

# Research on retail demand forecasting based on deep learning

Shuai Mi<sup>a</sup>, Jintian Ge<sup>\*bc</sup>, Hongmei Yan<sup>c</sup>, Chunxin Dong<sup>a</sup>

<sup>a</sup>School of Information Science and Engineering, University of Jinan, Jinan 250022, Shandong, China; <sup>b</sup>Industry School of Standardization, University of Jinan, Jinan 250002, Shandong, China;

<sup>c</sup>Business School, University of Jinan, Jinan 250002, Shandong, China

## ABSTRACT

In the retail sector, accurate and efficient forecasting of sales for a wide range of products is a vital part of business operations, especially for subsequent sales optimization. This makes extremely high demands on forecasting methods or models. To enable more accurate predictions, we propose an AggDeepar model. Real sales data are used to make predictions. Experimentally, the AggDeepar model has greater applicability and higher accuracy in sales volume time series forecasting.

**Keywords:** Retail demand forecasting, DeepAR, time-series forecasting

## 1. INTRODUCTION

In the retail sector, timely response to customer demand and control of inventory costs are vital to the retail sector. Inaccurate forecasting can be costly to a company's operations, including lost sales and excess inventory due to stock-outs<sup>1</sup>. And the cost of lost sales is much higher than the cost of excess inventory<sup>2</sup>. So, how predicting multiple products accurately and efficiently is, therefore, a key issue.

Many characteristics of retail products are represented by static variables, i.e., their values do not change over time. So, it is necessary to construct a model that can fuse multiple time series features from different products and shops to learn from the global model. However, many of the methods currently studied follow the classical approach, i.e., the parameters of the new model need to be fitted separately for each time series without including these variables<sup>3</sup>. In addition, many of the machine learning models used in these studies require extensive manual tuning of parameters to accommodate the task of learning global models from multiple time series, this makes it difficult to generalize complex systems and to process large numbers of time series quickly, and the results of machine learning models are very dependent on how the inputs are pre-processed. So, using machine learning models for retail demand forecasting is not appropriate.

Due to the dominance of deep learning in applications such as image recognition, there has been recent interest in the field of time series prediction. The use of deep learning also reduces the complexity of the system and improves maintainability from a system perspective. Therefore, an AggDeepar model is proposed to forecast sales data for retailer demand forecasting. This paper develops an experiment on data from a retailer for the years 2014-2019 and experimentally compared with various models such as LightGBM<sup>4</sup>, ARIMA<sup>5</sup>, LSTM<sup>6</sup>, and Holt-Winters<sup>7</sup>. The results show that the proposed AggDeepar model has better results in forecasting the company's sales.

## 2. RESEARCH METHODOLOGY AND THEORY

### 2.1 Retail supply chain

Depending on the objective of the forecast, all demand forecasts in the retail sector depend on the degree of aggregation of products, locations, or periods. For the retail sector, forecasts are generated based on suppliers, distributors, retailers, individual product categories, and individual SKUs. While these different dimensions of forecasting share many of the same issues, such as seasonality and trends. However, the focus of their forecasts will differ depending on the subject of the forecast, they will have different objectives, and data characteristics.

2.1.1 Supplier Level Demand Forecasting. Supplier level forecasts relate to the total sales of the entire business and the period for this level of forecasting can be monthly, quarterly, or annually. The supplier level is shown in Figure 1. To enable suppliers to plan and operate their retail business effectively at a strategic level, it is necessary to forecast demand at the

\* se\_gejt@ujjn.edu.cn

supplier level<sup>8</sup>. Supplier level demand data shows strong trends and seasonal variations. Unlike at the supplier level, data at the distributor and retailer levels involve irregular observations over time and a large number of zero values.

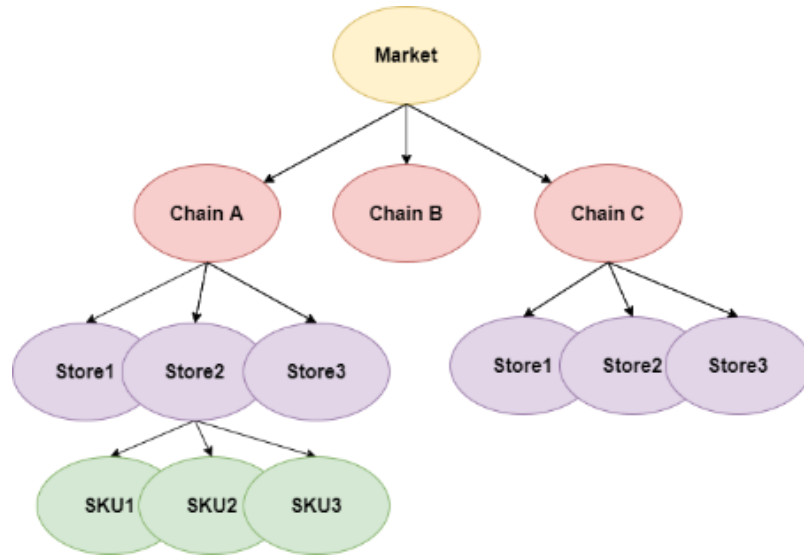


Figure 1. Supplier level.

## 2.2 DeepAR

DeepAR is a deep neural network model which builds a single global model<sup>9</sup>. Traditional time series forecasting methods (ARIMA, Holt-Winters) tend to model the single time series, and it is difficult to use extra features. By comparison, DeepAR models can easily take additional features into account. In the retail sector, for example, if future product sales probability distributions are known, the optimal purchasing volume for different business objectives can be derived using operations optimization methods to aid decision-making. DeepAR accepts as input the past series and its covariates. Let  $z_{i,t}$  denotes the value of the  $i$ th series at time step  $t$ , and  $x_{i,t}$  are covariates. The time point  $t_0$  is used as the division time, indicating the start of the forecast. The objective of the model is to obtain the joint conditional probability distribution. As shown in Figure 2, DeepAR predicts the probability distribution of  $z_{i,t}$  based on an autoregressive recurrent neural network, denoted by the likelihood function  $l(z_{i,t}|\theta_{i,t})$ . In turn, the parameters of the likelihood function are calculated  $\theta_{i,t} = \theta(h_{i,t})$ . Network parameters are updated by maximizing natural functions.

$$P(z_{i,t_0:T} | z_{i,t_0-1}, x_{i,1:T}) \tag{1}$$

$$L = -\sum_i \sum_t \log l(z_{i,t} | \theta(h_{i,t})) \tag{2}$$

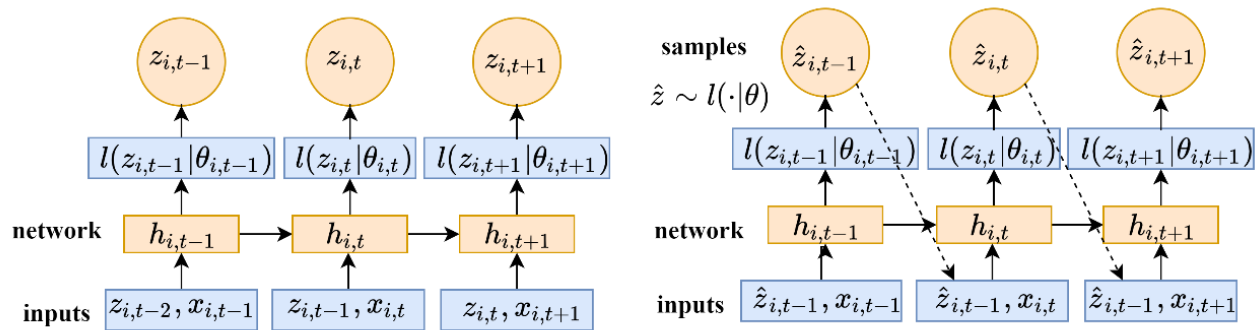


Figure 2. DeepAR network architecture.

### 2.3 Tweedie loss

At the retailer level, some of the time series includes many zero values, whereas a Poisson or binomial distribution may underestimate the probability of a value of zero. To alleviate the problem of too many zero values in the time series, the Tweedie distribution<sup>10</sup> is used as the target distribution in this paper. As shown in Table 1,  $p=1$ , Tweedie is the Poisson distribution, when  $p=2$ , Tweedie is the Poisson distribution. When  $1 < p < 2$ , it becomes a composite distribution of Poisson and Gamma.

Table 1. Tweedie distribution.

Tweedie EDMs	$p$	$V(\mu)$	$\phi$
Normal	0	1	$\sigma^2$
Poisson	1	$\mu$	1
Poisson-Gamma	$1 < p < 2$	$\mu^p$	$\phi$
Gamma	2	$\mu^2$	$\phi$

We use Poisson-Gamma distribution in this paper. The distribution is the sum of  $N$  independent identically distributed random variables sampled from the gamma distribution, where the number of samples to be added ( $N$ ) is the Poisson distribution variable.

$$\begin{aligned}
 N &\sim \text{Poisson}(\lambda) \\
 X_n &\sim \text{Gamma}(\alpha, \beta) \\
 Y &= \sum_{n=1}^N X_n
 \end{aligned} \tag{3}$$

## 3. AGGDEEPAR MODEL CONSTRUCTION

### 3.1 Dataset

In this paper, five consecutive years of sales records of a retailer are selected and aggregated at different levels to produce a time series of sales for each distributor and a time series of sales for the manufacturer.

As shown in Figure 3, the sales data show a clear trend and a certain seasonal pattern, with some randomness. The sales data is then extracted from the distributors and averaged over a 7-day window for the demand data and rolled up. The smoothed data shown in Figure 4 gives a visual indication of the trend.

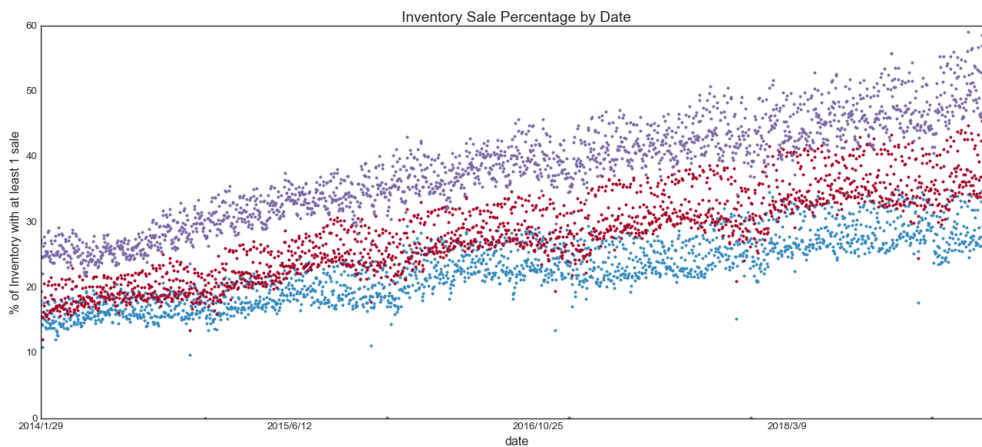


Figure 3. Inventory sale percentage by date.

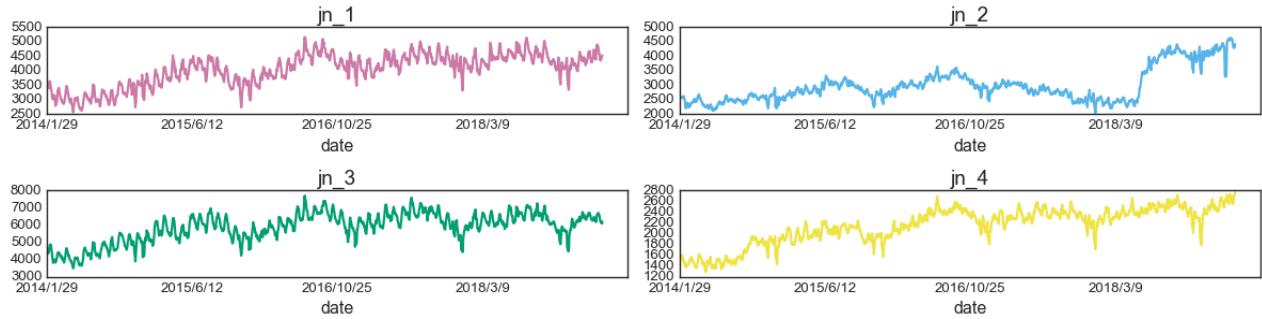


Figure 4. Rolling 7-day demand count by store.

As shown in Figure 5, there are a large number of zero values in the specific SKU sales data, reflecting the intermittent demand for that SKU.

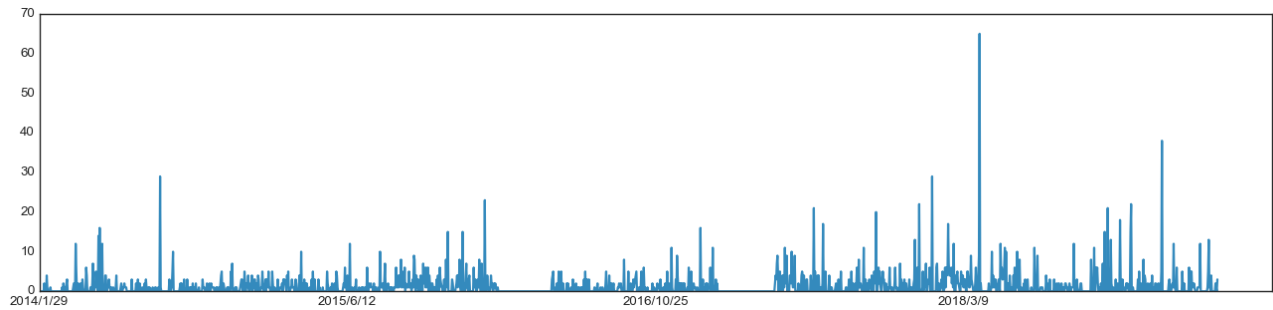


Figure 5. Sales data for an SKU.

### 3.2 AggDeepar model construction

After analyzing the data of this enterprise, this paper proposes the AggDeepar model by taking advantage of the gradient advancement model and the autoregressive neural network model for the data characteristics of different levels of data and the special situation of the retail industry. The hierarchy is divided into two levels according to the needs of different retailers at different levels. The first level is the sales of individual SKUs, as well as individual distributors, and the data at this level shows intermittent trends similar to counting. The second hierarchy is the manufacturer's sales, where the data shows a continuous trend.

For the first level, we use an autoregressive neural network, modeled directly with a modified DeepAR, which uses a recursive prediction strategy in the forecasting process, i.e., samples drawn from the parametric distribution at the previous time step are considered as input to infer the distribution parameters at subsequent time steps. Based on the samples extracted from the parameterized distribution, the quantile of each time step in the prediction horizon was calculated. The target parameter distribution is the Tweedie distribution.

$$X_i \sim \text{Gamma}(\alpha, \beta) \quad (7)$$

$$\sum_{i=1}^N X_i \sim \text{Gamma}(\alpha, \beta) \quad (8)$$

As shown in Figure 6, for the second level, to improve the forecast values, we trained a model, which was trained using Poisson loss, assuming that the retail data are close to a Poisson distribution. The predictions were combined with the deeper model.

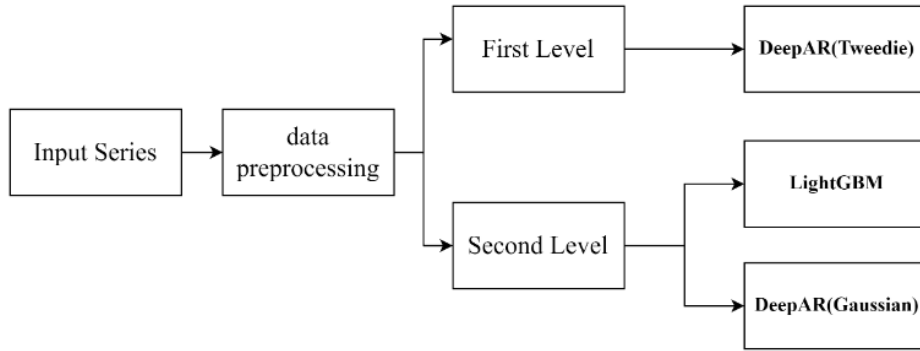


Figure 6. AggDeepar model.

### 3.3 Experiment

In this paper, we choose the RMSE as the criterion for evaluating the model.

$$RMSE = \left(\frac{1}{n} \sum_{i=1}^n |y - \hat{y}|^2\right)^{1/2} \quad (9)$$

The parameters of the model are set as shown in Table 2. To further verify the application performance of the AggDeepar aggregation model, Holt-winter, ARIMA model, LSTM model, and DeepAR (Gaussian) were used as the comparison models in this paper.

Table 2. DeepAR (Gaussian) and DeepAR (Tweedie) model parameters.

Hyperparameter	DeepAR (Gaussian)	DeepAR (Tweedie)
RNN Layers	2	4
RNN cells per layer	40	100
Dropout rate	0.1	0.1
Learning rate	0.001	0.001
Batch size	128	128

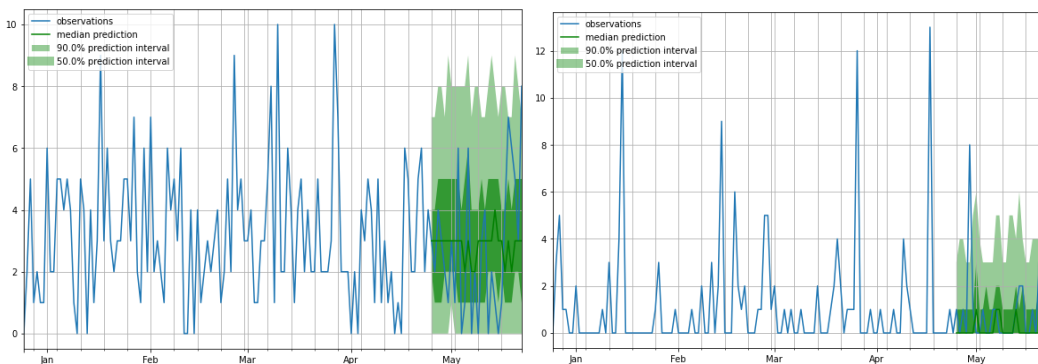


Figure 7. Forecast results for two commodities.

As shown in Table 3 and Figure 7, the AggDeepar model has improved relative to the other models, indicating that the combined forecasting results of the model are relatively valid and more applicable than a single model in forecasting the sales of the business. Traditional statistical models do not predict multiple time series very well and do not capture the characteristics of the time series very well. Experiments have demonstrated that traditional models can only predict a single time series, and difficult to operate with a large number of time series. This proves that for retail sales forecasting, the

traditional approach is no longer suitable and that AggDeepar, a deep learning model that learns a global model, is more practical. AggDeepar also fits similar products to new products based on their characteristics for the cold start problem.

Table 3. Experimental results.

<b>Model</b>	<b>RMSE</b>
Holt-winter	30.283
ARIMA	19.023
LSTM	4.324
DeepAR (Gaussian)	2.423
LightGBM	2.243
AggDeepar	<b>2.083</b>

#### 4. CONCLUSION

With the explosion of data volumes and the digital development of the enterprise, the retail industry is changing dramatically, which makes data-driven demand forecasting a source of decision making for companies. In this paper, the proposed AggDeepar model, based on sales data from one company, has significantly higher forecasting performance than typical time series forecasting models. In the retail industry, there are multiple similar time series across a representative set of units. AggDeepar goes through multiple similar time series and uses a neural network to learn related properties within similar time series. In the future, microdata will be integrated into more aggregate demand forecasts as more and more data are associated with “big data” generated from observed consumer behavior.

#### REFERENCES

- [1] Kourentzes, N., Trapero, J. R. and Barrow, D. K., “Optimising forecasting models for inventory planning,” *International Journal of Production Economics* 225, 107597 (2020).
- [2] Radasanu, A. C., “Inventory management, service level and safety stock,” *Journal of Public Administration, Finance and Law* (09), 145-153 (2016).
- [3] Salinas, D., Flunkert, V., Gasthaus, J. and Januschowski, T., “DeepAR: Probabilistic forecasting with autoregressive recurrent networks,” *International Journal of Forecasting* 36(3), 1181-1191 (2020).
- [4] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al., “Lightgbm: A highly efficient gradient boosting decision tree,” *Advances in Neural Information Processing Systems* 30, (2017).
- [5] Box, G. E. and Jenkins, G. M., “Some recent advances in forecasting and control,” *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 17(2) 91-109, (1968).
- [6] Siami-Namini, S., Tavakoli, N. and Namin, A. S., “A comparison of ARIMA and LSTM in forecasting time series,” 2018 17th IEEE Inter. Conf. on Machine Learning and Applications (ICMLA), 1394-1401 (2018).
- [7] Chatfield, C., “The Holt-winters forecasting procedure,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 27(3), 264-279 (1978).
- [8] Alon, I., Qi, M. and Sadowski, R. J., “Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods,” *Journal of Retailing and Consumer Services* 8(3), 147-156 (2001).
- [9] Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M. and Callot, L., “Criteria for classifying forecasting methods,” *International Journal of Forecasting* 36(1), 167-177 (2020).
- [10] Tweedie, M. C., “An index which distinguishes between some important exponential families,” *Statistics: Applications and New Directions: Proc. Indian Statistical Institute Golden Jubilee Inter. Conf.* 579, 604 (1984).