Analyzing Demand Forecasting Methods for Large-Scale Logistics Warehouses

Thinh NguyenQuang [1,2] Kosuke Matsuyama [1] Keisuke Shimizu [1] Hiroki Sugano [1] Eiji Kurimoto [1] Hasitha Muthumala Waidyasooriya [2] Masanori Hariyama [2] Tomohisa Okada [2] Kenta Sawamura [2] Masaru Hitomi [2] Masayuki Ohzeki [2]

[1]Sharp Corporation, Yamatokoriyama, Nara, Japan
[2]Graduate School of Information, Tohoku University, Sendai, Miyagi, Japan

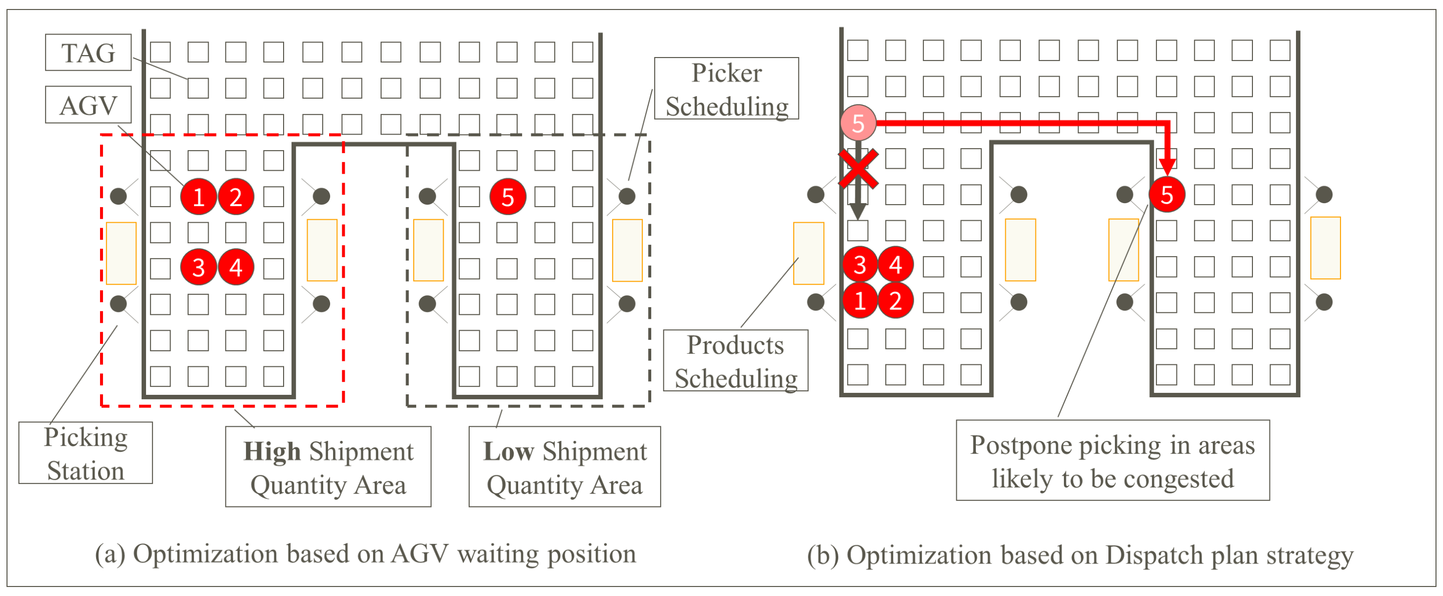
1. Background and Objectives

Previous works of Logistics Warehouses optimization and their problems:

- Optimizing warehouse operations increasingly depends on Automated Guided Vehicles (AGVs) [1].
- Accurate demand forecasting is vital for AGV allocation, dispatching, and picker scheduling (Figure 1), with several methods proposed in prior studies [2–4].
- This study focuses on forecasting challenges in warehouses handling 0.5–1.5 million products, where large-scale data makes prediction and real-time control more difficult.
- The goal is to analyze issues in prediction accuracy, training time, and to integrate forecasting results into AGV optimization system.

Objectives:

- Analyze the challenges in demand forecasting for products ranging from 500,000 to 1.5 million, focusing on prediction accuracy, learning processing time, and estimation processing time, to evaluate whether they meet the practical operation.
- Incorporate the forecasting results into a route optimization system to enhance processing capabilities. Effective demand forecasting allows for precise pre-dispatch control of AGVs, Picker/Product scheduling, and real-time adjustments to dispatch planning as Figure 1.
- Investigate potential solutions for future analytical challenges.


TAG
AGV
1
2
3
4
5
Picker Scheduling
Products Scheduling
Picking Station
High Shipment Quantity Area
Low Shipment Quantity Area
Postpone picking in areas likely to be congested

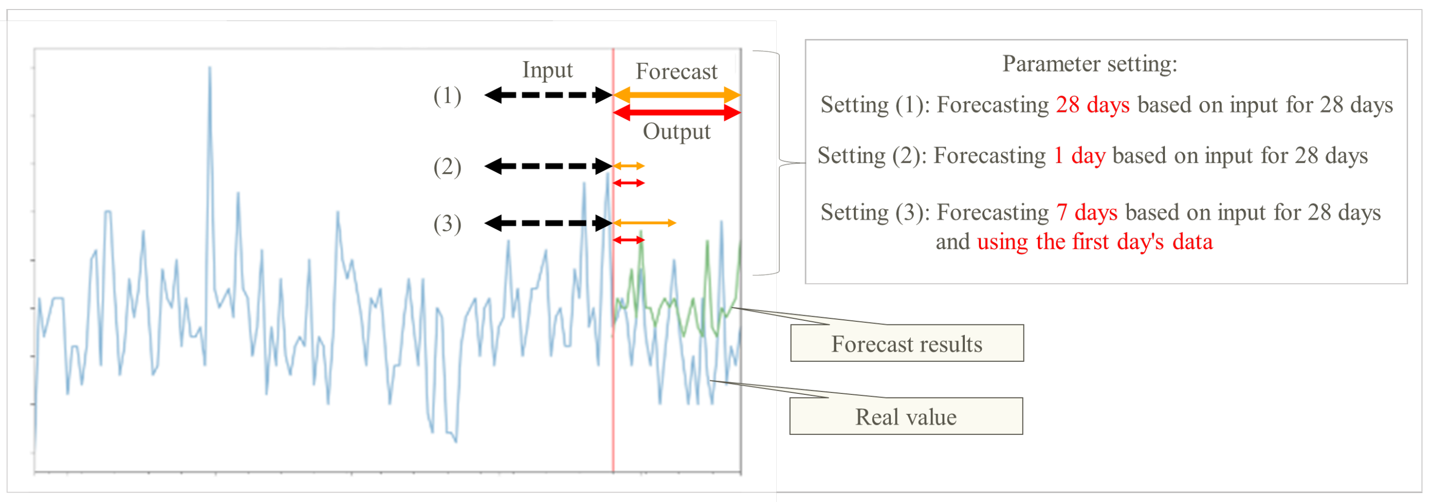
(a) Optimization based on AGV waiting position

(b) Optimization based on Dispatch plan strategy

Figure 1. AGV control Optimization based on demand forecast.

2. Demand Forecasting Methods

- The Forchestra model [3] improves large-scale demand forecasting by combining ensemble learning with deep neural networks, integrating multi-source data for real-time, accurate, and scalable predictions in logistics, retail, and supply chains.
- The DeepAR model [4] employs RNNs to generate probabilistic forecasts from historical time series, capturing uncertainty and adapting to trends for robust performance across various domains.



Parameter setting:

Setting (1): Forecasting 28 days based on input for 28 days

Setting (2): Forecasting 1 day based on input for 28 days

Setting (3): Forecasting 7 days based on input for 28 days and using the first day's data

Figure 2. Demand Forecasting parameters setting.

3. Analysis of Demand Forecasting for a Large-Scale of Products

- The analysis uses four datasets: M5 (30K), Synthetic data are Dataset 1 (100K), Dataset 2 (500K), and Dataset 3 (1M), as shown in Table 1.
- Two models, Forchestra and DeepAR, are tested with setting (1-3) in Figure 2.
- DeepAR runs on an i9-14900KF + RTX 4090, Forchestra on a Threadripper 7965WX + RTX 3070 Ti, and an additional setup includes a Xeon Gold 5320, 512 GB RAM, and 2×A100 (80 GB) GPUs.

4. Investigation of Multi-Method Demand Forecasting

Demand forecasts target 30 to 120 minutes per batch order. Table 1 shows 87 minutes for 500K products and 215 minutes for 1M products, with a 1.797 RMSE for the latter. Prediction Accuracy is Needed for forecast integration into route optimization.

Actual Forecasting Process Time and Prediction Accuracy:

- Forecasting Time: Table 1 shows 87 minutes for 500K products and 215 minutes for 1M products, requires GPU parallel processing.
- Prediction Accuracy: RMSE is 1.797 for 1 million products.

Methods for Reducing Forecasting Time and Improving Prediction Accuracy:

- Previous studies suggest more data leads to better results [2-4], but the validation results in Table 2 show that varying the number of base predictors has little effect.
- Larger datasets offer limited accuracy gains [2–4], and Table 2 reveals that RMSE does not significantly improve with base predictors ranging from 2 to 50. Therefore, the General Transform (GT) technique may enhance accuracy [5], and its implementation is under consideration.

Table 1. Comparison Results of Forecasting Methods(*).

Method	Parameters	Dataset	RMSE	Training time (s)	Forecast time (s)
Forchestra	Setting (1)	M5	2.229	19,155	1
Forchestra	Setting (2)	M5	2.035	308,497	43
Forchestra	Setting (3)	M5	2.049	423,054	21
DeepAR	Setting (1)	M5	2.409	505	42
DeepAR	Setting (2)	M5	2.945	704	361
DeepAR	Setting (3)	M5	1.962	671	646
Forchestra	Setting (3)	Dataset 1	1.651	211,544	104
Forchestra	Setting (3)	Dataset 2	1.653	666,864	269
Forchestra	Setting (3)	Dataset 3	-	-	-
DeepAR	Setting (3)	Dataset 1	-	-	-
DeepAR	Setting (3)	Dataset 2	1.789	456	5,236
DeepAR	Setting (3)	Dataset 3	1.797	405	12,912

(*) DeepAR test PC: i9-14900KF + 1 RTX 4090. Forchestra test PC: Threadripper 7965WX + 1 RTX 3070 Ti.

Table 2. Comparison effect Results of the number of base predictors(**) .

Base model	Number of base predictors	Dataset	RMSE	Training time (h)
RNN	1000 random sample from M5	2	2.229	0.95
RNN	1000 random sample from M5	50	2.035	14.50
LSTM	1000 random sample from M5	2	2.336	1.10
LSTM	1000 random sample from M5	25	2.345	12.00

(**) Setup: Intel Xeon Gold 5320 CPU @ 2.20GHz (26 cores, 1 unit), 512GB DDR4, 2 x A100 80GB GPUs (additional GPUs).

5. Conclusions

The findings from the experiments and future works are summarized below:

- Summary: Experimental results confirmed comparable performance to published studies using M5 and synthetic datasets, while analyzing training and inference times for large-scale products.
- Future work: Combine demand forecasting with route optimization and apply GT techniques to enhance feature extraction and accuracy.

6. References

[1] T. NguyenQuang et al., "Large-scale AGV routing based on multi-FPGA SQA acceleration," ASPDAC 2025, USA, pp. 1188–1194.

[2] S. Punia and S. Shankar, "Predictive analytics for demand forecasting," Knowledge-Based Systems, vol. 258, 2022.

[3] Y. Park et al., "A Large-Scale Ensemble Learning Framework for Demand Forecasting," ICDM 2022, pp. 378–387, doi: 10.1109/ICDM54844.2022.00048.

[4] D. Salinas et al., "DeepAR: Probabilistic forecasting with autoregressive networks," arXiv, 2019, arXiv:1704.04110.

[5] G. Budiutama et al., "General Transform: A unified framework for adaptive transform," arXiv, 2025, arXiv:2505.04969.

7. Acknowledgements

This work is based on results obtained from a project, JPNP23003, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).