Innovation in Smart Logistics Decision - Making Driven by Multimodal Data Fusion

Yishu Wang[a*], Minghai Lv[b*]

College of Business Administration, Liaoning Technical University, Huludao, 125105, China

[a*]1361149321@qq.com, [b*]845721911@qq.com

Abstract. In the wave of digitalization, the logistics industry faces the dual challenges of business growth and cost control. This article focuses on the application of multimodal data fusion in smart logistics decision-making, analyzes the principles of how it optimizes decision-making, plans logistics routes, and improves resource utilization efficiency. Through case studies of enterprises such as JD.com, it demonstrates the achievements and confirms that multimodal data fusion can enhance efficiency, reduce costs, and increase competitiveness. At the end of the article, it looks ahead to future trends such as technological innovation and scenario expansion, providing impetus for the development of smart logistics.

Keywords: Multimodal Data Fusion; Smart Logistics Decision - Making; Logistics Efficiency Improvement

1 Introduction

In the current era of rapid transformation in information technology and logistics, the emergence of multimodal data has become an important development trend. Logistics firms face cost and service pressures, and multimodal data fusion is a solution[1]. This forward - looking perspective will help the logistics industry anticipate changes, address new challenges, and fully tap the potential of multimodal data fusion to drive the continuous innovation and development of the smart logistics field.

2 Theoretical Basis

2.1 Concepts and Types of Multimodal Data Fusion

Multimodal data fusion involves integrating data from various sources, sensors, or modalities to produce more comprehensive, accurate, and reliable information. In smart logistics, this process is vital for enhancing decision-making.

For example, combining GPS data for location tracking, temperature sensor data for cold chain monitoring, and video surveillance data provides a holistic view of logistics

A. Mallik et al. (eds.), *Proceedings of the 2025 6th International Conference on Management Science and Engineering Management (ICMSEM 2025)*, Atlantis Highlights in Economics, Business and Management 4,
https://doi.org/10.2991/978-94-6463-845-5_12

operations. This allows managers to optimize routes, ensure proper handling of goods, and quickly address any issues.

Different fusion methods, such as feature-level, decision-level, and data-level fusion, each have unique characteristics. Feature-level fusion combines extracted features, decision-level fusion combines model outputs, and data-level fusion combines raw data. Each method can be tailored to specific logistics needs, contributing to more robust and efficient decision-making.

2.1.1 Statistical Methods.

Bayesian estimation is a sensor data fusion method that provides reliable posterior distributions by integrating prior knowledge with new evidence. In logistics, it predicts goods' arrival times by combining historical and real-time traffic data, enabling more accurate scheduling.

Kalman filtering is a robust recursive algorithm that fuses vehicle and map data to pinpoint vehicle positions and statuses, even in interference zones[2]. It continuously updates the vehicle's location based on its movement and map data, ensuring precise tracking and control.

Both methods enhance the accuracy and reliability of data fusion in logistics, improving operational efficiency.

2.1.2 Artificial Intelligence Methods.

Neural networks have emerged as highly potent AI methods for multimodal data fusion. Various models within the neural network realm, such as multi-layer perceptrons and convolutional neural networks, possess remarkable capabilities in extracting and fusing features from diverse data sources. These sources can encompass a wide range of modalities, including but not limited to data from sensors, images, and texts. In the specific context of logistics goods classification, the application of convolutional neural networks (CNNs) is particularly noteworthy. CNNs are adept at extracting intricate and meaningful image features from the visual representations of goods. They can also effectively process and incorporate textual information related to the goods. By seamlessly fusing these extracted image features with the textual data, CNNs achieve highly accurate classification of logistics goods. This fusion process not only enhances the overall accuracy of the classification but also significantly boosts the efficiency of the entire multimodal data fusion operation. The ability to integrate and leverage information from multiple modalities in such a sophisticated manner underscores the immense potential of neural networks in advancing the field of logistics and supply chain management.

2.2 Technology Integration and Decision-making Algorithms

In intelligent logistics, multiple techs collaborate. Sensor fusion gathers sensor data; in logistics vehicles, GPS locates, IMU adjusts attitude, lidar and cameras sense the environment for an accurate transport model. Stream processing with Flink and Storm detects anomalies in real-time logistics data. CEP flags goods detention and vehicle deviation. Data integration employs a federated database for enterprise data to aid decision-making. For decision-making, linear and integer programming focus on logistics routes and vehicle scheduling, using cost and time to model for optimal solutions[3]. Genetic and ant colony algorithms mimic biology to solve route and location issues. Q-learning and DQN let the system learn from real-time info to optimize transport and inventory.

3 Application of Multimodal Data Fusion in Smart Logistics

3.1 Smart Practices of JD Logistics

This section examines JD Logistics' notable achievements in in-warehouse automation and delivery optimization through multimodal data fusion. JD Logistics has effectively integrated data from various sources, such as warehouse management systems, AGVs, drones, and real-time traffic information.In in-warehouse automation, JD Logistics uses AGVs that navigate with precision, thanks to fused sensor and mapping data. Drones also assist in inventory checks, enhancing accuracy and speed. For delivery optimization, JD Logistics employs algorithms that combine real-time traffic data, historical patterns, and customer locations to optimize routes and schedules dynamically.These efforts highlight how multimodal data fusion can significantly enhance logistics operations, setting a high standard for the industry

3.1.1 In-warehouse Robot Technology.

The in-warehouse robots of JD Logistics skillfully utilize multimodal data to improve efficiency. They collect the appearance images of goods through vision, and combine with detailed text attributes, such as weight at the gram level and fragility level of 1 - 5, to help robots select handling and storage strategies according to the characteristics of the goods(see table 1).

Table 1. Practices of Multimodal Data Facilitating Efficiency Enhancement in Logistics Scenarios of Three Cities

City	Application Scenario	Application of Multimodal Data	Outcome
Beijing	Handling Electronic Display Screens	Visually collect appearance, combine 5kg weight & 5-level fragility text, then slowly move to buffer & humidity-controlled area..	Ensure the safety of the goods.
Shang-hai	Warehousing during	Visually collect data and combine it with refined textual attributes. The	The storage efficiency reaches 6

City	Application Scenario	Application of Multimodal Data	Outcome
	"Double 11"	monitoring accuracy of temperature and humidity is ±0.5°C and ±3%.	times that of manual labor.
Guang-zhou	Daily Ware-housing	Visually capture goods state, handle & store per 300g weight, 4-level fragility text. Control temp & humidity.	Avg storage eff is 5.2 times manual.

3.1.2 Applications in the Distribution Link.

JD Logistics uses multimodal fusion to boost services. It combines 2 - min traffic updates, 5 - m positioning & 1 - yr delivery data for route planning[4]. In 618 Beijing, it aided 50,000 + employees, dodged 200 + jams, cut delivery time by 1.5 hrs and loss rate to 0.3% from 0.8%. Using 3 - yr shopping & location data, it has an “front - store, back - warehouse” model. In SZ & HZ, 80% of orders arrive within 2 hrs. Its digital twin team monitors chains and offers 20 + aids, achieving great results and setting a benchmark.

3.2 Cainiao's Smart City Distribution Logistics System

3.2.1 Cognitive Decision-making Technology of Logistics Multimodal AI.

Cainiao incorporates the "logistics multimodal AI cognitive decision-making tech" throughout to boost user experience. In a Hangzhou community, the "building code" tech combines multimodal data. Image recognition measures parcel size (error < 5%) and shape, and with precise house number and 0.2 kg weight info, it aids sorting, hiking delivery efficiency per person by 10.5%. The "real door-to-door performance recognition model" uses AI for real-time monitoring and, combined with over 80% feedback recovery rate, lifts Hangzhou's customer satisfaction from 85% to 92%.

3.2.2 Solutions for Cost Reduction and Efficiency Improvement.

The smart city distribution logistics system of Cainiao achieves cost reduction and efficiency improvement through the fusion of multimodal data(see table 2).

Table 2. Achievements of Multimodal Data Application in Key Logistics Links

Application Link	Multimodal Data	Achievements
Order Allocation	Two-year order data, community coordinates (error < 10 meters), traffic information updated every 3 minutes	The average delivery time in Nanjing is reduced by 1.2 hours, and the cost is reduced by 18%.
Intelligent Packaging	Images with 15 million pixels, dimensions with an accuracy of	The loading rate of vehicles from Beijing to Shanghai is

	0.5 cm, weight with an accuracy of 0.1 kg	increased to 85%, and the cost is reduced by 15%.
Logistics Monitoring	More than 20 parameters such as temperature within ±0.4℃, humidity within ±3% + images	The probability of goods delivery reaches 99.5%.
Network Layout Optimization	Big data + AI analysis of multimodal data	The average distance between nodes in 10 cities is reduced by 10%, and the resource utilization rate is increased by 20%.

4 Addressing Multimodal Data Fusion Obstacles & Inherent Limits

4.1 Solutions for Overcoming Data Heterogeneity

In multimodal logistics, data types like images, text, and sensor data are structurally and semantically diverse. Direct fusion is difficult, as combining, say, cargo image features with temperature values yields little value. Finding an association model for such varied data is challenging.

To tackle this, techniques such as Multimodal Adversarial Networks (MAN) can be used for adversarial training to align features in a shared latent space. Alternatively, building a knowledge graph to link cold chain data and preservation text can strengthen the fusion basis.

4.2 Breakthroughs in Model Interpretability

Deep-learning multimodal fusion models are like “black boxes”. Logistics enterprises struggle to understand their decision-making, such as the unclear basis of GNN logistics route optimization models. Use Graphviz to display learning dynamics[5], extract rules with decision trees, and generate “if-then” transport rule sets based on cargo weight, distance and weather to boost model transparency and practicality.

4.3 Strategies for Ensuring Real-time Performance

The logistics industry requires high real-time performance, and quick decision-making is crucial for multimodal data fusion. However, large HD images can slow down model inference and cause delays.

To solve this, we can use 5G for fast transmission of big modal data. Also, edge devices can preprocess and initially fuse data near the source, reducing the data to be transmitted and processed centrally. Additionally, developing incremental learning algorithms to fine-tune logistics forecast models with new orders ensures timely and accurate forecasts.

5 The "Intelligent Transformation" in the Logistics Field

Multimodal data fusion has significantly impacted the logistics field. It has shifted employment from basic manual jobs to emerging roles requiring digital and intelligent skills. This technology also standardizes data formats, enhances system security, and boosts supply chain collaboration[6]. Additionally, multimodal data fusion enhances risk response capabilities by enabling early detection of potential disruptions through integrated data from various sources like weather forecasts and traffic conditions. This allows for proactive adjustments in transportation routes, inventory management, and delivery schedules, reducing the impact of unforeseen events. Overall, multimodal data fusion boosts logistics operations' efficiency and reliability, preparing the industry to meet future challenges with advanced analytics and diverse data integration.

6 Conclusion

This article examines the application of multimodal data fusion in smart logistics decision-making, highlighting its crucial role in enhancing the logistics industry. By integrating diverse data types from multiple sources, multimodal data fusion provides a comprehensive view that significantly improves operational efficiency, reduces costs, and boosts competitiveness. Case studies of leading enterprises like JD Logistics and Cainiao demonstrate these benefits, showing how this technology optimizes warehouse automation, delivery routes, and supply chain visibility. Despite challenges such as data heterogeneity, model interpretability, and real-time performance, advanced algorithms like Bayesian estimation and Kalman filtering are being developed to address these issues. Looking ahead, multimodal data fusion is expected to drive the intelligent transformation and innovation in the logistics field, making it an essential tool for companies to stay competitive in a rapidly evolving market.

References

1. Guo Jian, Liu Hua, Zhao Shouyang. The Future Trends of Expressway Inspection — Multimodal Data Fusion and Analysis Based on Artificial Intelligence[J]. Transportation Science & Technology and Management, 2024, 5(24): 1-4.
2. Zhang Hucheng, Li Leixiao, Liu Dongjiang. A Review of Multimodal Data Fusion Research[J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(10): 2501-2520.
3. Fan P ,Wu Q .Advances in computer AI-assisted multimodal data fusion techniques[J].Applied Mathematics and Nonlinear Sciences,2024,9(1):
4. Feng T ,Mao X .Multimodal Data fusion for SRGPS antenna motion error reduction[J].Multimedia Tools and Applications,2017,76(9):12035-12050.
5. Simanek J ,Kubelka V ,Reinstein M .Improving multi-modal data fusion by anomaly detection[J].Autonomous Robots,2015,39(2):139-154.
6. Du Pengju. Research on Multimodal Data Fusion and Retrieval Technology[J]. Automation & Instrumentation, 2023, (10): 23-26. DOI: 10.14016/j.cnki.1001-9227.2023.10.023.