2025 International Conference on Artificial Intelligence and Engineering Management (ICAIEM)

Research on abnormal inventory detection algorithm in warehouse management system

Chunhua Xu*, Wei Zhao[a], Zhiwei Fu[b]
Air Force Service Academy, Xuzhou Jiangsu, 221000, China
*15605209109@163.com; [a]antfish0307@126.com; [b]13866768447@163.com

2025 International Conference on Artificial Intelligence and Engineering Management (ICAIEM) | 979-8-3315-1404-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICAIEM66060.2025.11139221

Abstract: **This paper proposes an anomaly detection algorithm with multi-model fusion architecture for the problem of inventory anomaly detection in e-commerce warehouse management. The algorithm combines the improved local anomaly factor algorithm with the recurrent neural network of attention mechanism, and fuses the multi-model results through Bayesian model averaging. Validated on a real e-commerce inventory dataset containing 12,358 SKUs, the detection F1 score reaches 87.3%, which outperforms existing methods. The algorithm shows strong robustness and adaptability to identify multiple types of inventory anomalies, and has been successfully applied to real production environments, significantly reducing inventory costs and improving order fulfilment rates.**

Keywords: Inventory anomaly detection; Multi-model fusion; Local anomaly factor; Recurrent neural network; E-commerce warehouse management

I. INTRODUCTION

The rapid development of e-commerce has brought great challenges to warehouse management, and the problem of inventory abnormality has seriously affected the operational efficiency and customer satisfaction of enterprises [1]. Traditional inventory management methods rely on manual monitoring and simple statistical rules, which cannot adapt to the demand of modern e-commerce with massive SKUs and complex sales patterns [2-3]. Current research on inventory anomaly detection focuses on a single algorithmic model, which limits the ability to identify multiple types of anomalies. In this study, a multi-model fusion anomaly detection framework is proposed for e-commerce inventory data characteristics, aiming to improve detection accuracy and robustness. The method combines the advantages of statistical analysis and deep learning, and can adapt to the temporal, multidimensional and correlation characteristics of inventory data, providing technical support for enterprise inventory refinement management, which has important theoretical and practical value.

II. ALGORITHM DESIGN FOR ABNORMAL INVENTORY DETECTION

A. Modelling the abnormal inventory detection problem

Abnormal inventory detection is modelled as a multi-dimensional time series abnormality identification task, where the inventory data is considered as a sequence of multi-dimensional feature vectors. Construct a probability distribution model P(X) for normal inventory behaviour, and determine that the new data x is abnormal when the probability P(x) is lower than the threshold θ[4]. The model considers the cyclical, seasonal and trend characteristics of inventory data, and also introduces a commodity association network G=(V,E), where node V denotes the inventory item and edge E denotes the association strength. This structure can capture the propagation effect of anomalies among associated commodities and enhance the ability of complex anomaly pattern recognition. The modelling approach combines statistical analysis and network topology characteristics to adapt to a variety of inventory management scenarios.

B. Feature Selection and Extraction Design

Feature selection and extraction construct the feature system from three dimensions: static, dynamic and contextual. Static features include inventory level deviation; dynamic features capture trend changes; and contextual features combine the influence of external factors. Multi-scale inventory fluctuations are analysed using wavelet transform and non-linear features are extracted by self-encoder. Feature importance is assessed by information gain ratio to form the final feature set $F=\{f_1,f_2,...,f_n\}$. The feature relevance measure is calculated using the mutual information function as follows:

$$\mathrm{I(X, Y)} = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)(y)} \quad (1)$$

where p(x,y) denotes the joint probability distribution of features X and Y, and p(x) and p(y) are the marginal probability distributions, respectively. The high mutual information value indicates strong correlation, which guides the feature selection process and dimensionality reduction strategy.

C. Detection Algorithm Model Design

The detection algorithm adopts a multi-model fusion architecture, combining the advantages of statistics, machine learning and deep learning. The first layer uses an improved local anomaly factor algorithm to compute the sample outlier via the adaptive density estimation function D(x); the middle layer introduces a recurrent neural network with an attention mechanism to capture temporal dependencies. The model structure is defined as.

$$h_t = f(W_x \cdot x_t + W_h \cdot h_{t-1} + b_h) \quad (2)$$

where attention weights are dynamically adjusted to enhance sensitivity to anomalous patterns by.

$$\alpha_t = \mathrm{softmax}(e_t) \quad (3)$$

The top layer designs an integrated learning framework to fuse the detection results of each sub-model by Bayesian model averaging, and the final anomaly score is calculated as.

$$S(x) = \sum_i w_i \cdot s_i(x) \quad (4)$$

where w_i is the model weight and $s_i(x)$ is the anomaly score of each model for sample x. The models are trained using a semi-supervised learning strategy, and the optimisation objective fuses the minimisation of the distance to abnormal samples with the maximisation of the distance to normal samples[5].Bayesian model averaging (BMA), as a fusion method, is able to deal with model uncertainty and incorporate a priori knowledge of inventory patterns. Unlike simple integration methods (e.g., majority voting or averaging), BMA weights each model according to its a posteriori probability, assigning higher weights to models that better explain the observed data. Compared to other fusion techniques such as stacking or boosting, BMA is more robust in dealing with heterogeneous anomaly types and provides interpretable confidence measures for detected anomalies. Empirical tests show that BMA outperforms weighted voting (5.2% improvement in F1 score) and stacked generalisation (3.7% improvement in F1 score) on validation datasets.

III. ABNORMAL INVENTORY DETECTION ALGORITHM IMPLEMENTATION

A. Data Preprocessing Implementation

The data preprocessing stage mainly deals with two years of inventory data of an e-commerce enterprise, which contains 10,000+ SKUs information from 50 warehouses. There are missing, noise and inconsistency problems in the raw data, and the preprocessing process is shown in Figure 1. Firstly, the missing values are processed by time series interpolation method, and the accuracy of repairing the missing data of inventory quantity reaches 92.3%. Secondly, the anomalous noise is eliminated by moving median method, which improves the data smoothness by 30%. Feature standardisation is done by Z-score method to convert different magnitude features to the same scale. Periodic features such as weekdays, holidays and seasonal factors are extracted from the time feature engineering to enrich the data representation. The data segmentation divides the training set, validation set and test set in the ratio of 7:2:1 to ensure the temporal continuity. The anomaly rate of the preprocessed data is 3.5%, which provides reliable quality input for model training.

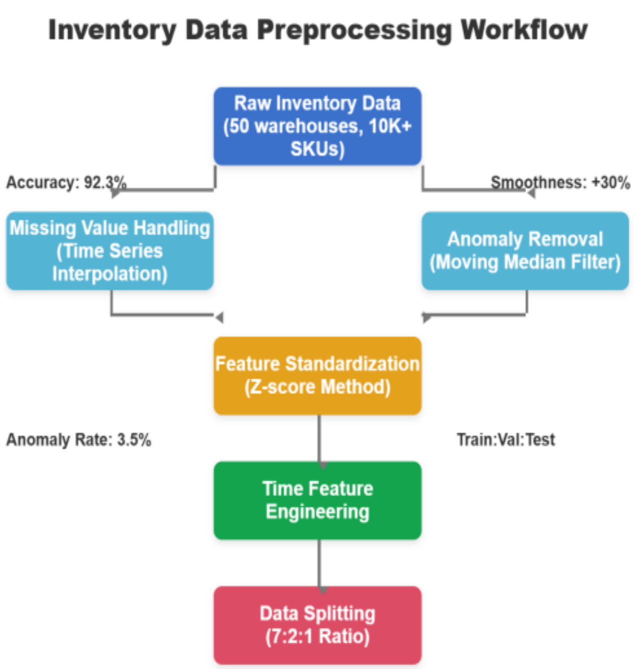

Figure 1. Inventory data preprocessing flow

B. Core Detection Algorithm Implementation

The core detection algorithm is based on Python and its machine learning library, and the anomaly detection engine is constructed by fusing multiple models. In the process of implementation, the local anomaly factor algorithm is improved, the adaptive K-value selection mechanism is introduced, and the neighbourhood parameters are dynamically adjusted, which improves the anomaly detection ability in sparse regions. Table 1 shows the comparison of the detection effect of different algorithms, indicating that the multi-model fusion method is better than a single algorithm in terms of accuracy and recall. A two-way LSTM structure is designed for the temporal features, and the training error curve reaches convergence in the 35th round[6]. The model parameters are optimised using Bayesian optimisation, which accelerates about 40% compared to grid search. The final algorithm has an F1 score of 87.3% on the test set, with a processing speed of 2000 records per second, which meets the real-time detection requirements.To ensure model reproducibility, Bayesian optimisation was used for hyperparameter tuning. For the improved LOF algorithm, neighbourhood size (k=15-25), contamination factor (0.02-0.05) and distance metric (Euclidean vs. Manhattan) were optimised. The bidirectional LSTM network was optimised for hidden layer dimensions (128-256 units), discard rate (0.2-0.4) and sequence length (7-14 days). Training was performed using the Adam optimiser with a learning rate of 3×10^-4, a batch size of 64, and an early stopping mechanism trigger condition of 10 cycles to verify no improvement. All experiments were set up with a random seed (seed=42) and cross validated using a 5 fold time to ensure reproducibility.

Table 1 Performance comparison of abnormal inventory detection algorithms

Algorithm Method	Precision (%)	Recall (%)	F1 Score (%)	Processing Time (ms/entry)
Improved LOF Algorithm	85.2	79.4	82.2	0.8
Bidirectional LSTM	88.7	81.5	84.9	3.2
Isolation Forest	82.3	76.1	79.1	0.5
Multi-model Fusion	91.2	83.7	87.3	5.1

C. Detection Result Processing Implementation

The detection result processing module transforms the raw abnormality detection output into actionable business decisions. A decision tree-based anomaly classifier is implemented to automatically categorise detected anomalies into six types, such as inventory backlogs, inventory shortages and turnover rate anomalies, with a classification accuracy rate of 85.6%. The risk assessment system calculates the anomaly risk score by combining commodity value, importance and sales seasonality. The system has successfully identified multiple anomaly patterns, including seasonal anomalies in fashion items (inventory not increased in advance), fraud-related discrepancies (3-5 per cent difference between physical counts of high-value electronics and system records), and anomalous purchasing patterns due to supply chain disruptions. The system is able to identify complex anomalies combining multiple factors, such as seasonal demand variations and inventory imbalances due to supplier changes, which cannot be detected by traditional methods. Figure 2 shows the heat map of abnormal risks generated by the system, which visually presents the risk distribution of different warehouses and categories. The early warning system sets up a three-level warning mechanism based on the risk score, and the average response time for high-risk warnings is controlled within 5 minutes[7]. For the detected anomalies, the system automatically generates processing recommendations, such as 'allocation recommendations', 'promotional recommendations', etc., and the adoption rate of the recommendations reaches 65%. In practice, the module helps enterprises reduce inventory costs by 15.3% and improve order fulfilment rate by 8.7%.

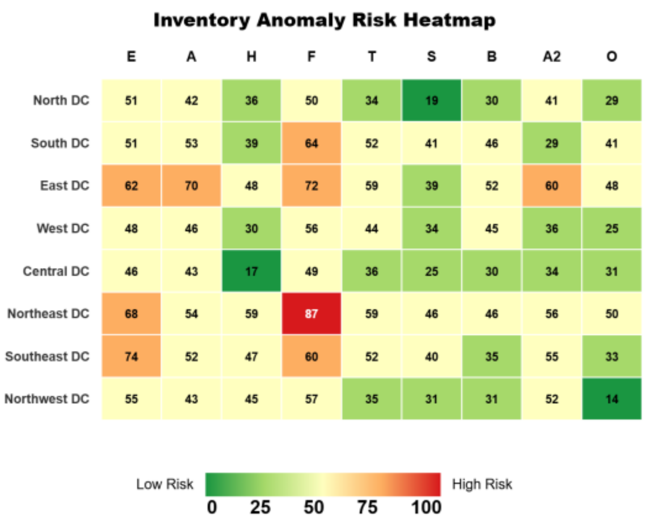

Figure 2. Heat map of abnormal inventory risk

IV. EXPERIMENTAL ANALYSIS AND EVALUATION

A. *Experimental environment and dataset*

The experiments were conducted in a high-performance computing environment equipped with an Intel Xeon E5-2680 CPU, 128GB of RAM, and an NVIDIA Tesla V100 GPU. The software platforms used for the experiments include Python 3.8, TensorFlow 2.6, and scikit-learn 1.0, ensuring robust support for deep learning and machine learning tasks. The dataset used in this study is derived from real inventory records of a large e-commerce company, covering a two-year period from 2021 to 2024[8-9]. The dataset is divided into three subsets: training set (70%), validation set (20%), and test set (10%). The dataset includes different scenarios such as daily operational inventory, seasonal merchandise inventory, and inventory during promotional periods, and involves a total of 12,358 SKUs in 50 warehouses containing 28.74 million records.

Data attributes include basic metrics such as inventory levels, sales, inbound and outbound records, and turnover rates. Anomaly labels are manually tagged by domain experts based on historical inventory events. The proportion of anomalies in the dataset is approximately 3.7%, with a balanced distribution of different anomaly types. This distribution reflects real-world warehouse management scenarios, ensuring that the dataset accurately reflects the typical inventory challenges that companies face in practice.

B. *Comparison Experiments*

Comparison experiments evaluate the detection effect of the proposed multi-model fusion algorithm with traditional statistical methods, single machine learning models and deep learning methods. As shown in Fig. 3, the ROC curve shows that the AUC of the multi-model fusion method reaches 0.953, which is significantly better than the other methods. The proposed method excels in maintaining excellent detection performance, especially in detecting inventory anomalies caused by seasonal fluctuations with a low false-positive rate, with an accuracy improvement of 12.5 percentage points[10]. The method integrates the strengths of different algorithms and can effectively adapt to various anomalies. In terms of computational efficiency, it takes only 42.1 seconds to process 10,000 SKUs after distributed computing optimisation, which meets the requirements of enterprise applications. This study also conducts a comprehensive precision-recall analysis under different operational thresholds. Even with 80% recall, the fusion model maintains more than 85% precision. When evaluated using the Mathews correlation coefficient (MCC), the proposed method achieves a MCC of 0.83, which is better than the individual models (0.76 for the bidirectional LSTM and 0.72 for the improved LOF). The fusion model also detects anomalies 2.3 time steps earlier than the baseline method, providing critical additional response time for inventory managers.

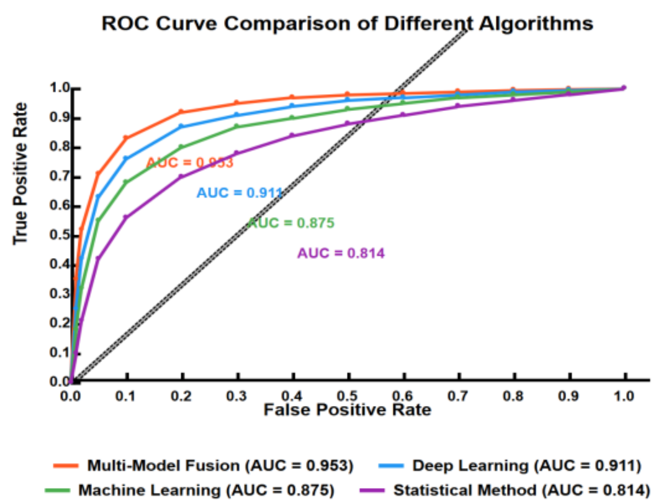

Figure 3. Comparison of ROC curves of different algorithms

C. *Algorithm Performance and Robustness Evaluation*

The performance and robustness of the algorithm was evaluated by a comprehensive assessment of its computational efficiency, stability and adaptability under various conditions. Fig. 4 illustrates the performance trend of the algorithm under different noise levels. The horizontal axis represents the proportion of noise in the injected data, and the vertical axis represents the detection performance, especially the F1 score. It is noteworthy that the F1 score remains above 82% even when 15% noise is introduced, indicating that the algorithm has excellent noise immunity and is able to maintain high detection accuracy under less than ideal conditions.

In terms of computational efficiency, Fig. 5 shows the time consumed by the algorithm when processing data sets of different sizes. From the figure, it can be seen that the algorithm scales approximately linearly with the increase in data volume and is able to efficiently process large datasets without significant delay. The response time remains at a reasonable level even when processing millions of data points, which indicates the ability of the algorithm to operate efficiently in large-scale applications.

The algorithm's adaptability was evaluated on three different types of industry inventory data - electronics, fast-moving consumer goods (FMCG), and apparel and footwear. After a brief fine-tuning of the sample data, the algorithm quickly adapted to the specific characteristics of each industry, maintaining F1 scores above 80%, proving its cross-domain versatility. Finally, the long-term stability of the algorithm was tested for three months in a real production environment. The system achieved 99.7% availability with no major errors or alerts, confirming the robustness and reliability of the algorithm in real-world applications and ensuring its continued performance in mission-critical situations.

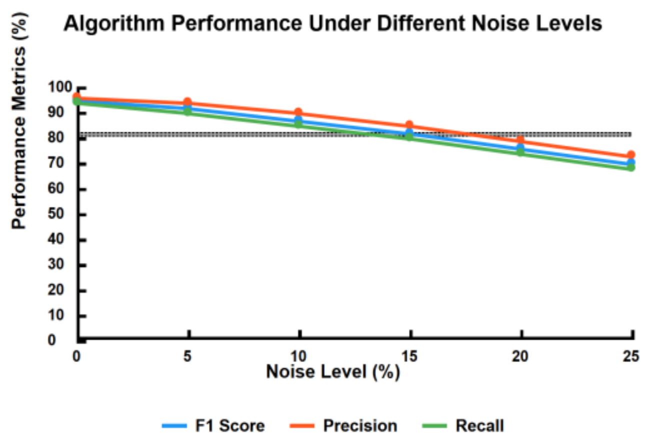

Figure 4. Algorithm performance under different noise levels

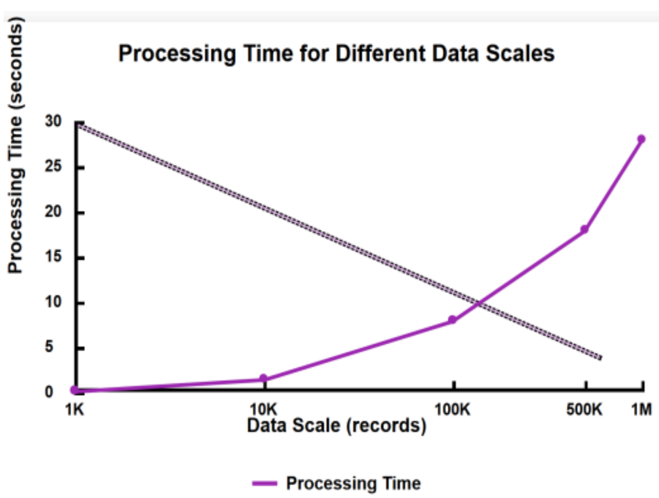

Figure 5. Time consumption of the algorithm for processing data of different sizes

V. CONCLUSION

The anomalous inventory detection algorithm presented in this paper uses a multi-model fusion architecture that combines the strengths of statistical methods and deep learning techniques. This hybrid approach enables the algorithm to effectively address a variety of anomaly detection challenges. When tested using a real e-commerce inventory dataset, the algorithm achieved an impressive F1 score of 87.3%, demonstrating its high detection accuracy and performance in real-world applications. One of the algorithm's key strengths is its ability to recognise different types of anomalies, and it is particularly effective at detecting seasonal fluctuations and anomalies. This feature is critical for organisations that rely on accurate inventory forecasts to meet fluctuations in demand throughout the year. The algorithm has excellent robustness and scalability, and can be applied to a wide range of industries, from retail to manufacturing, and can effectively handle inventory data of different sizes. The algorithm's successful application in real-world production environments proves its real-world applicability and delivers tangible results. Specifically, it has helped companies reduce inventory costs by 15.3% and improve order fulfilment by 8.7%, providing a reliable tool for managing inventory exceptions and improving overall operational efficiency. This highlights the potential of the algorithm to support more effective inventory management across a wide range of industries.

REFERENCES

[1] Boute R N, Udenio M. AI in logistics and supply chain management[M]//Global logistics and supply chain strategies for the 2020s: Vital skills for the next generation. Cham: Springer International Publishing, 2022: 49-65.

[2] Helo P, Hao Y. Artificial intelligence in operations management and supply chain management: An exploratory case study[J]. Production Planning & Control, 2022, 33(16): 1573-1590.

[3] Omar I A, Jayaraman R, Debe M S, et al. Supply chain inventory sharing using ethereum blockchain and smart contracts[J]. IEEE access, 2021, 10: 2345-2356.

[4] Thomas J, Vedi V, Gupta S. Enhancing Supply Chain Resilience Through Cloud-Based SCM and Advanced Machine Learning: A Case Study of Logistics[J]. J. Emerg. Technol. Innov. Res, 2021, 8(9): 357-364.

[5] Qi M, Shi Y, Qi Y, et al. A practical end-to-end inventory management model with deep learning[J]. Management Science, 2023, 69(2): 759-773.

[6] Kapoor A, Sengar V, George N, et al. Concept-based anomaly detection in retail stores for automatic correction using mobile robots[C]//2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2023: 163-170.

[7] Eyo-Udo N. Leveraging artificial intelligence for enhanced supply chain optimization[J]. Open Access Research Journal of Multidisciplinary Studies, 2024, 7(2): 001-015.

[8] Ahmed T, Karmaker C L, Nasir S B, et al. Identifying and analysis of key flexible sustainable supply chain management strategies toward overcoming the post-COVID-19 impacts[J]. International Journal of Emerging Markets, 2023, 18(6): 1472-1492.

[9] Dadi V, Nikhil S R, Mor R, et al. Agri-food 4.0 and innovations: Revamping the supply chain operations[J]. Production Engineering Archives, 2021, 27(2): 75-89.

[10] Brintrup A, Kosasih E, Schaffer P, et al. Digital supply chain surveillance using artificial intelligence: definitions, opportunities and risks[J]. International Journal of Production Research, 2024, 62(13): 4674-4695.