**Research on the construction of dynamic decision-making model for complex business systems based on multi-modal data fusion**

Xiyu Wei1,a, Kaining Shen , Yitong Wang\*

1，The University of Texas Health Science Center ，School of Public Health ， Houston, 77030, Texas, United States

2，Amazon，425 106th Ave NE, Bellevue, WA 98004 USA

3，Business intelligence consulting department, Shanghai Yinfinity Co., Ltd. 200041, Shanghai,china

aEmail： [weixiyu2025@126.com](mailto:weixiyu2025@126.com)

Emai：[k.shen.5574@westcliff.edu](mailto:k.shen.5574@westcliff.edu)

\*Email：wangyt0115@163.com

Abstract: To address the challenges of integrating multimodal data and dynamic response in complex business system decision-making, this study develops a framework for multimodal data fusion and a dynamic decision-making model. By integrating heterogeneous data such as text and images, the study designs mechanisms for data preprocessing, feature fusion, and timeliness updates. By combining system dynamics with reinforcement learning, it constructs a closed-loop model of 'input-computation-feedback.' This research provides methodological support for real-time decision-making in high-dimensional and dynamic business environments, expanding the application of multimodal data in areas such as consumer analysis and trend forecasting.

Key words: multimodal data; business system; dynamic decision

**foreword**

Traditional single-modal models struggle to handle high-dimensional heterogeneous data and the nonlinear interactions within systems. This study focuses on the integration of multi-modal data and the development of dynamic decision-making models, aiming to address the dimensional limitations and response delays of traditional models through cross-modal feature alignment, time management, and closed-loop modeling. It provides adaptive decision-making tools for supply chain optimization and market trend prediction in enterprises.

**1. Basic concepts and technologies of multimodal data fusion**

(1) Definition and characteristics of multimodal data

Multimodal data fusion is an interdisciplinary frontier that leverages information from diverse sources, such as text, images, audio, and video, to extract more valuable insights. Text data, sourced from social media comments, industry reports, and corporate documents, is rich in meaning but lacks structural organization. Image data, including product images, surveillance footage, and satellite remote sensing images, represents spatial information through visual features. Audio data, such as customer service call recordings and meeting transcripts, conveys emotional energy through voice patterns and intonation. Video data, which integrates images and audio, reflects the temporal dynamics of objective events and is used for consumer behavior analysis and market trend analysis. Each type of data provides unique insights from different perspectives, and this complementarity makes them suitable for various business decision-making scenarios. To achieve this, different modalities of data are transformed into a unified representation space to facilitate integration. The technical approaches for applying different modalities of data across various dimensions include data-level fusion, feature-level fusion, and decision-level fusion. Techniques like attention mechanisms and joint embedding learning can enhance the efficiency of cross-modal information fusion and improve the value of decision-making.

(2) Data fusion methods

In the field of multimodal data fusion, common methods can be categorized into early fusion, late fusion, and hybrid fusion based on the differences in processing stages and fusion levels. Early fusion, also known as data layer fusion, involves merging different modalities of data directly after collection without deep processing, followed by unified feature extraction and model training. This method maximizes the use of raw data information, reducing information loss, but it requires a high level of compatibility with data preprocessing. Late fusion, or decision layer fusion, involves independent feature extraction and model training for each modality, with the final results integrated through strategies such as voting or weighted averaging. This approach is advantageous for its flexibility in adapting to the characteristics of different modalities and its high computational efficiency, although it may result in insufficient information interaction due to independent processing in the early stage. Hybrid fusion combines the strengths of both early and late fusion, integrating some modalities at the feature layer while maintaining independent processing paths for other modalities. By adopting a hierarchical and phased integration strategy, it balances information integrity with computational complexity. In complex business system decision-making scenarios, hybrid fusion can dynamically adjust its integration strategy based on data characteristics and decision objectives, achieving efficient integration and utilization of multimodal information [1].

(3) Application of multimodal data in business decision making

Multimodal data holds significant potential in the decision-making of complex business systems, offering a new approach to accurately understand market dynamics and consumer behavior through its ability to integrate diverse information. In the field of consumer behavior analysis, text data can parse the emotional tendencies and pain points in social media comments and user feedback, while image and video data capture consumer shopping routes offline and online page heat maps, providing a clear view of behavior patterns and visual focal points. When combined with customer service tone and emotion from voice data, this can create a comprehensive consumer profile, uncovering potential preferences. For market trend prediction, the integration of multimodal data can combine industry report texts on policy and technology trends, satellite images and traffic logistics data reflecting regional economic activity, and news videos and public opinion data conveying social hotspots. By identifying nonlinear relationships between data, it can predict market supply and demand changes and emerging consumption trends in advance, helping companies dynamically adjust their strategic layouts[2].

**Second, analysis of dynamic decision-making problems in complex business systems**

A complex business system is a dynamic system with a highly complex structure and operational mechanism, characterized by diversity, nonlinearity, time variability, and uncertainty. Diversity refers to the system being a collection of various elements, including the company's production operations, financial management, human resources, and external factors such as market competition, changes in customer needs, and policy adjustments. Nonlinearity means that the causal relationships between components are not linear; a minor change at one point can affect the entire system or even the entire business chain. Time variability indicates that the overall environment and elements are constantly evolving, such as the cyclical fluctuations in market supply and demand, continuous technological advancements, and changing consumer attitudes, all of which can cause the system to transition from one state to another. Uncertainty arises from information asymmetry, making sudden events unpredictable, and any system can experience uncertain events.

In this complex and dynamic situation, the greatest challenge in dynamic decision-making is achieving timely and dynamic decisions in an uncertain and changing environment. Using a deterministic decision model to handle unstable information is extremely challenging. Traditional decision models are primarily based on known structured data and linear assumptions, which are not suitable for addressing the high-dimensional, heterogeneous, and complex issues in business systems. For example, when dealing with multi-modal data such as market sentiment information in text form, user behavior logs in image form, and time series sales data, traditional decision models cannot adequately integrate the vast amount of data with various relationships. Consequently, these models fail to accurately describe the relationships and extract valuable insights from the data, leading to biased decision outcomes and the inability to make correct or quick decisions[3].

**3. Design of multi-modal data fusion framework ​**

(1) Data acquisition and preprocessing ​

This study constructs a multi-source data ecosystem network that deeply integrates dynamic information from both internal and external sources of enterprises. Internally, it connects to the global supply chain ERP system to obtain high-precision sensor data in real time. For example, a multinational fresh food company has a cold chain logistics network with 5,600 temperature and humidity monitoring nodes, generating a set of time-series records every 30 seconds (with an average daily data volume exceeding 2TB). These records are also synchronized into the CRM system for customer complaint texts (accumulating 4.8 million entries annually). Externally, the data sources include satellite remote sensing and social media. The European Space Agency's Sentinel-2 satellite provides weekly 10-meter resolution images of ports, and during the 2022 Shenzhen epidemic lockdown, the container density at Yantian Port increased by 187% year-on-year. On social media, through the Twitter API, the frequency of keywords such as' supply cut-off 'and' out of stock' in public welfare topics surged by 320% over three days. To handle such heterogeneous data, the preprocessing phase innovatively implements a cross-modal spatiotemporal alignment strategy: for satellite images, an improved YOLOv5 model is used to identify container accumulation heat maps and dynamically correlate them with ERP logistics delay records (with a time error controlled within 15 minutes). For text data, a BERT word segmentation model enhanced by an industry-specific terminology database (F1 score 0.91) is used to extract supply chain disruption features, combined with the SnowNLP quantified panic sentiment index, ultimately establishing a coupling network of text sentiment, warehouse temperature, and transportation delay through a graph attention mechanism (Pearson correlation coefficient r=0.83, p<0.001).

The adaptability of preprocessing techniques in crisis response has demonstrated significant value. For instance, in the context of cold chain temperature and humidity time series data, using the Symlet5 wavelet basis function for noise reduction effectively filters out sensor errors (increasing SNR by 12.6dB). Additionally, Fourier transform is used to detect abnormal fluctuations, successfully warning of a sudden temperature spike in a cold chain warehouse due to power failure (47 minutes earlier than traditional threshold alarms). In particular, during the cross-modal fusion phase, the dynamic time warping algorithm aligns container stacking peaks from satellite images, supply chain disruption topics from social media, and enterprise ERP order delays, improving the accuracy of the fresh produce loss rate prediction model to 92.7% (an improvement of 18.4 percentage points over single-text analysis). This deep preprocessing based on real business scenarios lays the technical foundation for the collaborative application of multi-modal data in decision-making models [4].​

(2) Feature engineering and fusion strategy ​

In the framework of multi-modal data fusion, feature engineering and fusion strategies are crucial for integrating cross-modal information. Cross-modal feature alignment aims to eliminate differences in representation space between different modalities. Attention-based feature mapping effectively captures semantic associations between text, images, and other modalities by adaptively focusing on key information. Joint embedding learning maps multi-modal data into a unified low-dimensional space, enabling efficient integration of features from different modalities. In terms of fusion levels, data-level fusion combines raw data directly, preserving complete information but requiring high data compatibility and complex computation. Feature-level fusion concatenates or compresses features after extraction, balancing information loss and computational efficiency, making it suitable for scenarios with less structural variation in multi-modal data. Decision-level fusion independently processes each modality's data and integrates results at the model output stage, offering strong flexibility and scalability but potentially leading to insufficient interaction between modalities due to independent processing. Considering the diversity of data and the real-time requirements of dynamic decision-making in complex business systems, this study adopts a strategy primarily based on feature-level fusion, providing high-quality integrated feature inputs for dynamic decision-making models through hierarchical feature extraction and optimized mapping. ​

(3) Dynamic data update mechanism ​

In complex business systems, the dynamic and complex nature of data imposes stringent requirements on multi-modal data fusion frameworks. To ensure the timeliness and adaptability of these frameworks, a dynamic data update mechanism is crucial. Incremental learning algorithms, which update model parameters online, avoid retraining the entire dataset, significantly reducing computational costs and time. For instance, variants of stochastic gradient descent (SGD), such as Adagrad and RMSProp, can gradually adjust model parameters in batches or single samples based on new multi-modal data, such as real-time transaction records from e-commerce platforms, public opinion texts from social media, and image streams from monitoring devices, quickly adapting to changes in data distribution. Additionally, a time decay factor is introduced to construct a data timeliness evaluation system. This factor assigns weights based on the time of data generation, giving recent data higher priority while historical data's weight decreases exponentially over time. For example, the Weibo platform generates over 430 million posts daily. Without timeliness filtering, outdated information could interfere with model predictions. By using the time decay factor, the system can accurately capture current hot trends and user sentiments. The specific rules for adjusting data timeliness weights are detailed in Table 1:

| **Data generation time (based on the current moment)** | **Time decay factor weight** | **Application scenario example** |
| --- | --- | --- |
| 0-1, within an hour | 0.95 | Sudden public opinion monitoring |
| 1-6 hours | 0.85 | Short-term sales forecast |
| Within 6-24 hours | 0.7 | Market trend analysis |
| Within 1-7 days | 0.5 | Periodic strategy adjustment |
| More than 7 days | 0.2 | Long-term strategic planning |

(Table 1 Data timeliness weight)

Through the cooperative operation of incremental learning algorithm and time decay factor, the multi-modal data fusion framework can continuously absorb new information, dynamically optimize model parameters, maintain high sensitivity to the dynamic changes of business system data, and provide accurate and timely data input support for subsequent dynamic decision-making models [5].

**4. Construction of dynamic decision-making model for complex business system ​**

(1) System modeling and dynamic analysis ​

In the process of constructing dynamic decision-making models for complex business systems, system modeling and dynamic analysis are crucial steps in understanding the internal mechanisms of these systems. Complex business systems can be abstracted as dynamic networks involving multiple entities, resource flows, and environmental variables. As the core entity, a company's strategic decisions are influenced by consumer demand preferences, supplier supply stability, and competitor strategy adjustments. Resource flows, which run through all entities, maintain the system's operation. For example, the efficiency of capital flow directly impacts the company's production scale, while the speed of information flow determines the company's responsiveness to market changes. Structural Equation Modeling (SEM) can quantitatively analyze the direct and indirect effects among these elements, revealing potential causal relationships, such as the nonlinear relationship between customer satisfaction and corporate brand value. Granger causality tests can identify lead-lag relationships in time series data, assessing whether changes in one variable can effectively predict future trends in another. To account for the time-varying nature of business systems, time-varying parameters are introduced to describe the dynamic interactions between elements. For instance, the elasticity coefficient of market demand and product price can change with economic cycles and technological advancements. By updating these parameters in real-time, the system's operational patterns at different stages can be accurately captured, providing a solid theoretical foundation and analytical framework for dynamic decision-making models. ​

(2) Decision model architecture of multimodal fusion ​

The architecture of the multi-modal fusion decision-making model must fully integrate data fusion outcomes with dynamic decision-making needs to form a closed-loop decision-making system. At the input layer, the model receives multi-modal fusion feature vectors that have been processed through feature engineering and fusion strategies, converting heterogeneous data such as text, images, and time series into a structured format suitable for model computation. The core computing layer employs Recurrent Neural Networks (RNNs) and their variants, Long Short-Term Memory (LSTMs), which, with their unique memory units and gating mechanisms, effectively capture the temporal dependencies of commercial data, such as seasonal fluctuations in corporate sales and cyclical changes in market sentiment. Additionally, reinforcement learning algorithms are used to optimize decision-making strategies, with the goal of maximizing the benefits and minimizing the risks of commercial systems as the reward function. Through continuous trial and error and interaction with the environment, the model dynamically adjusts its decision-making approach, enabling it to autonomously learn the optimal decision path in complex business scenarios. In the output and feedback layer, the model outputs specific decision recommendations, such as product pricing strategies, inventory allocation plans, and market launch plans. Based on the deviation between actual decision outcomes and expected goals, feedback signals are fed back to each layer of the model, allowing for real-time parameter adjustments. This forms a complete closed loop of 'data input—feature processing—dynamic decision-making—feedback optimization,' continuously enhancing the model's decision-making accuracy and adaptability in complex business environments. ​

**epilogue**

In summary, by integrating heterogeneous data, designing timeliness mechanisms, and building closed-loop models, we can enhance the precision and real-time performance of decision-making in complex business systems. This outcome provides a decision-making framework for scenarios such as e-commerce and supply chains. In the future, we can explore the application of edge computing and generative AI in these models to expand into multi-objective decision-making scenarios, including ESG.

**reference documentation**

1. Ren Zhenhua. Research and Implementation of Abstract Generation for Multimodal Data [D]. University of Electronic Science and Technology of China, 2024.
2. Shi Zhihui and Lu Minfeng. Cross-boundary Empowerment: Insights from the Multi-modal Power Model on Intelligent Risk Control in Commercial Banks [J]. Agricultural Bank Journal, 2025, (02):34-39.
3. Zhu Chengxiang. Research on an E-commerce Image and Text Retrieval Method Based on Multi-modal Feature Fusion [D]. Harbin University of Commerce, 2024.
4. Dai Le. Business Data Mining and Application Research Based on Heterogeneous Enterprise Networks [D]. University of Science and Technology of China, 2024.
5. Ma Jingling. A Study on Financial Risk Early Warning of Listed Companies Based on Multimodal Machine Learning [D]. Hefei University of Technology, 2023.