ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20240318001408
Original scientific paper

Utilizing Internet Big Data and Machine Learning for Product Demand Forecasting and Analysis of Its Economic Benefits

Guangwei RUI*, Menggang LI

Abstract: In the context of digitalization and big data-driven advancements, the accuracy of demand forecasting in supply chain management has become a key competitive factor for businesses. This paper introduces a hybrid model combining Graph Convolutional Networks (GCN), Long Short-Term Memory networks (LSTM), and attention mechanisms, which enhances forecasting performance by integrating internet big data. The model extracts key information from multiple data sources, uses GCN to capture complex relationships within the supply chain, and employs LSTM for processing time-series data, while the attention mechanism boosts sensitivity to critical time points and relationships, significantly improving prediction accuracy. Moreover, the model optimizes production plans and inventory management, reduces the risk of supply chain disruptions, and enhances market adaptability and competitiveness.

Keywords: digital transformation; demand forecasting; economic benefits; internet big data

1 INTRODUCTION

In the context of the digital economy, supply chain management is undergoing a revolutionary transformation. The rapid development of big data technology has become the core force driving this change, bringing unprecedented transparency and efficiency to supply chain management. Through big data analytics, businesses are able to collect and process information from global supply networks in real-time, thereby monitoring supply chain dynamics, predicting market trends, and making more accurate decisions. This data-driven approach enables businesses to optimize inventory levels, reduce waste, and improve the performance of each segment of the supply chain [1-3].

Meanwhile, the integration of artificial intelligence technology has further enhanced the capabilities of supply chain management. AI-powered tools and algorithms can automatically process complex data sets, identify patterns and anomalies, thus achieving more efficient inventory management, demand forecasting, and logistics planning. For example, through machine learning models, businesses can accurately predict sales trends in different regions and adjust production and distribution plans accordingly, maximizing resource utilization and customer satisfaction. The combination of these technologies not only significantly improves the responsiveness and flexibility of the supply chain but also greatly enhances the efficiency of the entire economic system. Businesses can respond more quickly to market changes, reduce the risks of surplus or shortage, and improve customer service quality. Additionally, intelligent supply chain management also helps businesses achieve more environmentally friendly operations by optimizing routes and reducing unnecessary transportation, thus lowering energy consumption and carbon emissions. With continuous technological advancements and decreasing costs, it is expected that digital supply chain management will become more widespread in the future, continuing to drive businesses and the entire economy towards more efficient and sustainable development [4, 5]. This process not only enhances economic efficiency but also promotes the health and stability of global supply chains, providing strong support for facing future market challenges and opportunities.

In the current field of supply chain management, big data technology has become a key tool for improving the accuracy of product demand forecasting. By integrating and analyzing information from multiple data sources, including point-of-sale data, inventory records, logistics information, and socio-economic indicators, businesses can build comprehensive demand forecasting models. These data sets usually cover a wide range of historical and real-time data, enabling businesses to reveal demand patterns and changes in consumer behavior through advanced analytics. In academic research, the application of supply chain big data mainly focuses on how to effectively utilize these vast data sets to improve forecasting accuracy. Researchers explore how to identify key influencing factors through data mining techniques, such as the specific impacts of seasonal changes, market trends, and promotional activities on demand [6, 7]. Additionally, big data analytics allows businesses to conduct more detailed market segmentation, identifying specific needs of different consumer groups, thus enabling more personalized product supply strategies.

In supply chain management, leveraging big data and machine learning for product demand forecasting presents significant technological challenges, especially in processing unstructured information and deciphering the nonlinear relationships that often characterize product supply dynamics.

Incorporating Unstructured Supply Chain Information: Unstructured data, rich with insights critical for supply chain management, poses a notable challenge for traditional machine learning models which are typically designed to handle structured data. Efficient preprocessing methods are crucial for making this data usable. Techniques such as Natural Language Processing (NLP) are employed to extract vital information from textual data, transforming it into a structured format that machine learning algorithms can process more effectively. Furthermore, the design of hybrid models that seamlessly integrate both structured and unstructured data is essential for conducting thorough data analysis and achieving more precise demand forecasting.

Mining Nonlinear Relationships in Product Supply: The relationship between product supply and demand is frequently affected by complex, nonlinear interactions driven by market volatility, policy shifts, and unforeseen events. Traditional linear models fall short in capturing these complexities. Advanced machine learning approaches like deep learning and complex network theories are therefore critical, as they are capable of uncovering and learning from these hidden patterns. Deep neural networks, for instance, utilize their multi-layered architectures to learn high-level features and intricate relationships within the data. However, the implementation of these sophisticated models comes with its own set of challenges, including significant demands for data volume, computational power, and the complexities involved in model interpretability. These challenges underscore the need for careful consideration and strategic planning in the application of these technologies to ensure they are practical and effective in real-world settings.

To overcome the difficulties in handling unstructured information and mining nonlinear relationships in supply chain big data, this article proposes an integrated machine learning framework that uses information embedding techniques, graph convolutional networks (GCN), long short-term memory networks (LSTM), and attention mechanisms. Through this multi-technological fusion approach, we can more effectively utilize big data to predict product demand and have conducted an in-depth analysis of its economic benefits.

Our research contributions can be summarized as follows: 1) We have integrated internet big data into machine learning to enhance demand forecasting. This involves extracting key information from various data sources, providing a solid foundation for our forecasting framework and supplying a rich dataset for subsequent deep learning models. 2) We developed a sophisticated hybrid model that merges the capabilities of GCN and LSTM with attention mechanisms to process and analyze complex data sets. GCNs are adept at identifying intricate relationships within data, which are often missed by more traditional linear models. On the other hand, LSTMs excel in managing time-series data, capturing temporal dependencies crucial for understanding demand patterns over time. The addition of attention mechanisms significantly refines the model's ability to pinpoint and prioritize critical time points and interdependencies within the data, thereby substantially enhancing prediction accuracy. This hybrid approach not only consolidates the strengths of each individual method but also addresses their respective limitations when used in isolation. Through extensive comparative experiments across multiple simulated datasets, we have demonstrated that this integrated model substantially outperforms traditional methods that typically do not capitalize on both spatial and temporal data relationships simultaneously. 3) We have analyzed the economic benefits brought by the model. This includes reducing inventory costs through accurate demand forecasting, optimizing production plans, and reducing the risk of supply chain disruptions. Furthermore, we have also examined the model's contributions to improving supply chain transparency, response speed, and flexibility, all of which directly impact a business's competitiveness and market performance.

2 RELATED WORKS

2.1 Big Data-Based Product Demand Forecasting Model

In the context of the digital age, demand forecasting in supply chain management is undergoing significant transformations driven by big data and machine learning technologies. Recent research indicates that by integrating detailed consumer data from omnichannel sources, such as the basket data-driven model proposed by Omar et al. [8], significant improvements can be achieved in cross-channel sales forecasting, which is crucial for optimizing inventory management and enhancing retailers' market responsiveness. Additionally, Zhu et al. [9] demonstrate that incorporating upstream supply chain information into demand forecasting, especially in the pharmaceutical industry, can effectively address market fluctuations and policy changes, highlighting the potential of machine learning to solve industry-specific challenges. Furthermore, Seyedan and Mafakheri [10] explore the application of big data predictive analytics in revealing market patterns and anomalies, providing a basis for enterprises to make data-driven strategic decisions. Huber and Stuckenschmidt [11] focus particularly on the impact of calendar-specific special days on retail demand forecasting. Their research enhances the adaptability and flexibility of forecasting models by adjusting machine learning algorithms to accommodate sales fluctuations during these periods. Collectively, these studies demonstrate how big data and machine learning can improve forecasting accuracy, deepen understanding of consumer behavior, and make supply chain management more agile in responding to rapidly changing market conditions. As these technologies continue to evolve, they are expected to play an increasingly critical role in future supply chain strategies.

2.2 Economic Benefits of Digital Technologies

In the context of Industry 4.0, the contribution of digital technologies to economic growth and environmental sustainability has garnered widespread attention. Li et al. [12] demonstrated that digital technologies significantly enhance both the economic and environmental performance of enterprises by optimizing resource usage and improving energy efficiency, revealing the crucial role of digital transformation in promoting sustainable development. Furthermore, Liu et al. [13] explored how digital technologies support the dual goals of economic and environmental objectives from the perspective of the circular economy, emphasizing the importance of building sustainable ecosystems through enhanced resource recycling and supply chain transparency. At the macro level, Solomon and van Klyton [14] illustrated that although digital technologies have facilitated economic growth in Africa, there is significant variation in the impact of technology application across different regions and industries, highlighting the necessity of formulating region- and industry-specific strategies. These studies collectively underscore the multifaceted value of digital technologies in the modern economy, capable not only of driving economic benefits but also of advancing environmental protection and sustainable development globally. Therefore, devising prudent policies and strategies to optimize the application of these technologies is key to maximizing their economic and environmental benefits.

3 METHODS
3.1 Overall Framework

This study aims to construct an advanced demand forecasting model specifically tailored for the dynamic complexities of supply chain management by integrating GCNs, attention mechanisms, and LSTMs. GCNs are utilized for their capacity to model complex relationships within interconnected data, making them ideal for representing the intricacies of supply chains where nodes represent key entities such as products and suppliers. Attention mechanisms complement this by prioritizing the most informative connections between these entities, enhancing the model's focus and accuracy in response to changing market dynamics. LSTMs contribute by processing time-series data, capturing temporal dependencies essential for predicting trends over time. Together, these technologies are orchestrated to handle both the spatial and temporal challenges inherent in supply chain data, optimizing the model's performance through iterative keyword refinement, semantic relationship mapping, and trend analysis. The integration of these methods not only boosts the model's predictive accuracy but also its adaptability, making it a robust solution for demand forecasting in complex environments.

3.2 Embedding Internet Big Data

Search trend data, derived from analyzing online user search behaviors, reflects the intensity of interest users have in specific topics or keywords and how this interest evolves over time [15-17]. This data is crucial for demand forecasting as it provides deep insights into market dynamics, changes in consumer behavior, and cultural trends. Effective monitoring and analysis of these trends allow businesses to respond quickly to market changes, optimizing product and service supply plans to better meet consumer needs. In supply chain management and market analysis, leveraging search trend data significantly enhances the accuracy of demand forecasting. This data helps analysts identify which products or services are experiencing growing demand and which may be declining, thereby enabling businesses to make strategic decisions in advance. For instance, by tracking the search volume changes related to specific products, companies can predict sales trends and adjust inventory and production plans accordingly.

In practice, identifying keywords related to the target market is the initial step in the information embedding process. This study employs an iterative method to accurately extract search terms highly relevant to demand forecasting. Initially, a set of base terms is selected, which are then expanded through search engine queries to uncover more related queries. This process is repeated until no new terms are extracted, gradually refining and optimizing the keyword set to ensure it comprehensively covers and reflects market demand and consumer interests.

Algorithm: Keyword Acquisition
Input: Initial set of base terms
Output: Optimized set of keywords
Initialize the keyword set K = base terms
2. Repeat
3. New keyword set = empty
4. For each term in the keyword set
5. Query the term in search engines
6. Extract related queries
7. New keyword set = New keyword set ∪ related queries
8. End for
9. Update the keyword set = New keyword set
10. Until the new keyword set of terms does not change from the last iteration
11. Use MIC to filter out keywords with low correlation to demand
12. Return the optimized keyword set

The search volume data obtained typically contains multiple terms. Although deep neural network frameworks are used for analyzing this data, and these frameworks do not require manual feature engineering, terms with low correlation to the target variable still need to be excluded. Given the Pearson correlation coefficient's inadequacy in capturing nonlinear relationships, the maximum information coefficient (MIC) is employed to exclude low-correlation variables [18-20]. The *MIC* formula is as follows:

$$\mathrm{MIC}(X,Y)=\max_{k,l}\left(\frac{I_k(X;Y)}{\log(\min(k,l))}\right), \tag{1}$$

where X and Y represent the search trends and demand, respectively, and $I_k(X;Y)$ is the mutual information at grid size $k\times l$. The *MIC* effectively captures nonlinear relationships between variables and is therefore selected to filter candidate queries with the strongest predictive power. This step ensures the quality of data input into the model, enhancing the accuracy and reliability of the forecasting model.

3.3 Graph Convolutional Attention LSTM

In handling complex supply chain data, especially for time series-based market demand forecasting, combining GCN with LSTMs can effectively capture and analyze spatial and temporal relationships in the data [21-24]. This section details how to use the Graph Convolutional Attention LSTM model to process and analyze search trend data. A Graph Convolutional Network is an efficient method for handling graph-structured data, capable of capturing complex relationships between nodes. In our model, the specific meanings of nodes and edges are as follows:

Each node in the graph represents a keyword. The features of the nodes are the search trend data for these keywords, such as search frequency, seasonal variations, or their correlations with other keywords. Through node features, we can understand the search behavior and trends for each keyword. Edges represent the connections between nodes (keywords), indicating the semantic distance or similarity between the keywords. This similarity can be quantified using methods such as word vector similarity (e.g., cosine similarity). The weight of an edge reflects the semantic closeness between two keywords. For example, if two keywords frequently appear together in search queries, the edge connecting them will have a higher weight.

The basic operation of GCN can be expressed by the following formula:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right), \quad (2)$$

where $H^{(l)}$ is the node feature matrix at the l-th layer, $\tilde{A}$ is the adjacency matrix with added self-connections, $\tilde{D}$ is the degree matrix of $\tilde{A}, W^{(l)}$ is the weight matrix of the layer, and σ is the activation function, typically ReLU.

Introducing an attention mechanism into GCN further enhances the model's focus on important nodes, thereby improving prediction accuracy. The attention mechanism assigns a weight (attention score) to each edge, reflecting the importance of the connected nodes for the current task.

The attention score is calculated using the following formula:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(a^T\left[Wh_i \,||\, Wh_j\right]\right)\right)}{\sum_{k\in\mathcal{N}(i)}\exp\left(\text{LeakyReLU}\left(a^T\left[Wh_i \,||\, Wh_k\right]\right)\right)}, \quad (3)$$

where h_i and h_j are the feature vectors of nodes i and j respectively, a is the parameter vector of the attention mechanism, || denotes the concatenation operation, W is the weight matrix, and α_{ij} represents the influence of node j on node i.

After extracting the temporal features of the keywords, LSTM is used to process these time series data to capture the temporal dependencies of keyword search trends. The structure of LSTM is particularly suited for handling time series data and can effectively mitigate long-term dependency issues.

The core operations of LSTM involve controlling the flow of information through several gates:

forget Gate:

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right), \quad (4)$$

input Gate:

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right), \quad (5)$$

candidate Memory Cell:

$$\tilde{C}_t = \tanh\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right), \quad (6)$$

cell State Update:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \quad (7)$$

output Gate:

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right), \quad (8)$$

and hidden State:

$$h_t = o_t * \tanh\left(C_t\right), \quad (9)$$

Through these gates, LSTM can selectively remember or forget information, thereby effectively capturing long-term dependencies in time series data.

In this study, we first use GCN to extract the semantic relationships between keywords and enhance the weights of important nodes through the attention mechanism. The extracted node features and semantic relationships are then fed into LSTM to capture the temporal dependencies.

To effectively integrate GCN and LSTM in a model, the GCN first processes the input to extract spatial features from graph-structured data, where each node represents an entity and edges define their relationships. These features are then transformed into sequential data, aligning with the LSTM's requirement for time-series input. The LSTM subsequently processes these sequences to analyze temporal dynamics, learning from the evolution of node characteristics over time. This integration allows the model to harness both spatial and temporal data, enhancing its ability to forecast future trends based on complex interdependencies captured through the GCN and the sequential patterns identified by the LSTM.

By combining the strengths of GCN and LSTM, we can construct a powerful predictive model for accurately forecasting supply chain demand. This method not only captures the complex relationships and temporal dependencies between keywords but also highlights key influencing factors through the attention mechanism, thereby improving the accuracy and reliability of the predictive model.

3.4 Parameter Optimization

Parameter optimization is a critical step in developing effective machine learning models. The overall task is to solve an optimization problem that minimizes the mean squared error (MSE). The objective function is defined as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2, \quad (10)$$

where y_i are the actual values, $\hat{y}_i$ are the predicted values, and n is the number of samples. The goal is to minimize MSE to optimize the model's predictive performance.

The decision variables in this optimization problem are the parameters of the neural network, including weights and biases. The optimization problem can be formulated as:

$$\min_{\theta}\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_{\boldsymbol{i}}\right)^2, \quad (11)$$

where θ represents the set of all neural network parameters.

To solve this optimization problem, we use the Adam optimizer, which combines the benefits of momentum and adaptive learning rates. The Adam optimizer updates the parameters as follows [24, 25]:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + \left(1-\beta_1\right) g_t \\ v_t &= \beta_2 v_{t-1} + \left(1-\beta_2\right) g_t^2 \\ \hat{m}_t &= \frac{m_t}{1-\beta_1^2} \\ \hat{v}_t &= \frac{v_t}{1-\beta_2^t} \\ \theta_t &= \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} - \epsilon} \end{aligned} , \qquad (12)$$

where m_t and v_t are the first and second moment estimates of the gradients, β_1 and β_2 are the decay rates for these moment estimates, α is the learning rate, and ε is a small constant to prevent division by zero.

4 EXPERIMENTAL RESULTS
4.1 Data

The data utilized for this study comprises two main types: (1) automobile sales volumes and (2) Baidu index data, both pivotal in reflecting market dynamics and consumer interests.

Data Source and Size: Automobile Sales Volumes: We utilized data encompassing monthly cumulative sales for each of the 30 car products, spanning from June 2014 to June 2024. This dataset is sourced from relevant automobile sales databases and internal company sales records, ensuring a reliable and authentic basis for our analysis.

Baidu Index Data: As a complementary dataset, we employed the Baidu index, a robust trend analysis tool that leverages search data to gauge search popularity and trends associated with specific keywords linked to car products. This data helps in understanding shifts in consumer interest and market demand.

Preprocessing Steps: Data Collection: Initially, we gathered the required automobile sales data and Baidu index search trend data. This step ensured that all data used aligns with the specific time frame of our study, covering both historical and recent trends.

Data Cleaning: Subsequent to collection, the data underwent a rigorous cleaning process. This included the handling of missing values to avoid any bias or errors in prediction outcomes. We also removed outliers that could skew the results and standardized the data to maintain consistency across all data points.

Feature Extraction: We then proceeded to extract relevant keywords from the Baidu index data, focusing on those directly related to the car products. For each keyword, we calculated monthly search trends, providing a dynamic measure of consumer interest over time.

Data Integration and Splitting: After aligning the search trend data with the automobile sales data, we combined them into a single comprehensive dataset rich with multiple features essential for an in-depth analysis. This dataset was then divided into three parts: training (70%), validation (15%), and testing (15%), ensuring a robust framework for training and evaluating the model.

Through these meticulous steps, we have constructed a detailed and reliable dataset that not only supports the validation of our model's feasibility, accuracy, and robustness but also enhances the integrity and reproducibility of our findings.

4.2 Experimental Implementation

The experimental setup for this study utilized TensorFlow 2.0 and Python 3, running on an NVIDIA RTX2080 GPU and a quad-core Intel i7-7700 processor, providing the necessary computational power to handle the complex model training. To ensure optimal performance of our Graph Convolutional Attention LSTM model, we meticulously fine-tuned several hyperparameters. Specifically, the learning rate was carefully selected at 0.001 to balance between training speed and convergence stability. The batch size was determined to be 32, allowing for efficient memory utilization and gradient estimation.

To provide the model with sufficient training to learn intricate patterns without causing significant computational delay, we set the training duration to 100 epochs. This allowed the model to iteratively adjust weights and biases to minimize error effectively. The LSTM layer was configured with 64 hidden units to adequately capture the temporal dependencies without unnecessarily increasing the model's complexity. Additionally, the model incorporated two attention heads, enhancing its ability to focus on relevant features and relationships critically important for accurate demand forecasting.

To mitigate the risk of overfitting - a common issue in deep learning models dealing with complex data - a dropout rate of 0.5 was applied during training. This technique randomly omits a portion of feature detectors on each training case, thus reducing the chances of co-adaptations and ensuring that the model generalizes well to new, unseen data. These settings were systematically tested through several preliminary runs to determine their effectiveness in improving model performance, and adjustments were made based on observed validation loss and accuracy metrics.

To evaluate the model's performance, we used two primary metrics: Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} , \qquad (13)$$

where y_i are the actual values, $\hat{y}_i$ are the predicted values, and n is the number of samples. MAPE is calculated as:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_{\boldsymbol{i}}}{y_i}\right| . \qquad (14)$$

These metrics were chosen to provide a comprehensive evaluation of the model's accuracy in both absolute and relative terms. By leveraging the computational power of the RTX2080 GPU and the flexibility of TensorFlow 2.0 [26], we were able to efficiently train and evaluate our model, ensuring robust and reliable predictions for the automobile sales dataset.

4.3 Comparisons with Baseline Model

In this study, we selected several common time series forecasting models and deep learning models as baseline models to verify the effectiveness and superiority of the proposed Graph Convolutional Attention LSTM model. The chosen baseline models include SARIMAX, LSTM, GRU, GCN, and GCN-LSTM. SARIMAX is a traditional time series forecasting method that can capture seasonal and trend components in the data. LSTM and GRU are widely used deep learning models known for their ability to capture long-term dependencies in sequential data. GCN is effective in capturing spatial relationships in graph-structured data, and GCN-LSTM combines the strengths of GCN and LSTM to handle both spatial and temporal dependencies. By comparing the performance of these baseline models, we can better understand the advantages of the proposed method.

The results, as shown in Tab. 1, indicate that the proposed Graph Convolutional Attention LSTM model outperforms all other baseline models in terms of both RMSE and MAPE. Specifically, the SARIMAX model has the poorest performance with an RMSE of 0.250 and a MAPE of 0.060, due to its linear nature which limits its ability to handle complex nonlinear relationships. The LSTM and GRU models improve upon SARIMAX but still fall short, with RMSEs of 0.220 and 0.210, and MAPEs of 0.050 and 0.048, respectively, because they do not fully leverage spatial relationships in the data. The GCN model shows better performance with an RMSE of 0.200 and a MAPE of 0.045 by capturing the semantic relationships between keywords. The GCN-LSTM model further enhances the predictive accuracy by combining spatial and temporal dependencies, achieving an RMSE of 0.190 and a MAPE of 0.042. However, the proposed Graph Convolutional Attention LSTM model, which incorporates an attention mechanism to assign importance scores to edges, achieves the best results with an RMSE of 0.152 and a MAPE of 0.032. This demonstrates its significant advantage in capturing complex spatiotemporal relationships and improving forecasting accuracy. Overall, the experimental results validate the effectiveness and superiority of the proposed model in forecasting monthly cumulative sales of automobile products, providing a more accurate and robust method for demand forecasting in supply chain management.

Table 1 Comparison with baseline models

Method	RMSE	MAPE
SARIMAX	0.250	0.060
LSTM	0.220	0.050
GRU	0.210	0.048
GCN	0.200	0.045
GCN-LSTM	0.190	0.042
Proposed method	0.152	0.032
SARIMAX	0.250	0.060

4.4 Parametric Analysis

This section presents a parametric analysis to understand the impact of different hyperparameters on the model's performance. Specifically, we analyze the effect of the number of keywords and the MIC threshold on the model's RMSE and MAPE.

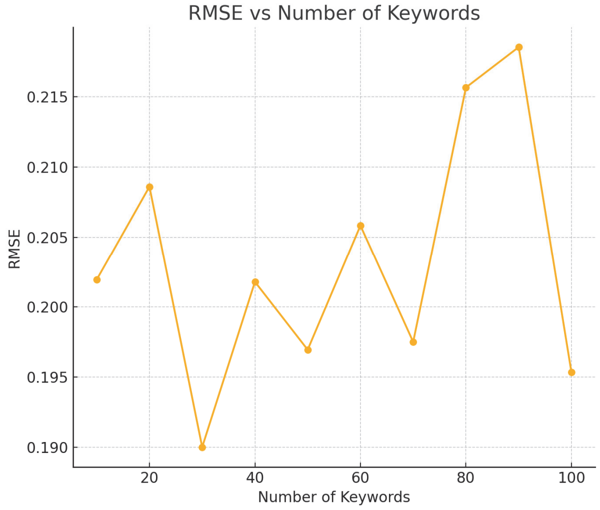

Figure 1 The RMSE varies with the number of keywords

First, we analyzed the effect of the number of keywords on the model's performance. The results demonstrate that RMSE and MAPE both decrease initially and then increase as the number of keywords changes. As shown in Fig. 1, RMSE varies with the number of keywords, and the optimal performance is achieved when the number of keywords is 30, with the lowest RMSE at 0.190. Similarly, as shown in Fig. 2, MAPE also varies with the number of keywords, with the lowest MAPE at 0.045 when the number of keywords is 30. This indicates that there is an optimal number of keywords that balances capturing sufficient relevant information and avoiding the introduction of excessive noise. Beyond this optimal number, the inclusion of additional keywords introduces noise and redundant information, leading to a decline in model performance. Therefore, appropriately selecting the number of keywords is crucial for enhancing the model's predictive accuracy.

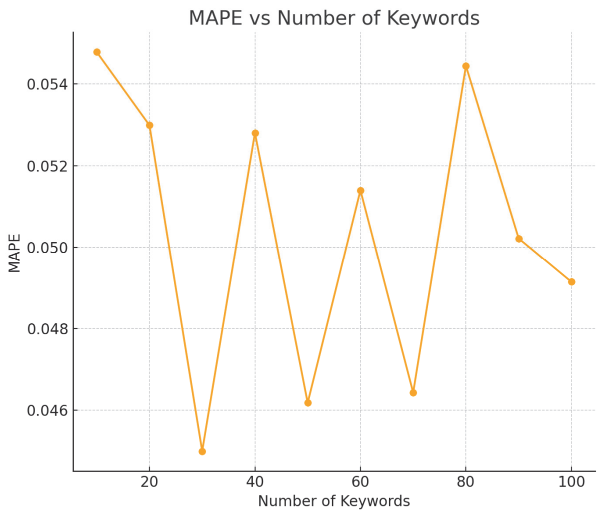

Figure 2 The MAPE varies with the number of keywords

Second, we examined the effect of the MIC threshold on the model's performance. The results indicate that RMSE and MAPE decrease initially and then increase as the MIC threshold changes. As shown in Fig. 3, RMSE varies with the MIC threshold, with the best performance achieved when the MIC threshold is 0.7, resulting in the lowest RMSE at 0.185. Similarly, as shown in Fig. 4, MAPE varies with the MIC threshold, with the lowest MAPE at 0.042 when the MIC threshold is 0.7. This suggests that selecting an appropriate MIC threshold effectively filters out keywords that are highly correlated with demand, thereby enhancing the model's predictive accuracy. A MIC threshold that is too low introduces too many irrelevant or weakly correlated keywords, while a threshold that is too high may exclude valuable information. Thus, setting an optimal MIC threshold is essential for optimizing model performance.

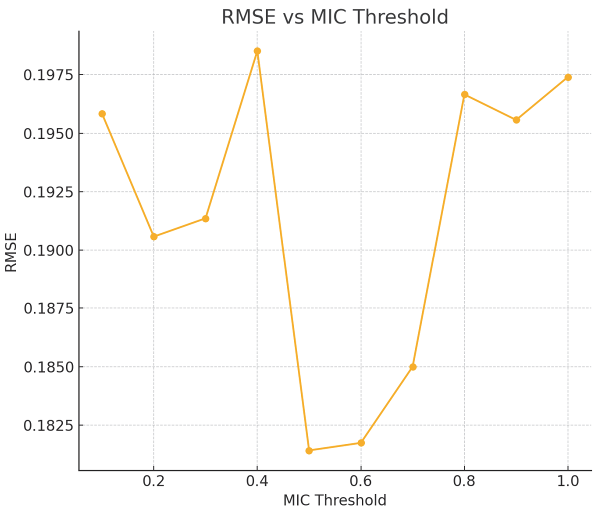

Figure 3 The RMSE varies with the MIC threshold

Data Quality and Preprocessing: One crucial factor that significantly affects model performance is the quality of the input data. Inaccuracies, missing values, or noise in the data can lead to misleading predictions. We will analyze how our preprocessing steps, such as normalization and outlier removal, enhance data quality and subsequently model performance.

Model Architecture and Hyperparameters: The specific configurations of the GCN, LSTM, and attention mechanisms, including the number of layers, the number of neurons per layer, learning rates, and dropout rates, play a critical role in model behavior. We will discuss how variations in these hyperparameters can lead to different performance outcomes, highlighting the optimal settings that emerged from our experiments.

External Environmental Factors: Factors external to the model, such as changes in market conditions, policy changes, or unforeseen economic events, can also impact the effectiveness of demand forecasting models. We will consider how such factors might have influenced the model's performance during the testing phase and discuss strategies to make the model more robust against such variabilities.

Training and Validation Data Split: The method of splitting the dataset into training, validation, and test sets can affect model training and its ability to generalize to unseen data. We will explore how different splits and data sampling techniques impact model accuracy and discuss the best practices we followed to ensure comprehensive training and evaluation.

In conclusion, we address the fine-tuning of hyperparameters such as the number of keywords and the MIC threshold, which indeed are crucial for optimizing model performance. We now clarify how these adjustments help balance the complexity of the model with its predictive accuracy, leading to significant enhancements in performance. However, we also delve into the limitations that arise when these hyperparameters are not optimally tuned, such as potential overfitting with too many keywords or underfitting with an excessively high MIC threshold. Furthermore, we discuss specific scenarios where our model may underperform, such as in data environments with highly volatile trends or sparse data points, which can challenge the model's capacity to accurately forecast demand. By discussing these limitations, we provide a more nuanced view of the conditions under which our model is most effective and where caution should be exercised. These discussions are supplemented with suggestions for mitigating these issues, such as implementing more sophisticated data preprocessing techniques to handle outliers or using ensemble methods to improve stability in volatile environments.

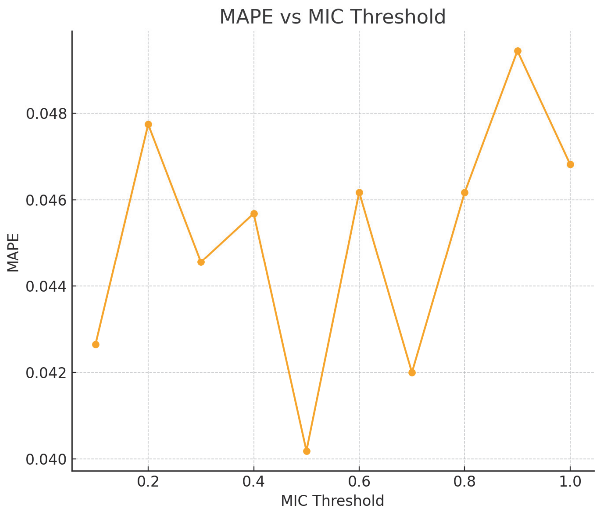

Figure 4 The MAPE varies with the MIC threshold

By incorporating these detailed analyses and discussions, our paper not only provides specific guidance for model optimization but also critically assesses the model's performance across a range of scenarios, thereby enriching its contribution to the field of demand forecasting. This comprehensive approach ensures that our findings are robust, actionable, and directly applicable to real-world supply chain challenges.

5 ANALYSES OF ECONOMIC BENEFITS

In this study, we introduced a hybrid model framework combining GCN, attention mechanisms, and LSTM aimed at enhancing demand forecasting accuracy in supply chain management. This model framework profoundly impacts operational efficiency and economic benefits for businesses by accurately analyzing and predicting market demand dynamics.

Firstly, the model directly enhances decision-making efficiency by improving forecasting accuracy. Accurate demand forecasts allow businesses to effectively plan production and inventory, avoiding issues of surplus or insufficient stock, which is particularly important in sectors dealing with high-value or rapidly changing products. For instance, by reducing unnecessary surplus inventory, businesses can free up capital for other value-added activities such as new product development or market expansion. Additionally, precise demand forecasts help enterprises optimize production schedules, reducing the need for emergency orders, thereby lowering costs associated with overtime and expedited logistics.

Secondly, the model framework helps reduce operational costs by optimizing inventory management. Adequate inventory levels can reduce the risks associated with stockpiling and product expiration, thus lowering capital tied up in inventory and storage costs. Moreover, more accurate demand forecasting reduces the occurrences of emergency purchases, which typically involve high costs due to expensive express shipping fees and potential quality compromises.

Furthermore, the application of this model significantly improves customer satisfaction and market competitiveness. Accurate inventory and demand management ensure timely order fulfillment, reducing instances of stock shortages and delivery delays, thereby enhancing customer satisfaction and loyalty. Over the long term, these improvements not only help retain existing customers but also attract new customers through positive word-of-mouth, increasing market share. Simultaneously, the model enables businesses to respond quickly to market changes, effectively adjusting marketing strategies and product supply, enhancing market sensitivity and responsiveness, thus gaining an edge in the competitive market.

In summary, by improving decision-making efficiency, reducing operational costs, enhancing customer satisfaction, and boosting market competitiveness, the hybrid model framework brings significant economic benefits to businesses. With ongoing technological development and optimization, it is expected that this model will be applied more broadly in supply chain management, further promoting efficiency improvements and cost reductions for businesses, thereby creating greater economic value.

6 THEORETICAL AND PRACTICAL IMPLICATIONS

In this study, we conducted an in-depth exploration of demand forecasting within supply chain management by implementing and testing an advanced hybrid model that integrates GCN, attention mechanisms, and LSTM. This research not only expands the existing knowledge base of predictive models theoretically but also demonstrates significant economic benefits in practical applications. Here are some insights summarized from both theoretical and practical perspectives:

Firstly, from a theoretical standpoint, the successful integration of this model underlines the critical role of data integration and the fusion of multiple modeling techniques in solving complex systems issues. By incorporating spatial structural information using GCN, processing time series data via LSTM, and focusing on critical data points with an attention mechanism, this research illustrates the effectiveness of hybrid models in enhancing prediction accuracy. This approach not only validates the concept but also sets a new direction for methodological advancements in related fields.

Secondly, the practical implications of this study are substantial, demonstrating that technological innovations can yield significant economic advantages for businesses. By implementing this hybrid model, companies can reduce inventory costs, refine production scheduling, improve customer satisfaction, and bolster market competitiveness. These enhancements contribute to greater operational efficiency and adaptability, ultimately boosting financial performance. Consequently, companies considering new technologies should evaluate their potential impact on overarching business strategies, ensuring these innovations align with long-term objectives and market needs.

Lastly, this research highlights the necessity of cross-departmental collaboration for successful technology implementation. Effective deployment of advanced technologies requires not only IT and data science expertise but also the collaborative efforts of operations, sales, and marketing teams. Establishing cross-functional teams and engaging diverse expertise from the start of the project can drive the initiative effectively, maximizing economic returns.

In conclusion, by applying this hybrid model to supply chain demand forecasting, we have validated its efficacy and offered a practical blueprint for businesses to leverage this technology in strategic decision-making. The insights and implications drawn from this study provide valuable guidance for companies aiming to enhance business growth through technological innovation. These points will be elaborated upon to clearly demonstrate the model's benefits and its applicability across various business scenarios.

7 CONCLUSIONS

This manuscript introduces a hybrid model that integrates GCN, attention mechanisms, and LSTM aimed at enhancing demand forecasting accuracy within supply chain management. By amalgamating these three sophisticated technologies, our model adeptly captures and analyzes complex patterns and spatial relationships embedded within time series data, thereby facilitating an efficient response to dynamic market fluctuations.

Experimental evaluations indicate that our hybrid model surpasses traditional singular models, such as standalone implementations of LSTM or GCN, across multiple datasets by delivering superior demand forecasting accuracy. Furthermore, an economic impact analysis underscores the practical application value of the model, highlighting significant benefits such as reductions in inventory costs, enhancements in operational efficiency, and improvements in customer satisfaction.

Despite the promising outcomes, the study is not devoid of limitations. We acknowledge that integrating our advanced model, which utilizes GCNs, LSTMs, and attention mechanisms, can be complex and resource-intensive. This integration may require significant computational resources and specialized expertise, leading to high initial costs and potential operational adjustments. In the revised manuscript, we will detail these aspects, offering potential users a clearer perspective on the practicality and financial implications of adopting this technology, aiding in a more informed decision-making process regarding its implementation and return on investment. Additionally, the robustness of the model against extreme market volatilities remains to be thoroughly validated and optimized. Our future research will focus on several key areas to enhance the performance and applicability of our model. We plan to refine the model's architecture to reduce its computational demands, making it more efficient for practical use. Additionally, we aim to develop innovative methods to improve the model's ability to handle anomalous data and enhance its robustness, particularly in managing data fluctuations and adapting to new scenarios without extensive retraining. Furthermore, we will expand the testing and application of the model across a diverse range of industries to assess its universality and adaptability. By broadening its application horizon, we intend to demonstrate the model's utility in various contexts, from healthcare to retail, ensuring it can meet different industry needs. These efforts will not only advance our model's capabilities but also contribute to the broader field of predictive analytics, making it a valuable tool for numerous applications.

8 REFERENCES

[1] Aci, M. & Yergök, D. (2023). Demand Forecasting for Food Production Using Machine Learning Algorithms: A Case Study of University Refectory. *Tehnički vjesnik*, *30*(6), 1683-1691. https://doi.org/10.17559/TV-20230117000232

[2] Wang, W. & Huang, Q. (2023). Research on Urban End-Delivery Paths Considering the Consumer's Delivery Time Demand. *Tehnički vjesnik*, *30*(4), 1209-1216. https://doi.org/10.17559/TV-20221212144331

[3] Wang, Y. L., Yin, X. M., Zheng, X. Y., Cai, J. R., & Fang, X. (2022). Supply chain coordination contract design: The case of farmer with capital constraints and behavioral preferences. *Advances in Production Engineering & Management*, *17*(2), 219-230. https://doi.org/10.14743/apem2022.2.432

[4] Singh, S. K., Cha, J., Kim, T. W., & Park, J. H. (2021). Machine Learning Based Distributed Big Data Analysis Framework for Next Generation Web in IoT. *Computer Science and Information Systems*, *18*(2), 597-618. https://doi.org/10.2298/CSIS200330012S

[5] El Filali, A., Lahmer, E. H. B., & El Filali, S. (2022). Machine learning techniques for supply chain management: a systematic literature review. *Journal of System and Management Sciences*, *12*(2), 79-136. https://doi.org/10.33168/JSMS.2022.0205

[6] Chang, D., Wang, Y., & Fan, R. (2022). Forecast of Large Earthquake Emergency Supplies Demand Based on PSO-BP Neural Network. *Tehnički vjesnik*, *29*(2), 561-571. https://doi.org/10.17559/TV-20211120092137

[7] Kim, Y. J. & Ha, B. C. (2022). Logistics service supply chain model. *Journal of Logistics, Informatics and Service Science*, *9*(3), 284-300. https://doi.org/10.33168/LISS.2022.0320

[8] Omar, H., Klibi, W., Babai, M. Z., & Ducq, Y. (2023). Basket data-driven approach for omnichannel demand forecasting. *International Journal of Production Economics*, *257*, 108748. https://doi.org/10.1016/j.ijpe.2022.108748

[9] Zhu, X., Ninh, A., Zhao, H., & Liu, Z. (2021). Demand forecasting with supply - chain information and machine learning: Evidence in the pharmaceutical industry. *Production and Operations Management*, *30*(9), 3231-3252. https://doi.org/10.1111/poms.134

[10] Seyedan, M. & Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *Journal of Big Data*, *7*(1), 53. https://doi.org/10.1186/s40537-020-00329-2

[11] Huber, J. & Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. *International Journal of Forecasting*, *36*(4), 1420-1438. https://doi.org/10.1016/j.ijforecast.2020.02.005

[12] Li, Y., Dai, J., & Cui, L. (2020). The impact of digital technologies on economic and environmental performance in the context of industry 4.0: A moderated mediation model. *International Journal of Production Economics*, *229*, 107777. https://doi.org/10.1016/j.ijpe.2020.107777

[13] Liu, Q., Trevisan, A. H., Yang, M., & Mascarenhas, J. (2022). A framework of digital technologies for the circular economy: Digital functions and mechanisms. *Business Strategy and the Environmen*t, *31*(5), 2171-2192. https://doi.org/10.1002/bse.3015

[14] Solomon, E. M. & van Klyton, A. (2020). The impact of digital technology usage on economic growth in Africa. *Utilities policy*, *67*, 101104. https://doi.org/10.1016/j.jup.2020.101104

[15] Li, M. & Xu, T. (2023). Short and long term tourism demand forecasting based on Baidu Search Engine Data. *Journal of Humanities, Arts and Social Science*, *7*(3), 529-539. https://doi.org/10.26855/jhass.2023.03.010

[16] Chen, J., Ying, Z., Zhang, C., & Balezentis, T. (2024). Forecasting tourism demand with search engine data: A hybrid CNN-BILSTM model based on Boruta Feature Selection. *Information Processing & Management*, *61*(3), 103699. https://doi.org/10.1016/j.ipm.2024.103699

[17] Pan, B., Chenguang Wu, D., & Song, H. (2012). Forecasting hotel room demand using search engine data. *Journal of Hospitality and Tourism Technology*, *3*(3), 196-210. https://doi.org/10.1108/17579881211264486

[18] Pan, W. (2021, March). Feature selection algorithm based on maximum information coefficient. *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2600-2603.

[19] Shao, F. & Liu, H. (2021). The theoretical and experimental analysis of the maximal information coefficient approximate algorithm. *Journal of Systems Science and Information*, *9*(1), 95-104. https://doi.org/10.21078/JSSI-2021-095-10

[20] Zhao, X., Deng, W., & Shi, Y. (2013). Feature selection with attributes clustering by maximal information coefficient. *Procedia Computer Science*, *17*, 70-79. https://doi.org/10.1016/j.procs.2013.05.011

[21] Si, C., Chen, W., Wang, W., Wang, L., & Tan, T. (2019). An attention enhanced graph convolutional lstm network for skeleton-based action recognition. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1227-1236.

[22] Xu, S., Rao, H., Peng, H., Jiang, X., Guo, Y., Hu, X., & Hu, B. (2020). Attention-based multilevel co-occurrence graph convolutional LSTM for 3-D action recognition. *IEEE Internet of Things Journal*, *8*(21), 15990-16001. https://doi.org/10.1109/JIOT.2020.3042986

[23] Pan, C., Cao, H., Zhang, W., Song, X., & Li, M. (2021). Driver activity recognition using spatial-temporal graph convolutional LSTM networks with attention mechanism. *IET Intelligent Transport Systems*, *15*(2), 297-307. https://doi.org/10.1049/itr2.12025

[24] Zhang, Z. (2018, June). Improved adam optimizer for deep neural networks. *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*, 1-2.

[25] Khaire, U. M. & Dhanalakshmi, R. (2020). High-dimensional microarray dataset classification using an improved adam optimizer (iAdam). *Journal of Ambient*

Intelligence and Humanized Computing, *11*(11), 5187-5204. https://doi.org/10.1007/s12652-020-01832-3

[26] Singh, P., Manure, A., Singh, P., & Manure, A. (2020). Introduction to tensorflow 2.0. *Learn TensorFlow 2.0: Implement Machine Learning and Deep Learning Models with Python*, 1-24.

Contact information:

Guangwei RUI
(Corresponding author)
School of Economics and Management,
Beijing Jiaotong University, Beijing 100044, China;
Beijing Laboratory of National Economic Security Early-Warning Engineering,
Beijing Jiaotong University, Beijing 100044, China
E-mail: 20113009@bjtu.edu.cn

Menggang LI
Beijing Laboratory of National Economic Security Early-Warning Engineering,
Beijing Jiaotong University, Beijing 100044, China
E-mail: mgli1@bjtu.edu.cn