2019 IEEE International Conference on Big Data (Big Data)

Hierarchical Demand Forecasting for Factory Production of Perishable Goods

Can Chen[1], Yijun Wang[2], Guoan Huang[2], Hui Xiong[1]
[1]Rutgers, the State University of New Jersey, USA
[2]Hangzhou LineZone Data Analytics, LLC, China

***Abstract*—Demand forecasting factory production is of particular importance for retailers of perishable goods, as they are produced daily with a fixed production lead time. Over- or underestimating demand can result in loss of profits due to stock-outs or overstock. However, demand forecasting and production planning for perishable goods represent a significant challenge due to factors such as high volatility, significant variation, the dynamics of store-level product demand, and the need for L-day ahead forecasting that allows enough time for production planning. By collaborating with a leading perishable product retailer, we have analyzed (1) detailed internal supply chain data, including sales transaction records and day-end inventories, along with (2) environmental factors, including temperature, weather conditions and wind speed. With the aim of minimizing loss of profit caused by inaccurate forecasting, we propose the following three-stage hierarchical demand forecasting model that leverages the combined data for perishable goods production planning: 1. Identification of store-level demand patterns, 2. store clustering for aggregated production, and 3. a recurrent dynamic network based on a nonlinear autoregressive network with exogenous inputs (NARX) for L-day ahead demand forecasting. Finally, we validate the proposed approach by comparing the loss of profits using this model with other baselines along with the industry standard model used in the perishable goods industry. Our proposed model successfully reduces lost profits to 3.30% of total sales, representing a reduction of 1.71% when compared with the industry standard production system.**

***Index Terms*—Perishable goods; Deep learning; Hierarchical forecasting; Production planning**

I. Introduction

Recent years have the emergence worldwide of AI-powered supply chain management systems [1], which integrate machine learning techniques and optimization algorithms that leverage both internal supply chain internal data and external factors. Retailers, especially those offering perishable consumer goods, are faced with the challenge of how to accurately forecast product demand in order to manage their daily operations [2]. Perishable goods are typically ordered, produced, and delivered on a daily basis with high demand volatility and short shelf-life. Items that have not been sold by the end of the day result in demand waste. On the other hand, items that are sold-out early cause demand loss. An accurate forecasting model can help retailers of perishable goods to reduce lost profits by increasing product availability while limiting the day-end waste.

While the importance of accurate forecasting for daily operations is clear to managers dealing with the supply chain for perishable goods, actually designing an accurate forecasting model for production has remained a challenge for several reasons. Firstly, the dynamics of product demand inevitably give rise to high variance at individual store level. A strategy that predicts daily demand at store level and then aggregates the total demand tends to lead to inaccurate forecasts. Secondly, product demand relies heavily on consumers' purchasing habits as well as external factors, such as temperature and other weather conditions. It is necessary to combine the internal supply chain data (transactions, inventories, etc.) with external data (meteorology reports, special events, etc.) in order to develop a reliable forecasting model. Thirdly, the time period from the creation of the production plan to the time when the product is produced and delivered presents challenges in forecasting demand for the days during this lead time.

Recently, a number of studies have been carried out investigating the forecasting problem for factory production of perishable goods. Most industry standard forecasting models rely on a moving average model [3] at the individual store level combined with adjustments from experienced store managers [4]. However, retailers of perishable goods continue to face major problems with demand loss and demand waste over long-term operations. Du et al., (2013) propose a support vector machine (SVM) forecasting system to predict demand for perishable farm products. However, the SVM model neglects the time dependencies of daily demand. Van et al., (2016) [5] test several regression models for promotional demand forecasting for perishable goods. Recently, Huber et al., (2017) proposed a cluster-based hierarchical demand forecasting model at different organizational levels based on an agglomerative hierarchical clustering algorithm and an ARIMA model. However, this model is not able to support production planning for individual products taking into account production lead time and store-level aggregation.

The emergence of multi-source big data and well- organized internal supply chain data has enabled a new paradigm that can enhance demand forecasting for the production of perishable goods. This has allowed us to exploit internal supply chain data (transactions, inventories, etc.) and external data (meteorology reports, special events, etc.) to develop a new hierarchical demand forecasting model. Specifically, we first analyze store-level demand patterns from product demand time series and their sensitivity to external factors. A KMeans clustering algorithm is then used to aggregate stores with similar patterns to produce an aggregated production plan.

978-1-7281-0858-2/19/$31.00 © 2019 IEEE

Finally, a recurrent dynamic network based on a nonlinear autoregressive network with exogenous inputs (NARX) can be developed for $L-$day ahead demand forecasting that can meet the requirements of production lead time. Finally, we conduct extensive experiments on a real-world data set in collaboration with a leading Chinese retailer of perishable goods. Our proposed method successfully reduces profit loss to 3.30% of total sales, representing a reduction of 1.71% when compared with the industry standard production system.

The remainder of this paper is organized as follows. Section II provides the context and defines the problem of demand forecasting for production of perishable goods. After that, Section III examines the methodologies used to reveal demand patterns, store aggregation, and demand forecasting models. The results of the experiment are reported in Section IV demonstrating the superior performance of the proposed model. Finally, Section V summarizes some related work and Section VI briefly summarizes our contributions.

II. Problem

In this section, we set out the notations and definitions that appear throughout this paper. We then define the prediction problem for factory production of perishable goods.

A. Notation and Definitions

Notation	Definitions
L	Production lead time
T	Evaluation time period
t	Next t day forecasting, $t = 1, 2, ..., L$
$s \in \mathcal{S}$	Set of all retail stores
$\mathcal{C}_k$	k-th set of retailer stores, $\bigcup_k \mathcal{C}_k = \mathcal{S}$,
$r \in \mathcal{R}$	Set of store keeping units
$I^t_{s,r}$	Day-end inventory for item r in store s on day t
$y^t_{s,r}$	Sales of item r in store s on day t
w^t	Weather condition on day t
f^t	Temperature on day t
v^t	Wind speed on day t
y^t_i	Actual total demand of product i on day t
$\hat{y}^t_i$	Predicted total demand of product i on day t
λ_i	unit profit loss due to product waste of product i
η_i	unit profit loss due to demand loss of product i

Definition 1 (Store hierarchy): A 3-level store hierarchy is shown in Figure 1. It starts from the roof representing all the stores S that are supplied by the same factory. The first level is made up of different clusters of stores $\mathcal{C}_k, k = 1, 2, ..., n$. Typically, stores in the same cluster are located in similar urban functional zones and the demand in these stores displays similar sensitivity to external factors. The leaves represent the individual stores.

In order to provide forecasting analysis for factory production planning, analysis can be conducted at the roof level by analyzing the total demand of all retailer stores directly. The production planning systems currently in use rely on forecasting for individual retail stores at the leaf level, and then aggregate the estimated demand to determine factory production. Our proposed alternative, which we can call hierarchy aggregation, first clusters together stores with similar demand patterns and then conducts forecasting analysis at the cluster level before aggregating the total demand of all clusters.

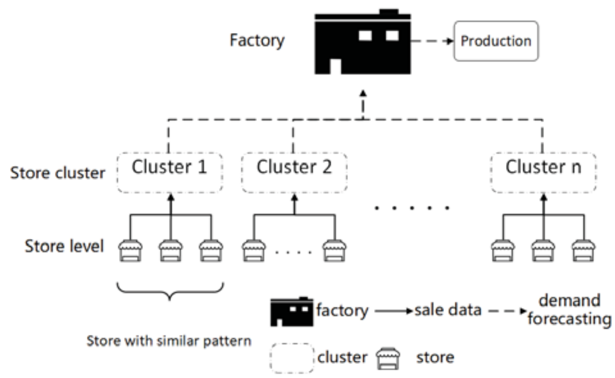

Fig. 1. Hierarchical structure for retail stores

Definition 2 (Product demand): The product demand $y^t_{s,r}$ is defined as the amount of product r that can be sold in store s on day t when the supply is sufficient. In some cases, product r is sold-out before the end of the day. Here we estimate the unrecorded demand when the store is out-of-stock by using the 1-Nearest-Neighbor predictor that utilizes historical sales data when the supply is sufficient during the day [3].

Definition 3 (Production lead time): Production lead time L is defined as the number of days a factory needs to complete its production plan. The production lead time for different products is set to be 3 days. To better decide the production volume for the t-th day, it is necessary to forecast the following L days.

Definition 4 (Profit loss): In real-world production planning, if the retailer forecasts a demand greater than the actual consumption, there is a risk that the left-over product will be wasted. If the forecast result is less than the actual demand, it may cause a demand loss. The combined product waste and demand loss due to inaccurate forecasting together as profit loss. Formally, the profit loss for product i on day $T + t$ is defined as follows:

$$\mathcal{F}_{i,t} = \sum_T \lambda_i max(\hat{y}_i^{T+t} - y_i^{T+t}, 0) + \eta_i max(y_i^{T+t} - \hat{y}_i^{T+t}, 0)$$

where the first term represents the wastage of product i on day $T+t$ leveraged by the cost per unit of product wasted λ_i. The second term represents the demand loss leveraged by the product price η_i.

B. Problem Description

Given a set of historical transaction records, store inventory levels, and meteorological reports, the demand forecasting problem for product production can be stated as follows: to predict the total demand of all stores supplied by the same factory with a lead time of L, with the objective of minimizing the total loss of profit due to forecasting inaccuracy.

As mentioned in Definition 1 above, there are three fundamentally different approaches to production plan forecasting: leaf-level production planning, roof-level production planning,

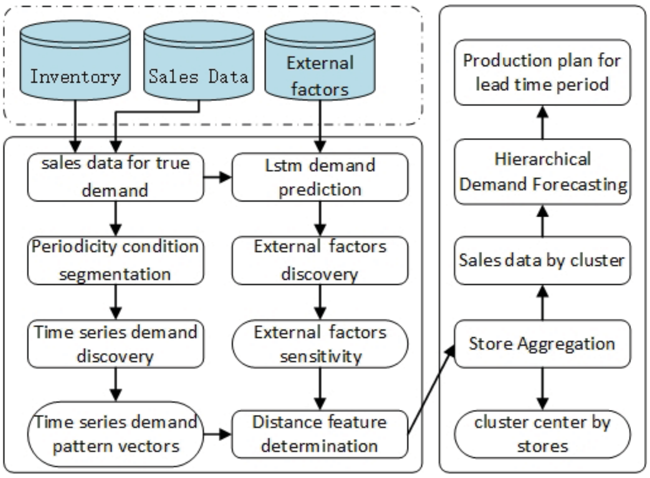

Fig. 2. Framework overview.

and hierarchy aggregation production planning. Leaf-level production planning forecasts the product demand at store-level and aggregates the predicted demand for production. This is the strategy currently employed by most of the leading retailers in China. Roof-level production planning works by directly predicting the total demand of the retail stores in a given region. The third approach, hierarchy production planning, first aggregates stores with similar demand patterns into groups and then predicts the total demand of these store clusters. The production plan is drawn up to meet the total demand of all clusters of stores.

C. Framework Overview

Figure 2 shows the framework of our proposed approach, consisting of three distinct phases: identification of store-level demand patterns, store aggregation, and hierarchical demand forecasting for production planning.

Identification of Demand Patterns. Demand patterns can be divided into two categories: time series demand patterns and demand sensitivity to environmental factors. We first analyze the store-level transaction history of a product in order to extract its time series demand pattern, which describes the product demand repeating purchase, weekly purchase period, and demand trend. Then the weather reports for each time slot can be leveraged for environmental factor pattern analysis. Here we implement a single LSTM unit to show demand sensitivity to environmental factors [6].

Store Aggregation. Store aggregation uses the store-level demand patterns as input, and aims to group the stores with similar patterns together, thereby forecasting product demand at a cluster level. Here we implement a KMeans clustering algorithm to identify clusters of stores. As a result, the demands of stores grouped into a cluster are aggregated into our forecast targets. The individual store level forecast and roof node level forecast are two special cases of $K = |\mathcal{S}|$ and $K = 1$ respectively.

Hierarchical Demand Forecasting Finally, we are able to predict the total demand of a cluster of stores based on a dynamic recurrent neural network with externals for $L - day$ ahead forecasting, with a loss function defined as the profit loss.

III. Methodology

In this section, we first consider the methodologies used in the above framework, including for the identification of store-level demand patterns and store aggregation. We then look at the hierarchical demand prediction model based on a nonlinear autoregressive network with exogenous inputs (NARX).

A. Store-level Demand Patterns Discovery

Time Series Demand Pattern. Motivated by time series pattern discovery [7], here we use daily product demand of one week, defined in 2 to represent the store-level time series demand pattern. Formally, the time series demand pattern of product r at store s ($XT_{s,r}$) is defined as follows:

$$XT_{s,r} = (\bar{y}^1_{s,r}, \bar{y}^2_{s,r}, ..., \bar{y}^7_{s,r})$$

where $\bar{y}^k_{s,r}$ is the average demand on the $k - th$ day of the week.

External factors Sensitivity Analysis. Here we use wind speed, temperature, and weather conditions as our external factors E. First we train a neural network with historical data for the past month and then calculate the XE sensitivity (y rate of change) of change E based on each factor plus or minus 5 percent variation. Here we assume that y changes $\triangle y$ is linear when the factors variation $\triangle x$ is small.

B. Stores Aggregation

The current production planning approach employed by our industry partners uses sales forecasting at store level without store aggregation. As a result, the forecast suffers from serious demand variance at the individual store level. We propose a KMeans clustering algorithm to aggregate stores according to their time series demand patterns and demand sensitivity to external factors as revealed in Section III-A.

Starting with the store time series patterns XT, and external patterns sensitivity XE as input, we first randomly initialize k points as the cluster centers. Then the distance $DS(k, j)$, defined by Expression 1, is calculated between each retail store j and each clustering center k. Finally, we partition each point to the cluster center with the nearest distance to complete a cycle. We recalculate the cluster center of the stores of the same label until the cluster centers do not change.

$$\begin{aligned} DS(k,j) &= (1 - Cor(k,j)) + Dis(k,j) \\ Cor(k,j) &= \frac{Cov(XT_k, XT_j)}{\sqrt{Var[XT_k]Var[XT_j]}} \\ Dis(k,j) &= \sum_i^N (XE_{ki} - XE_{ji})^2 \end{aligned} \tag{1}$$

Figure 3 shows the with store distribution in Hangzhou City and the aggregated result for product Yoghurt. The different colors represent different clusters, while the larger the dot size, the more stores there are in a cluster. 200 retail stores that

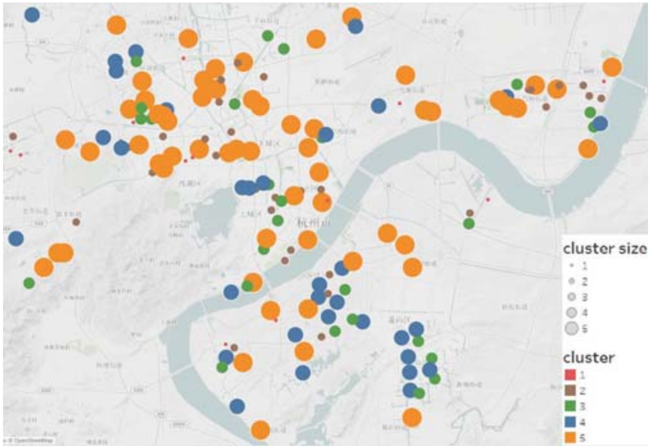

Fig. 3. The cluster of stores for product Yoghurt

selling the product are grouped into 5 clusters. The aggregated results and the cluster center characteristics show that stores located near residential areas (Cluster 5) and business areas (Cluster 4) have a high demand patterns that are highly sensitive to weekdays and weekends, while being less affected less by the external factors. Stores located near the West Lake or other famous attractions in Cluster 2 have a high sensitivity to external factors. Stores in Cluster 1, that are located far away from the central business areas, have low sensitivity to external factors. Others, including the stores located near subway stations in the central business district, have a clear time series demand pattern and a high sensitivity to external factors and are grouped in cluster 3.

C. Sales Forecasting

We propose a nonlinear autoregressive network with exogenous inputs (NARX) to forecast the sales demand based on the features extracted. This is a powerful class of models that has been demonstrated to be well-suited to modeling nonlinear systems and especially time series. Learning is more effective in NARX networks than in the other neural networks and convergence is much faster. [8]. Moreover, they outperform other widely used prediction algorithms for our sales distribution forecasting with significantly improved accuracy. We utilize the notation h represents the current time of the network, n_u and n_y denote delays to the input and output, S is the number of neurons in the hidden layer, f is the activation function of hidden layer neurons and the f_{M-1} is a linear function of the output neuron. The details of the M-layer feedforward network are summarized below:

Layer Input. Since the features are from different factors and have different ranges, they are normalized within [0,1] ranges by mapping $x = \frac{x-x_{min}}{x_{max}-x_{min}}$ in order to prevent the simulated neurons from being driven too far into saturation [9]. The input-output relationship at the moment of $h+1$ is shown as follows:

$$n_i(h+1) = \sum_{i=1}^{n_u} w_{ji}u(h-i) + \sum_{j=1}^{n_y} w_{ji}y(h-j) + b_i^{h+1} \quad (2)$$

Layer Output. The output in layer $k+1$ is mapped from n_i using a linear function:

$$y\left(h+1\right) = \Phi\left(n_i(h+1)\right) \quad (3)$$

The output layer is a linear layer for the regression problem of sales forecasting and the final output a^M is t_{sd} (continues variable).

Training Algorithm. The training task is to learn the associations between our training set of input-output pairs $\{(\overline{Y_1}, u_1), (\overline{y_1}, u_2), ..., (\overline{Y_h}, u_h)\}$ which aims to minimize the profit loss $\mathcal{F}_i$ caused by inaccurate forecasting as defined in Definition 4.

The Levenberg-Marquardt algorithm has been shown to be the fastest tool for training moderate-sized feedforward neural networks with the sum of squared error objective is applied for parameter training in our study [10], [11]. Moreover, a testing set is used to monitor validation error without affecting training parameters during the training process. When the neural network begins to overfit, the validation error will begin to rise. Our optimal training parameters are chosen at the time with the minimum validation error [12], [13].

IV. Experiment

A. Datasets

We use the supply chain data for retail stores in the cities of Hangzhou and Wenzhou, and the meteorological records from October 2016 to December 2018 in order to conduct model training and evaluation. The data for the previous year is used for training, and the evaluation is conducted as a rolling forecast for the next 3 months. A statistical summary for the data is provided in Table II. The experiments were conducted on a server with 24 Core CPU (Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz) and 120GB Memory.

Supply Chain Data. We use a complete set of supply chain data, including factory production volume, store-level day-end inventory, and product sales transactions. A total of 1500 retail stores from three major cities and 100 products are included in our research.

Meteorological Records. The meteorological report data consists of hourly weather reports, including time, weather conditions, temperature and wind speed, which are publicly available from Weather Underground[1]. The missing meteorological data is generated according to the preceding hourly recorded weather report and the missing wind speed is estimated using the average value of its previous and next reports.

B. Baselines

- **Industry Standard(IS)**: The Industry Standard model is currently employed by the production planning departments of leading retailers of perishable goods in China. The planner takes a historical moving average as its reference, with further consideration given to meteorological conditions, holidays, events, and other external factors they think may affect total demand.

[1]https://www.wunderground.com

TABLE I
DETAILS OF THE DATASETS

Data Source		Hangzhou	Wenzhou
Time Span		10/1/2016 - 12/31/2018	
Sales Data	# of stores	337	502
	# of articles	99	98
	# of trade(million)	115.64	223.01
	Shelf-life(day)	[1,14]	[1,30]
External Factors	Holidays	79 day	79 day
	Weather conditions	12 types	6 types
	Temperature(^oC)	[2, 37]	[5, 38]
	Wind speed(mph)	[0.6, 5.6]	[0.5, 6.3]

- **Random Forest (RF)** [14]: RF fits a number of decision trees on various samples of the original dataset and uses the averaged results for forecasting and over-fitting control.
- **ARIMAX** [15]: The ARIMAX model is an ARIMA model with additional explanatory variables. The ARIMA part consists of autoregressive (AR) and moving average (MA) components. In this paper, we set the parameters of ARIMA model (p,d,q) = (7,0,1) which can get the performance is optimal.
- **Sequence to Sequence (Seq2Seq)** [16]: Sequence to sequence learning has been successfully implemented in many tasks, including machine translation and demand forecasting. The Seq2Seq used in our experiment has 64 hidden units, a learning rate set at 0.01 and an input sequence length set at 56.

C. Evaluation metrics

We use total Sales normalized Demand Loss (SDL), Demand Waste (SDW), and Mean Profit Loss (SMPL) to evaluate the performances of our method and baselines for next t day sales forecasting, $t = 1, 2, 3$:

$$\begin{aligned} SDL_i^{+t} &= \frac{1}{N}\sum_i^N \sum_T \lambda_i max(\hat{y}_i^{T+t} - y_i^{T+t}, 0) \\ SDW_i^{+t} &= \frac{1}{N}\sum_i^N \sum_T \eta_i max(y_i^{T+t} - \hat{y}_i^{T+t}, 0) \\ SMPL_i^{T+t} &= SDL_i^{+t} + SDW_i^{+t} \end{aligned} \tag{4}$$

where T represents the evaluation time period.

D. Experiment Setup

The experiments consist of three stages: sales forecasting for individual stores, sales forecasting at roof node level, and aggregate sales forecasting in a region and aggregate production. The first stage deals with the comparative accuracy of the different methods at individual store level, with the aim of testing how localized demand patterns can affect forecasting. The second stage involves testing the forecasting results at the roof level of a whole city. Finally, a comparison of accuracy for the entire set of sales forecasts between the sum of individual stores and the aggregate method can be evaluated. This stage will prove the effectiveness of our aggregation production strategy. For all stages, we use the same evaluation metrics and steps for rolling forecasting. We use outputs $T + 3$ sales forecasting results for a 3-day lead time factory production plan, and evaluate the performance.

E. Experiment Results

Table II is a detailed performance evaluation including demand loss, demand waste and profit loss of the different sales forecasting models. This shows that our proposed model achieves the best performance over these three evaluation metrics. It can help reduce the profit loss to a level of 3.30% for $T + 3$ forecasting, which is a considerable reduction from the industry standard system.

TABLE II
DETAILED PERFORMANCE COMPARISON

Aggregate Type	Model	SDL	SDW	SMPL
Gap(Roof)	IS	1.53%	8.49%	5.01%
Roof node at city level	RF	3.57%	5.59%	4.58%
	Seq2Seq	4.07%	12.4%	8.24%
	ARIMA	3.52%	7.17%	5.34%
	NARX	3.64%	4.79%	4.22%
Individual store	RF	2.38%	6.99%	4.69%
	Seq2Seq	2.91%	9.97%	6.44%
	ARIMA	0.38%	11.28%	5.83%
	NARX	2.32%	5.30%	3.81%
Aggregate forecasting	RF	3.12%	5.8%	4.46%
	Seq2Seq	5.26%	6.45%	5.86%
	ARIMA	3.37%	7.61%	5.49%
	NARX	2.90%	3.70%	**3.30**%

V. RELATED

Identification of Demand Patterns. Many researchers focus on the identification of demand patterns for store keeping units for retailers. Jakob, etl (2017) [17] propose a list of promotion patterns, such as price discounting, featured promotions and shelf displays. Sanjita, etl. (2014) [18] focus on the demand bullwhip effect and its time series seasonality and stationary patterns. Vijayalakshmi and Bernard (2013) [19] investigate time series trends and seasonal and irregular demand patterns. Nari and Diane (2015) further considers the internal demand patterns, such as promotions, price reduction, stock outs, and external factors (e.g. weather, holidays). However, the identification of these patterns is mainly based on statistics. A model-based method of identifying demand pattern and aiming to aggregate retail stores with similar patterns to minimize demand variance needs to be carefully studied.

Sales Forecasting Complex temporal series forecasting for retail sales has attracted substantial academic attention and seen the application of multiple methods. The literature mainly focuses on two sets of problems. The first deals with sales forecasting for individual retail stores [1,2]. Du, et al., (2013) proposes a support vector machine (SVM) forecasting system for demand forecasting for perishable farm products. Van, et al., (2016) [5] test several regression models for promotional demand forecasting for perishable goods. Recently, Huber, et al., (2017) proposed a cluster-based hierarchical demand forecasting approach at different organizational levels based on an agglomerative hierarchical clustering algorithm and an ARIMA model. Others apply machine learning techniques, such as deep neural networks (DNN) [20], to predict sales volumes for specific products without any regional concept. As trends at the level of individual retail stores are not that strong, these methods achieve good performance but do not take full advantage of the temporal information. These approaches are not equipped to deal with regional forecasting with seasonal patterns.

Another group focuses on sales forecasting for regional retail sales, called aggregate forecasting [21]. This differs from the study at the individual store level, because this approach first groups stores at region or city level, and then conducts forecasting at this level only. However, this strategy relies on a good store aggregation method to lower the store-level demand variance while maintaining a high sensitivity to external factors.

VI. Conclusion

In this paper, we have developed a hierarchical demand forecasting model for factory production of perishable goods. Specifically, we first identified time series demand patterns and the degree of demand sensitivity to external factors at the individual store level. We then grouped together stores with similar demand patterns and sensitivity to generate our aggregation production forecast. Then, a dynamic recurrent neural network was used for aggregated demand forecasting. Various evaluation metrics including demand loss, demand waste, and profit loss were utilized. We compared three different forecasting strategies: individual store demand forecasting and aggregation, roof node demand forecasting for stores over a large region, and finally our proposed aggregation forecasting model. Finally, extensive experiments using the detailed internal supply chain data from the largest perishable goods retailer in Zhejiang Province (China) demonstrated the advantages of our approach. Our proposed method was able to reduce lost profits to 3.30% of total sales, representing a reduction of 1.71% reduction from the industry standard production system.

This work has some limitations that require future research. Firstly, we have focused on the forecasting element of an aggregation production plan. An inventory model with a daily production limit should be integrated for a more thorough evaluation. Another study is proposed to explore the effects of major events, such as street or school closures. Here we have presented only one complete data set from the retailer's internal data. However, our proposed model can easily be expanded to consider other external factors.

VII. Acknowledgements

This research was partially supported by the National Science Foundation (NSF) via the grant number: 1814510.

References

[1] T. Nguyen, L. ZHOU, V. Spiegler, P. Ieromonachou, and Y. Lin, "Big data analytics in supply chain management: A state-of-the-art literature review," *Computers & Operations Research*, vol. 98, pp. 254–264, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0305054817301685

[2] C. W. Holsapple and M. P. Sena, "Erp plans and decision-support benefits," *Decision Support Systems*, vol. 38, no. 4, pp. 575–590, 2005.

[3] N. S. Arunraj and D. Ahrens, "A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting," *International Journal of Production Economics*, vol. 170, pp. 321–335, 2015.

[4] K. H. van Donselaar, V. Gaur, T. van Woensel, R. A. C. M. Broekmeulen, and J. C. Fransoo, "Ordering behavior in retail stores and implications for automated replenishment," *Management Science*, vol. 56, no. 5, pp. 766–784, 2010. [Online]. Available: https://doi.org/10.1287/mnsc.1090.1141

[5] K. H. Van Donselaar, J. Peters, A. de Jong, and R. A. Broekmeulen, "Analysis and forecasting of demand during promotions for perishable items," *International Journal of Production Economics*, vol. 172, pp. 65–75, 2016.

[6] Y. Li and Y. Zheng, "Citywide bike usage prediction in a bike-sharing system," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2019.

[7] J. Liu, Y. Fu, J. Ming, Y. Ren, L. Sun, and H. Xiong, "Effective and real-time in-app activity analysis in encrypted internet traffic streams," pp. 335–344, 08 2017.

[8] E. Diaconescu, "The use of NARX Neural Networks to predict Chaotic Time Series," *Computer*, vol. 3, no. 3, pp. 182–191, 2008.

[9] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks:: The state of the art," *Int J of forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[10] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the marquardt algorithm," *Neural Networks, IEEE Transactions on*, vol. 5, no. 6, pp. 989–993, 1994.

[11] M. T. Hagan, H. B. Demuth, M. H. Beale *et al.*, *Neural network design*. Pws Pub. Boston, 1996.

[12] W. S. Sarle, "Stopped training and other remedies for overfitting," in *Proc. of the 27th symposium on the interface of computing science and statistics*, 1995, pp. 352–360.

[13] I. V. Tetko, D. J. Livingstone, and A. I. Luik, "Neural network studies. 1. comparison of overfitting and overtraining," *Journal of chemical information and computer sciences*, vol. 35, no. 5, pp. 826–833, 1995.

[14] Y. Grushka-Cockayne, V. R. R. Jose, and K. C. Lichtendahl Jr, "Ensembles of overfit and overconfident forecasts," *Management Science*, vol. 63, no. 4, pp. 1110–1130, 2016.

[15] B. M. Williams, "Multivariate vehicular traffic flow prediction: Evaluation of arimax modeling," *Transportation Research Record*, vol. 1776, no. 1, pp. 194–200, 2001. [Online]. Available: https://doi.org/10.3141/1776-25

[16] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[17] J. R. Trapero, N. Kourentzes, and R. Fildes, "On the identification of sales forecasting models in the presence of promotions," *Journal of the Operational Research Society*, vol. 66, no. 2, pp. 299–307, 2015. [Online]. Available: https://doi.org/10.1057/jors.2013.174

[18] S. Jaipuria and S. Mahapatra, "An improved demand forecasting method to reduce bullwhip effect in supply chains," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2395 – 2408, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417413007926

[19] V. Murlidharan and B. Menezes, "Frequent pattern mining-based sales forecasting," *Opsearch*, vol. 50, no. 4, pp. 455–474, 2013.

[20] J. Engström, "Forecasting campaign sales using artificial neural networks," *Master's Theses in Mathematical Sciences*, 2016.

[21] C. P. Da Veiga, C. R. P. Da Veiga, A. Catapan, U. Tortato, and W. V. Da Silva, "Demand forecasting in food retail: A comparison between the holt-winters and arima models," *WSEAS transactions on business and economics*, vol. 11, no. 1, pp. 608–614, 2014.