Improvement of Demand Forecasting Accuracy: A Method Based on Region-Division [*]

Guoshan Liu[1,†]   Yuanyuan Lu[1,2,‡]

[1]School of Business, Remin University of China, Beijing, P.R.China, 100872
[2]Mathematical College, Jilin Normal University, Siping, P.R.China, 136000

Abstract   The key factor to increase enterprise profits and reduce the logistic costs is scientific and reasonable logistics demand forecasting, the accuracy of which can directly influence the effect of decision-making for an enterprise. In some instance, a single customer's demand is irregular, while the demand in a region is comparatively more stable, so it can be better forecasted. In this paper, we have proposed a method to form regions in order that the demand in these regions can be more efficient forecasted while the error of transportation cost caused by replacement of customers by regions can be controlled. Each formed region consists of adjacent customers with similar unit transportation cost to all distributors. Numerical test with data form Northeast Subsidiary Company of China National Petroleum Corporation shows that the method can improve forecast accuracy efficiently.

Keywords   logistics; forecasting accuracy; region-division; clustering algorithm

1   Introduction

With the boost in social economics and science development, logistics has been considered as the "third source of profit" after the reduction of raw material consumption and improvement of productivity. To lower logistic costs and increase profit margin, an enterprise must be able to reasonably forecast her future logistics demand.

At present, there are three typical logistics demand forecasting methods: Time series model ([1]), Cause-and-Effect model ([2]) as well as Judgmental forecast model ([3]). Excepted for the classical forecasting method narrated above, many other new methods have been brought forward, for example, grey forecasting model([4]), combine forecasting model ([5]), and so on.

These forecasting methods, however, either ([6,7]) emphasize the improvement of forecasting technology, or focus on forecasting one single and specific customer demand, or just make partial cause and effect analysis. To some enterprises of large scale, say, enterprises with many distribution centers in petroleum industries, the problem of demand forecasting is far more complicated. Because the petroleum product is mainly delivered

---

[†]Email: luyuanyuan_78@163.com
[‡]Correspondence author. Email:liuguoshan@vip.sina.com.

by trainčňbulk buying becomes economical. To reduce freight, some customers of small demand would choose to purchase only once or several times a year and to purchase from the distribution center nearest to themselves, resulting in a huge fluctuation in demand. Even for those bigger customers, their demands are subject to a series of factors such as the sales, transportation, and particular environment, etc, thus making future demand difficult to predict. If the planning department of the enterprises forecasts future demand base on the existing techniques and individual customer's demand statistics, the forecast would be undoubtedly inaccurate, it would in turn cause either under-stock or overstock. An effective petroleum product demand forecasting method thus becomes even more important. At present, research dedicated to the improvement of forecasting accuracy of large-scale enterprises is still largely neglected.

To improve the low accuracy of demand forecasting, we proposed in the current article a method to form regions in order that the demand in these regions can be more efficiently forecasted while the error of transportation cost caused by replacement of customers by regions can be controlled. Each formed region consists of adjacent customers with similar unit transportation cost to all distributors. Using the MATLAB program, we will illustrate the effectiveness of our method based on the example of petroleum product distribution case of Northeast Subsidiary Company (hereinafter referred to as NSC) of China National Petroleum Corporation (CNPC).

2 Preliminary

NSC is a subordinate regional company of CNPC. It is also one of the largest commodity circulation enterprises in the northeast of China. One of its important tasks is to make the product oil dispatching plan (such as yearly plan, seasonal plan, and monthly plan). Since the main cost of the company comes from transportation expenditure, to make a good petroleum product dispatching plan becomes a top priority to the management. The prerequisite to such a plan is, however, based on the capability of accurately forecasting petroleum product demand.

In view of the obvious seasonal fluctuation of product oil demand, we adopt the Winters Seasonal Exponential Forecasting model ([8]) to forecast the monthly petroleum product demand of NSC. Established by P.R. Winters in the 1960s, the model is based on three mathematical equations, they are listed below:

$$S_t = \alpha(y_t/I_{t-L}) + (1-\alpha)(S_{t-1} + T_{t-1})$$
$$T_t = \lambda(S_t - S_{t-1}) + (1-\lambda)T_{t-1}$$
$$I_t = \gamma(y_t/S_t) + (1-\gamma)I_{t-L}$$

Forecasting formula of Winters Seasonal Exponential Forecasting model:

$$y_{t+m} = I_{t-L+m}(S_t + mT_t)$$

Where is the value of the time series at time t. S_t is the estimated level component of the series and is smoothed using the parameter α. The trend component is T_t, which is smoothed using the parameter λ. The seasonal factors are denoted by I_t, L is the number of periods in one season. The seasonal factors are smoothed separately from the level and

trend component by γ. Where m is a number of periods in the future. The forecasting error is below:

$$MSE = (\sum_{t=1}^{n} (Y_t - y_t)^2)/n$$

or

$$SDE = \sqrt{(\sum_{t=1}^{n} (Y_t - y_t)^2)/n}$$

3 Region-division Principle and Algorithm-Frame

For the purpose of improving the accuracy of the petroleum product demand forecasting, we put forward an idea of dividing all NSC's customers according to the principle of region-division which is listed below:

First, determine an area where distribution centers do not need to distribute goods and customers fetch their goods. We will call it special area. Each petroleum product distributor has the area of its own, and the area-size can be determined by a parameter. Customers within this area are not subject to sales region division. Second, cluster the customers with identical or similar unit freight into one region. Computer program will be designed to compare the unit freight difference from each distribution center to each customer. By calculating the degree of similarity in unit freight cost, we can cluster together those customers with similar unit freight to all the distributors. Third, divide sale regions in accordance with their geographic position. Each formed region consists of adjacent customers with similar unit transportation cost to all distributors. Forth, decide the proper size of the divided sales region. If the size of sales region is divided too small, it only includes single or a few customers, the demand forecast error will increase due to the comparatively irregular demand of the single customer in this region; whereas too big a sales region will considerably debilitate the estimation accuracy of total distribution cost as a result of the huge diversities in freight from each distribution center to numerous customers within the same region.

Denote the set of all the distributor centers as $\{a_1, a_2, \cdots\cdots, a_m\}$, the set of all the customers as $\{b_1, b_2, \cdots\cdots, b_n\}$. Let x_{ij} be the unit freight value from the distribution center i to the customer j, y_{ij} be the distance between the distribution center i to the customer j. We could thus calculate:

(1) For any two customers b_i and b_j, the similarity degree of the unit freight to all the distribution centers is:

$$r_{ij} = \sqrt{\sum_{k=1}^{m} (x_{ki} - x_{kj})^2}$$

(2) R stands for the neighbor relationship, any two customers b_i, b_j border each other if and only if $(b_i, b_j) \in R$, adjacency matrix is denoted by:

$$W(i, j) = \begin{cases} 1 & (b_i, b_j) \in R \\ 0 & else \end{cases}$$

If $\exists b_i \in D'$, $\exists b_j \in D''$ and $(b_i, b_j) \in R$, then we can get $(D', D'') \in R$.

(3) Demand forecasting error e_i of the region i and the weighted total error are defined as $e_i = \sqrt{\left(\sum_{j=1}^{t} (Y_j^{(i)} - y_j^{(i)})^2\right)/t}$ and $E = \sum_{i=1}^{n} (c_i/c)e_i$.

Where $Y_j^{(i)}$ is the demand observed value, namely the average of observed demand value of all the customers in the i-th region at the j-th period. $y_j^{(i)}$ is the forecast demand value, c_i is the sum of the customers' demand during the whole forecasting periods in the region i, c is the sum of all the customers' demand during the whole forecasting periods, n stands for the total number of sales regions.

(4) Unit freight difference s_l of the region l and the weighted total difference value are defined as $s_l = \max \{ |x_{ki} - x_{kj}| : \forall b_i, b_j \in D^l \}$ and $S = \sum_{l=1}^{n} (c_l/c)s_l$.

Where c_l is the sum of the customer demand in the region l, c is the sum of all the customers' demand, n stands for the total number of sales regions.

At the beginning of sales region division, each customer (except the customers in the special area) is regarded as a separated region. We then clustered together those who met the clustering requirement to form a new sales region and to calculate its demand forecasting error e, unit freight difference s, corresponding weighted demand forecast error E and weighted total unit freight difference S respectively. If the above parameters are within the acceptable limits, we can stop calculating and accept the new division results of sales regions. Otherwise the calculation process should not halt until we find a satisfying dividing way. Concrete steps as follow:

Step1: Suppose $\ddot{D}$ as the set of all customers, we can determine the special area D_i of each distribution center based on the distance parameter $\bar{\delta}$ given by the distributor:

$$D_i = \{b_j | y_{ij} \leq \bar{\delta}, j = 1, 2 \cdots n\}, i = 1, 2 \cdots m, D^1 = \ddot{D} - \bigcup_{i=1}^{m} D_i$$

Region D^1 is a big region that includes all the customers, and D_i is the special area .

Step2: Select the unit freight difference error η and demand forecasting error value β. Choose each customer in D^1 as a single-point region, and $\mu(1)$ is the totality of customers here. Let $k = 1$.

Step3: $D^k = \bar{D}^{k-1}$, $\bar{D}^0 = D^1$. The region D^k stands for a big region that we will divide according to region-division principle as follows, and $\bar{D}^k$ is the big region after division at step k. We can calculate the similarity between $r(D', D'')$ any two regions belong to D^k:

$$r(D', D) = \max\{r_{ij} | b_i \in D', b_j \in D\}(D', D'' \in D^k)$$

Search the minimum similarity value between any two adjacent regions in region D^k, if $\bar{r}_k$ is the minimum similarity value between region D' and D'', D' and D'' and are neighbors with each other, that is $(D', D'') \in R$, then the two regions can be combined to form a new small region, so we can get the new big region $\bar{D}^k$.

Step4: Calculate the observed demand value $Y_j^{(i)}(k)$, $j = 1, \cdots, t$, $i = 1, \cdots, \mu(k)$ of each small region in the region $\bar{D}^k$ ($\mu(k)$ is the variable that represents the amount of all the small regions in D^k).

Step5: Calculate the demand forecast value $y_j^{(i)}(k)$ and $e_i(k)$ in the big region $\bar{D}^k$ with the Winters seasonal exponential forecasting model, and compute the total demand forecasting error $E(k)$ at k of regional division.

Step6čžCalculate the unit freight difference value $s_i(k)$, $i = 1, \cdots \mu(k)$ of each small region in the region $\bar{D}^k$, and export the weighted total unit freight difference $S(k)$. If $E(k) \leq \eta$, $S(k) \leq \beta$ the procedure terminates, otherwise $k = k + 1$, and turn to step3.

4   An Application

The present paper was based on NSC's statistics of freight cost, transportation distance (including 213 customers and 13 petroleum product distribution centers in all), and gasoline sales statistics in 1999, 2000 and 2001. Monthly gasoline demand of each customer in 1999 was taken as the initial value to forecast the annual demand in 2000 and 2001. 100 was chosen as the parameter value of the special area. The values of α, γ and *ambda* were 0.3, 0.1, 0.1 respectively.

We then used MATLAB software package to automatically divide NSC's sales regions. According to the above-mentioned method, when the unit freight difference error η and demand forecasting error β value were big enough, MATLAB program could calculate out all region division results. In the present article, we chose the region division results from times 1 to 140. and traced out the trend of demand forecasting error varying with division times, as shown in Fig. 1; and the trend of unit freight difference error varying with division times, as shown in Fig. 2.

The MATLAB results showed that, at the first time of division, each customer consisted of a separate sales region, we had the greatest number of sales regions, the demand forecasting error was as large as 57562.3821 and the unit freight difference error was 0 (see table 1), whereas at the 140 divisions, the demand forecasting error was reduced to 496.6065, and unit freight difference error was enlarged to 3.9759. As shown in Fig.1, we could see that the demand forecasting error decreased in the division times, which shows the method could efficiently enhance the forecasting accuracy. However, as shown in Fig.2, the unit freight error also increased in the division times accumulated.

Table 1: Demand Forecasting Error and Unit Freight Difference.

times	Demand Forecasting Error	Unit Freight Difference Error
1	57562.3821	0
20	53950.9856	0.0087
40	35636.0608	0.0713
60	25496.3129	0.1742
80	21193.1344	0.3694
100	2454.7662	0.9143
120	834.2577	1.5951
140	496.6065	3.9759

The case of NSC illustrated that neither too few nor too many divisions were desirable: too few divisions would lead to big forecasting errors, resulted from the volatile demands from small size areas, whereas too many divisions would lead to huge freight
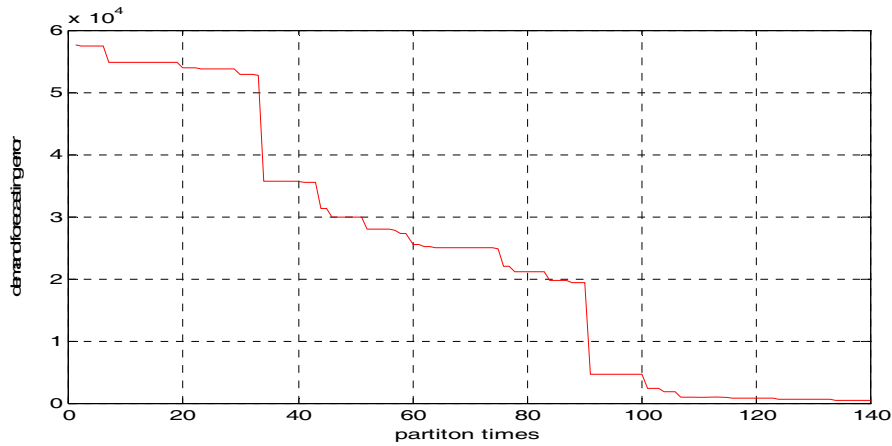
Figure 1: Change Trends of Demand Forecasting Error

difference from one distribution center to each customers within the same sales region. As a result, the estimate on the total freight in that area would be affected. To draw up a better dispatching plan, it was necessary to select appropriate η and β in accordance with specific situation of their own, so as to get a proper sales region division results based on the above-mentioned method.
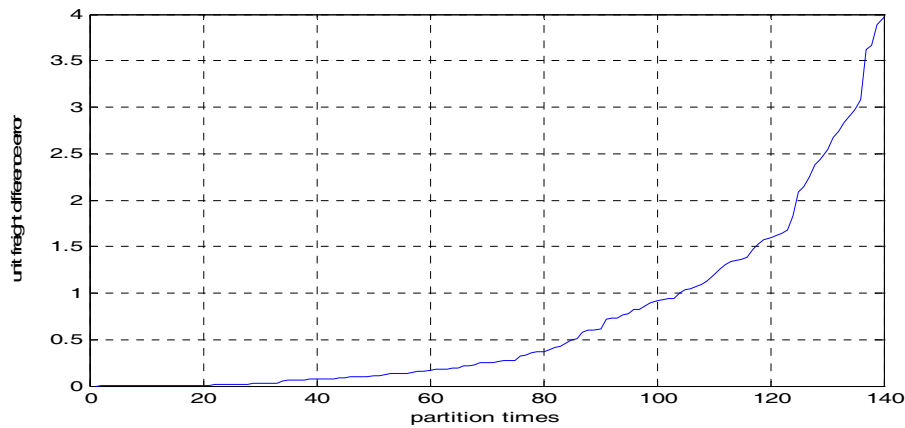

Figure 2: Change Trends of Unit Freight Difference Error

5 Conclusion

In the present paper, we proposed a new method of enhancing logistics demand forecasting accuracy by means of sales region division, and we discussed the problem of petroleum product demand forecasting in NSC. With the help of MATLAB program, our study illustrated that the region division method could greatly improve forecasting ac-

curacy efficiently. In addition, users could select region dividing results freely based on their specific situation. We therefore believed that our method is of great practical value for users.

References

[1] George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, 1999. *Time Series Analysis: Forecasting and Control*. China Statistics Press.

[2] Latif. Abdul, King. Maxwell L, 1993. Linear regression forecasting in the presence of AR (1) disturbances. *Journal of Forecasting*,12(6) 513-524.

[3] Corolyn E Clark, Paul L Foster, Karen M Hogan, George H Webster, 1997. Judgmental approach to forecasting bankruptcy. *The Journal of Business Forecasting Methods & Systems*, 16(2) 14 -18.

[4] M Mao, E C Chirwa, 2005. Combination of grey model GM (1, 1) with three-point moving average for accurate vehicle fatality risk prediction. *International Journal of Crashworthiness*, 10(6) 635-642.

[5] James H. Stock, Mark W, 2004. Combination forecasts of output growth in a seven-country data set Watson. *Journal of Forecasting*, 23(6) 405-430.

[6] Luh-Yu (Louie) Ren, 2007. Revised Mean Absolute Percentage Errors (Mape) for Independent Normal Time Series. *Journal of American Academy of Business*, 10(2) 65-70.

[7] Michael W. Babcock, Xiaohua Lu, 2002. Forecasting inland waterway grain traffic. *Transportation Research*, Part E (38) 65-74.

[8] Gardner, Everette S., Jr., McKenzie, Ed., 1989. Seasonal Exponential Smoothing With Damped Trends. *Management Science*, 35(3) 372-376.