloud-based computing for storing data, alongside artificial intelligence for analysis—possess the capacity to enhance fishery-dependent information systems by augmenting or optimizing gathering information, performing and encouraging data processing and evaluation, and enabling the dissemination of outcomes to the appropriate parties (Tagliarolo et al., 2021).

An alternative method for tackling the Future Location Prediction (FLP) issue involves machine learning techniques. Ibadurrahman et al. simulate the route of maritime boats and offer a service that forecasts the position of any specific vessel in near-real time using multiple layered perceptrons (MLPs) (Ibadurrahman et al., 2021). Yang et al. provide a machine learning (ML) approach that utilizes geographic time-series monitoring information from maritime boats to forecast potential paths based on real-time parameters (Yang et al., 2022). Various ML methods were evaluated for regression precision, quantified by mean absolute error, and processing duration. The perceptron was chosen because of its superior performance compared to all the other methods. Wang et al. offer a predictive method that generates the future path sequences of nearby boats in real-time using a Long Short Term Memory (LSTM) encoding-decoding design (Wang et al., 2020).

Technologies can enhance the dissemination and accessibility of knowledge for fishermen, enabling them to improve their fishing practices according to the most pertinent details and converting unidirectional communication (from fishermen to supervisor) into a collaborative, mutually advantageous cycle of gathering data, synthesizing, and collaborating (Eck et al., 2024). Fishermen, reliant on comprehending the ocean's movements for their livelihoods, are typically the initial observers of marine environment alterations, rendering the many vessels out of the sea the primary line of defense in monitoring shifting stock patterns and environmental fluctuations (Kroeker et al., 2020). Despite the abundance of efficient, affordable solutions and the capacity of modern technology to meet information needs in the fishing industry, the use of advanced fishery-dependent information systems is still the case and not the norm.

3 Big Data-related Research Methods

The research employs a lambda design that integrates sequential processing with data streams in a hybrid approach to manage the big data pertinent to proposed use case situations and to enable diverse processing and analytical tasks (Park et al., 2023). The objective of this strategy is to equilibrate latency, productivity, and tolerance for errors. This is accomplished using sequential processing for extensive past information analysis while concurrently employing streaming processing for immediately available insights. The two perspectives are combined before the exposition.

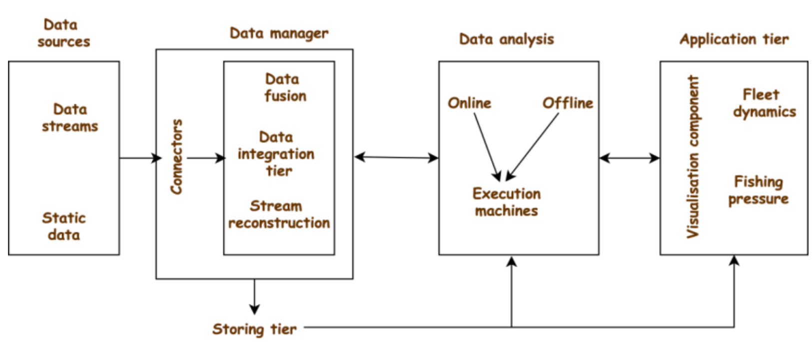

Figure 1: Architectural Design of the Proposed System

Figure 1 depicts the comprehensive design of the extensive data system. The design has six modules that handle and utilize both batches and online information to fulfill the objectives of maritime area monitoring and vessel traffic evaluation (Chen et al., 2020). The platform's input data point is the Supervisor, which gathers, integrates, and enhances the information that arrives. The Data Manager's result is recorded in permanent memory and then sent to online statistical analysis modules. The Storage Tier is tasked with the retention of generated data. The Offline Analytics component executes the necessary batch statistics. The data generated by this component is saved in the Storing Tier and employed by various web analytics programs afterward. The Online Analysis component obtains data from the receiving flows generated by the Data Management and the Storing Tiers, which houses the results of the Offline Analysis. The Application Tier component comprises numerous programs that leverage the proposed system to simulate fleet motion, assess fishing stress, and detect and forecast vessel operations, among other functions. The Application Tier encompasses visualization techniques that render data according to various application contexts.

All these components are essential for fulfilling the programs' needs. The primary objective of the proposed structure is the cooperation and amalgamation of databases of information, that includes the unification of vessel movement information with open physical ecological, and weather information, as well as the combination of traffic information from various sources (VMS and AIS information) about the same organizations (vessels). A further objective is to devise and execute techniques for enlargement, synopsis extraction, and identification of linguistically enriched paths from various movement information. The demands are executed within the Managiing Tier. A significant objective of the proposed framework is examining integrated traffic information, which seeks to formulate methodologies and instruments for analyzing real-time and past vessel traffic information.

A significant objective is the forecasting of fisheries vessel movements and tasks, which focuses on the formulation and advancement of methods for predictions about setting, schedules, and operations of fishing boats. The application must be implemented via the Data Analysis Component (offline and on the Internet). Lastly, a crucial application needed is the simulation of fleet motion, which focuses on using approaches for fleet movement models and calculating fishery stress. Both application situations are executed via the Application Tier.

Recommendation Algorithm Based on Collaborative Filtering

In recommender systems, collaborative filtering algorithms are one of the most common and mature technologies, mainly predicting items users are interested in based on their historical behavioral data. The community has conducted extensive, in-depth research on collaborative filtering. It has developed various methods, including neighborhood-based recommendation, hidden semantic model-based, and graph-based random wandering algorithms. Neighborhood-based recommendation methods have gained wide application and recognition in the industrial world due to their ease of understanding and implementation. The neighborhood-based recommendation can be further subdivided, and the specific procedures of the two methods are shown in Figure 2. It recommends items liked by other users with preferences for the target user. In contrast, other model recommends new items identical to those the target user liked in the past. These two approaches capture user preferences from different perspectives and provide personalized recommendations to users.

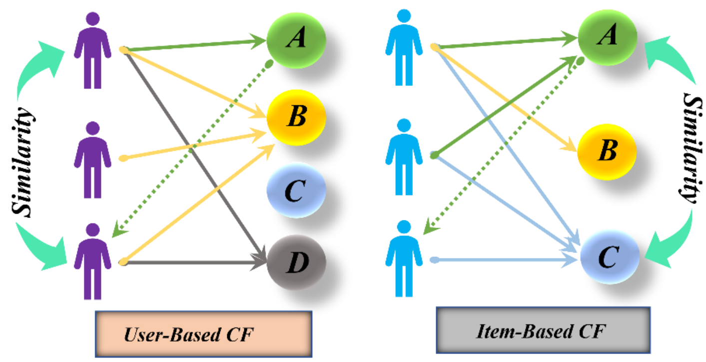

Figure 2: Two Collaborative Filtering Algorithms Recommendation Process

As shown in Figure 2, if user one purchases items A, B, and D, and user three purchases B and D, user one and user 3 have a high degree of similarity based on the analysis of purchasing behavior. Because user 1 purchased item A, it can be presumed that user three is interested in A. Therefore, the system provides A as a recommended item to user 3. This process embodies the mechanism of user-based CF, which aims to achieve personalized recommendations through the similarity between users. Item-Based CF indicates that user one and user two purchased items A and C, then A and C are considered similar; when user three purchases item C, the system will recommend item A to user three because users tend to purchase similar items based on historical data.

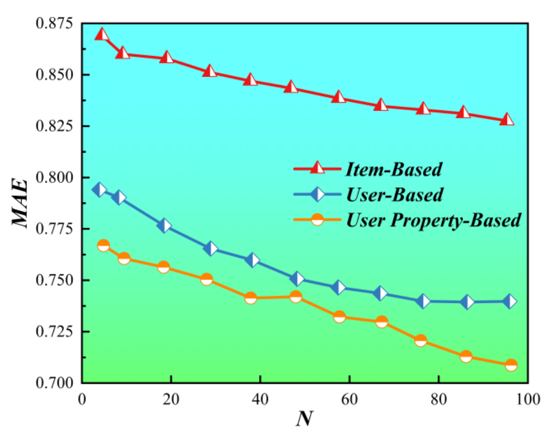

Figure 3: Comparison of MAE Values of Different Models

Such a recommendation process not only improves efficiency but also enhances the interpretability of the recommendation. The MAE values of the three models are plotted by taking out the N nearest users for prediction scoring and calculating the MAE values, as shown in Figure 3.

Spark Algorithm

Spark is a high-speed, multi-functional big data computing platform developed using Scala programming. It dominates distributed computing scenarios with its excellent fault-tolerance mechanism, high efficiency, and scalability, and is widely used to process large data sets. Spark's ease of use allows developers to get started quickly, significantly improving the programming efficiency, and the corresponding ecosystem is shown in Figure 4. Spark Core assumes the core responsibility of the framework and provides essential functional support; Spark SQL focuses on the processing of structured data and provides a convenient programming interface for programmers; Spark Streaming allows for analysis of data streams, ensuring the Spark MLlib is a built-in ML library that simplifies the implementation of ML tasks. Graphx is a distributed framework specifically designed for graphical data processing, making the operation of graphical data more efficient. Spark not only carries the distributed parallel processing capabilities of Hadoop MapReduce but also targets MapReduce's shortcomings. Spark not only carries the distributed parallel processing capability of Hadoop MapReduce but also optimizes for MapReduce's shortcomings. It reduces I/O operations on the HDFS file system by caching intermediate computation data in memory, thus increasing the speed of iterative operations. As the core of Spark's in-memory computation model, the Resilient Distributed Dataset (RDD) is a collection of distributed data, which reduces the cost of fault-tolerance mechanisms and accelerates the data retrieval process thanks to its caching capabilities. Spark performs better when dealing with various workloads with RDD.

4 Results and Discussions

Experimental Analysis

The studies were performed on a 10-node group utilizing Hortonworks Engine, with every element equipped with 8GB of RAM and 16-core processors. The storing component is integrated into the cluster, offering a cumulative capacity of 3.3 TB divided across seven nodes. The Spark execution machine is implemented over seven nodes, with Apache Hadoop YARN as the asset management.

The study employs two years of information from boats operating in seas. The mobility statistics are controlled information from the Department of Fisheries, Ministry of Marine, and isolated Rules. The dataset comprises over 50 million location records from over 1,000 fishing boats, using roughly 1GB storing format (gzip). The research utilized a more comprehensive collection of ground AIS information from the Directorate of Marine Security, Ministry of Naval Affairs, and Insular Policies of Greece. The AIS collection comprises around 3.8 billion positional data from over 10,000 boats, using 60GB of compacted (gzip) storage capacity. The research employed meteorology re-analysis data from the Centre of Meteorological Forecasting data site, including the maritime region of Greece, throughout the identical two-year timeframe. The data contains hourly readings for 11 critical metrics about maritime meteorological conditions. Supplementary data utilized chiefly for data enhancement encompasses a bathymetric spatial pixel database alongside polygonal shapes delineating marine areas established by fishing restrictions. These were analyzed and delivered to the consortium by the Centers for Ocean Studies, an investor in the proposed work.

Results

The trials were conducted on Apache Spark Streams utilizing four distinct configurations. The research mainly tested the trigging period choice assigned to 1 to 8 minutes, indicating that the portable data enhancement procedure is activated whenever a collection of information with lengths of 1 to 8 minutes is gathered. For each configuration, the research quantified the median duration required by Apache Spark Broadcasting to go through the input volumes and generate enhanced mobility information. The outcomes of the study are displayed in Table 1.

Table 1: Average Delay Analysis

Triggering period	Noise free information	Synopses
1	28.32	2.31
2	33.4	5.33
3	44.76	5.94
4	49.25	10.46
5	56.27	13.2
6	74.79	17.52
7	96.79	20.9
8	121.27	25.11

Figure 4 illustrates an example of hotspot analysis used in AIS information. The analytical variables include trawler as the vessel category, fisherman as the action, a boundary box encompassing the geographical spectrum, and February 2023 for the duration spectrum. The cell sizes are 0.02 levels in longitude and 0.02 in latitude, encompassing the whole chronological range in the axis. The boundary is established at 2. It is logical for the lengths of the two coordinates. Due to a singular timespan in the sense of time, every cell lacks adjacent cells in this area. The top 2000 units are presented. Red squares, cold spots, and tissues by light blue squares represent locations. As spots are depleted before reaching the 3k boundary, many neutral units are exhibited alongside all the places.

Figure 4: Hospost Analysis

Regarding the testing stage, utilizing the proposed design enables forecasting the vessels' subsequent positions through an online decentralized process. Simultaneous streaming data from several vessels is transmitted to the system, which a) generates forecasts concurrently and b) disseminates the latest data in a dispersed streamed manner to other tools inside the proposed system.

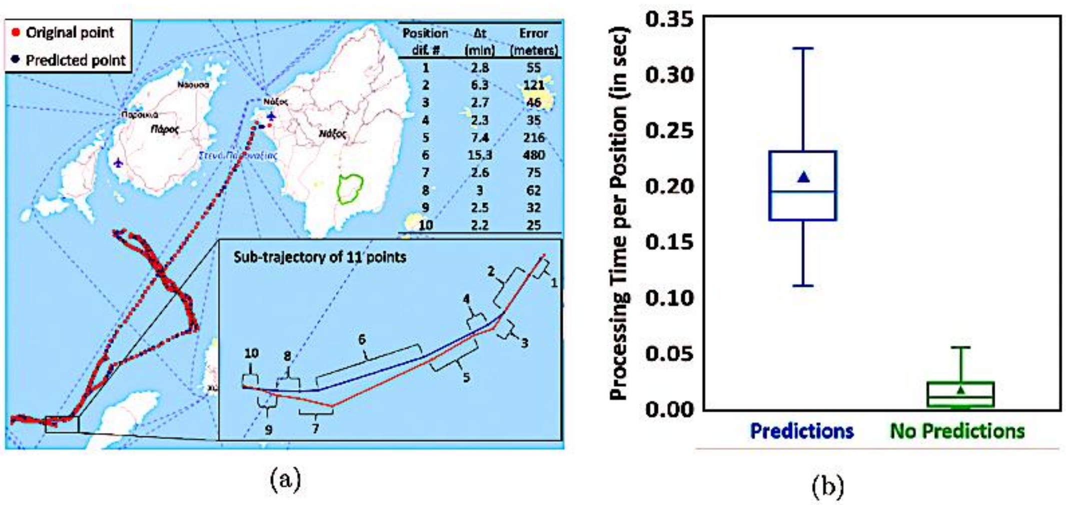

Figure 5: Processing Time Analysis

The study outcomes were assessed using the Mean Absolute Errors (MAE) based on the length among the source locations and the anticipated points in the measurement group. The findings indicate that the LSTM algorithm accurately predicts the subsequent location of fishing boats. The MAE for the forecast intervals of 2 to 30 minutes in the experiment set are 50, 100, 250, 500, and 1000 inches, respectively. Figure 5(a) illustrates the original positions of a boat used for fishing with the anticipated sites for each projection range. Figure 5(b) illustrates the temporal effectiveness of the LSTM-based FLP on the idea system, represented as time spent processing per prediction.

5 Conclusion

The expanding human population is escalating the need for seafood, making it essential to ensure equitable utilization of resources in all areas of fishing for seafood to protect food and livelihood stability and preserve the seas' natural integrity. An exhilarating opportunity exists to leverage technology to enhance fisheries information systems significantly. Innovative technologies can enhance collecting, analyzing, and dissemination to promote environmentally friendly use of resources by equipping participants with geographically and chronologically pertinent information for fishing and managing fisheries. The current application of technology in the fishing industry underscores its considerable potential to enhance outcomes across various fishing industries; the restricted creativity and widespread acceptance of novel technologies indicate that substantial challenges have hindered support for technological advancements in many fisheries around the globe. The research proposes integrative fishing governance to enhance the acceptance and adoption of novel fishery-dependent information technology. Through collaborative efforts towards a common objective, fisheries managers can leverage fishery-dependent technological innovations as a compelling proposition for fishers. By utilizing tools that simplify and automate gathering and analyzing information and providing users with insights on optimal fishing locations and enhanced market access, implementing novel data management systems can improve fishery outcomes for many different groups.

The research recognizes that establishing enhanced data streams entails problems, including an overload of information and excessive faith in models and their corresponding analytics. Information alone will not yield sustainable aquaculture, nor does it inherently facilitate improved decision-making; instead, it is an essential element of good management across all fishing sectors and management institutions. Adopting technical developments in fisheries information systems can result in mutually beneficial circumstances where managers and harvesting machines achieve enhanced fishing results through superior data.

References

[1] Affandi, A., & Sumpeno, S. (2020). Clustering spatial temporal distribution of fishing vessel based LON VMS data using K-means. *In IEEE 3rd International Conference on Information and Communications Technology (ICOIACT)*, 1-6.

[2] Alqurashi, F. S., Trichili, A., Saeed, N., Ooi, B. S., & Alouini, M. S. (2022). Maritime communications: A survey on enabling technologies, opportunities, and challenges. *IEEE Internet of Things Journal, 10*(4), 3525-3547.

[3] Asiaei, K., Bontis, N., & Zakaria, Z. (2020). The confluence of knowledge management and management control systems: A conceptual framework. *Knowledge and Process Management, 27*(2), 133-142.

[4] Brodie, S., Abrahms, B., Bograd, S. J., Carroll, G., Hazen, E. L., Muhling, B. A., & Jacox, M. G. (2021). Exploring timescales of predictability in species distributions. *Ecography, 44*(6), 832-844.

[5] Chen, M. Y., & Wu, H. T. (2022). An automatic-identification-system-based vessel security system. *IEEE Transactions on Industrial Informatics, 19*(1), 870-879.

[6] Chen, X., Xu, X., Yang, Y., Wu, H., Tang, J., & Zhao, J. (2020). Augmented ship tracking under occlusion conditions from maritime surveillance videos. *IEEE Access, 8,* 42884-42897.

[7] Eck, K., Soomai, S. S., & MacDonald, B. H. (2024). On the development of criteria for determining the effectiveness of technical working groups: A case study about information processes in marine conservation and fisheries management in Belize. *Environmental Science & Policy*, *160*, 103863. https://doi.org/10.1016/j.envsci.2024.103863

[8] Elisha, O. D., & Felix, M. J. (2020). The loss of biodiversity and ecosystems: a threat to the functioning of our planet, economy, and human society. *International Journal of Economics, Environmental Development and Society, 1*(1), 30-44.

[9] Ibadurrahman, Hamada, K., Wada, Y., Nanao, J., Watanabe, D., & Majima, T. (2021). Long-term ship position prediction using automatic identification system (AIS) data and end-to-end deep learning. *Sensors, 21*(21), 7169. https://doi.org/10.3390/s21217169

[10] Jaya, I., Satria, F., Nugroho, D., Sadiyah, L., Buchary, E. A., White, A. T., & Green, S. J. (2022). Are the working principles of fisheries management in Indonesia? *Marine Policy, 140,* 105047. https://doi.org/10.1016/j.marpol.2022.105047

[11] Kroeker, K. J., Bell, L. E., Donham, E. M., Hoshijima, U., Lummis, S., Toy, J. A., & Willis-Norton, E. (2020). Ecological change in dynamic environments: Accounting for temporal environmental variability in studies of ocean change biology. *Global Change Biology, 26*(1), 54-67.

[12] Park, S., Yang, C. S., & Kim, J. (2023). Design of vessel data lakehouse with big data and AI analysis technology for vessel monitoring system. *Electronics*, *12*(8), 1943. https://doi.org/10.3390/electronics12081943

[13] Tagliarolo, M., Cope, J., & Blanchard, F. (2021). Stock assessment on fishery-dependent data: Effect of data quality and parametrization for a red snapper fishery. *Fisheries Management and Ecology, 28*(6), 592-603.

[14] Wang, Z., Su, X., & Ding, Z. (2020). Long-term traffic prediction based on lstm encoder-decoder architecture. *IEEE Transactions on Intelligent Transportation Systems, 22*(10), 6561-6571.

[15] Yang, Y., Shao, Z., Hu, Y., Mei, Q., Pan, J., Song, R., & Wang, P. (2022). Geographical spatial analysis and risk prediction based on machine learning for maritime traffic accidents: A case study of Fujian sea area. *Ocean Engineering*, *266*, 113106. https://doi.org/10.1016/j.oceaneng.2022.113106