European Economic Letters
ISSN 2323-5233
Vol 13, Issue 1 (2023)
http://eelet.org.uk

Constructing a Collaborative Planning, Forecasting, and Replenishment based on Data Mining for Refrigerated Beverage Industries

Sheng Yuan Hsu

Graduate Institute of Services and Technology Management,Chienkuo Technology University,Changhua, Taiwan, ROC

jackhsu@ctu.edu.tw

Abstract—Based on the end demand forecast, this research proposes two executive strategies of CPFR. Strategy one, satisfy the demands of customer orders and takes consider of sales forecasting. Strategy two, only considers the sales forecasting. The simulation analysis is conducted respectively in two strategies, and the inventory level will be discussed. Before CPFR, the average inventory is 55,878 in production plant, 220,845 in sales channel. Both of the two strategies of CPFR can reduce the stock significantly. If the CPFR mechanism is applied according to the end demand forecasting (strategy 2), it is estimated that the inventory of production and inventory of channels can be effectively reduced to 37,915 (decrease 32%) and 24,234 (decrease 89%) respectively after the simulation analysis.

Index Terms—forecasting model, cold drinks, regression model

1. INTRODUCTION

As to the business management of the chain supermarkets, purchasing and predicting future sales are necessary daily work projects for managers. Thus, each store has introduced a Point of Sales System (POS system, or point of sale management system) to assist carrying out routine operations of each sale point, such as purchases, sales, and inventory. In addition, the system provides the function of product sales prediction as well as recommendation and estimation of popular products, and it reminds the store to adjust the inventory during the specific festivals or regional events. It reveals that the prediction of sales belongs to one of decision-making task in the management level. [1] pointed out that the correct demand prediction plays and extremely critical role in the retail industry, especially cold or refrigerated fresh food products. In other words, if the ability of sales prediction is not good, the insufficient quantity will not be able to meet the needs of consumers, which would be resulting in fleeting business opportunities, while the excessive quantity will cause a backlog of funds and a high inventory. As a result, inaccurate of prediction will have a significant impact on the profitability of a company and it would even weaken the market competitiveness.

Traditionally, supply chain members have tended to maintain safe inventory buffers on their own against possible changes in supply and demand at both ends of the operation. For example, a manufacturer can maintain a buffer of one to three to two weeks of production raw materials, while also maintaining a one to two-week inventory of finished goods. Under the framework of information opacity and transaction object relationship, retailers may also maintain a considerable level of inventory at the same time. Inventory prepared for safety throughout the supply chain may far exceed customer needs. pressure, which in turn erodes profits. In fact, inventory cost is the main indicator that reflects the ability to adapt to changes in supply and demand. The more unpredictable the demand, the greater the inventory risk to take. Likewise, the further away the inventory buffer links in the value chain process are from consumers, the more uncertain the demand facing the inventory buffer is ([2]). According to [3], data mining is part of Knowledge Discovery in Databases (KDD). It is a nontrivial process of pointing out valid, new and potential benefits in data, with the ultimate goal of understanding the style of the data. As defined by [4], data mining is the analysis of large amounts of data in an automated or semi-automated manner to find meaningful relationships or rules. It can be understood from the above definition that the basic spirit of data mining is to use relevant analysis techniques to discover new and unknown patterns or rules from data or databases, and to discover beyond the scope of data through the application of data mining. Induce external data relational schemas ([5]).

The application scope of data mining is more and more extensive. At present, the main field is mainly enterprise business application, and we can find favorable information in various ways. [6] divided the application areas into three areas: marketing management, risk management and fraud management. [7] solve the management forecasting problem using data mining. Under CPFR's business partner framework, through the sales and order forecasting process and the configuration operation in the freezing stage, the situation variables that members need to face are greatly reduced due to rigorous procedures and transparent information, thus reducing unnecessary inventory. buffer volume, while the full flow is released. Secondly, it can reduce the inventory pressure caused by the bull-whip effect ([8]).

To achieve this successfully, retail companies need to design their demand and supply planning processes to avoid customer service issues and mass write-offs due to increased inventory and expired products. These are sensitive issues for retail companies because of the complex relationship of demand data with considerable monthly fluctuations, the presence of many middlemen in the process, the variety of products, and the quality service consumers demand. In the recent literature, it is well known that there is a need to investigate the most appropriate prediction model for each situation, since there is no universal model that can make accurate predictions for all problems. Historically, many scientific papers have focused on modeling, analysis, and forecasting methods for time series that exhibit irregular variability, seasonality, and trends ([9], [10], [11] and [12]). In recent decades, many theoretical and heuristic models have been developed and empirically evaluated ([13] and [14]).

Most of the related literatures on sales forecasting use nerve-like BPN to construct sales forecasting models. The application is more complicated and the cost of analysis is too high as well. In this study, SPSS linear regression analysis software was used to construct a sales forecasting model. The main consideration of using SPSS is that SPSS is a linear regression software which is cheap, simple and easy to understand, easy to apply in practice, and easy to operate. Therefore, SPSS software is selected to predict the actual sales volume. The overall sales forecasting will be carried out from the manufacturing to improve the application scope and accuracy of the forecasting. At the end, it can be served as the basis for production and inventory scheduling.

2. Prediction Model Construction

After research and analysis, it is found that the sales status has a great relationship with the climate status. Therefore, in the section of data collection, the climate and sales information are collected separately. The information collection of the climate is mainly from Yahoo meteorology to collect the daily temperature, rain probability, and weather patterns by districts according to the channels of the warehouses. They are described as follows

Zone temperature: According to the channel of the warehouses, the warehouse is divided into different zones, and the high and low temperature of each zone are collected separately.

Probability of rain: According to the channel of the warehouses, the percentages of rain in each zone are collected separately.

Weather pattern: According to the channel of the warehouses, the weather patterns of each zone are collected by matching the weather type classification of Yahoo from Code 01 to 13. (01 Sunny and cloudy, 02 Sunny but cloudy in the afternoon, 03 Sunny but brief showers in the afternoon, 04 Cloudy and sunny sometimes, 05 Cloudy, 06 Cloudy but brief showers in the afternoon, 07 Cloudy with brief showers, 08 Cloudy with brief showers, 09 Cloudy with brief showers, 10 Cloudy with brief showers or thunderstorms, 11 Cloudy with showers or thunderstorms, 12 Overcast when cloudy, 13 Showers when cloudy)

In addition to the actual sales records, order records, number of order shops, and promotion types provided by the case company, sales data collection also includes channels, divisions (consolidated warehouses) and product items. The above data was respectively collected through the POS system and ordering system of the channels.

Definition of Prediction-related Variables

The prediction variables used in the prediction model are mainly weather-related variables, including predicted temperature, low temperature difference, high temperature difference, rain probability, weather patterns, weather pattern difference, and etc. The definitions are as follows：

Predicted temperature: predicted low temperature (low-temp) and predicted high temperature (high-temp)

Predicted average temperature (avg-temp)：(high-temp + low-temp)/2

Predicted low temperature difference: Predicted low-temp today – Predicted low-temp the day before

Predicted high temperature difference: Predicted high-temp today – Predicted high-temp the day before

Predicted percentage of rain (rain pert): 0~100%

Predicted weather pattern：01~13

Predicted weather pattern difference: Weather pattern today – Weather pattern the day before

Table 1. The definition of prediction-related variables

European Economic Letters
ISSN 2323-5233
Vol 13, Issue 1 (2023)
http://eelet.org.uk

prediction-related variables	definition
low-temp	l-temp
high-temp	h-temp
avg-temp	avg-temp
low temperature difference	l-temp-diff
high temperature difference	h-temp-diff
rain pert	rain-pert
wear-type	wea-type
wear-type_difference	wear-type_diff

3. Forecast Model Construction and Testing

After completing the preliminary data sorting and linear regression test through Excel software, SPSS statistical software was used to perform multivariate regression analysis to construct a complete forecast model in order to improve the accuracy and estimate forecast error assess model quality. Apply the forecast model to the data of each zone in the next year (2010) to predict and evaluate the error status. The error evaluation indicators used ate the mean absolute error value (Mean Absolute Deviation, MAD), the percentage of forecast error and the number of forecast error baskets, etc. The detailed definitions and error analysis examples are described as below：

Mean Absolute Error (MAD) = | Forecast Sales Value – Actual Sales Value|

The percentage of forecast error = (MAD/Actual Sales Value) * 100%

Forecast error (baskets) = MAD/35 bottles (1 basket = 35 bottles which is the base ship unit of the case company)

The error analysis takes the regression model of the peak season of the product A in Zone 1 of the 7-11 channel as an example：

y=52.172x1-79.905x2-69.885x3+176.56x4-1326.381

y：predicted value

x1：avg_temp(average temperature)

x2：l_temp_diff(low-temp difference)

x3：wea_type(weather pattern)

x4：l_temp(low-temp)

Substitute the weather information of the forecast day (x1：24，x2：-5，x3：09，x4：22) and you can get the sales forecast value of the day (y=3581). If the actual sales value is 3587, MAD, the percentage of the forecast error baskets would be calculated as below：

MAD=｜3581-3587｜=6

Forecast error = (6/3587)*100%=0.17%

Forecast error baskets = 6/35 = 0.17 baskets

Bring the collected historical data into the regression model, calculate the MAD, the percentage of forecast error, and forecast error baskets of each district, and conduct model error analysis. Table 2-3 show the results of forecast error analysis on the channels of 7-11 and Family Mart.

European Economic Letters
ISSN 2323-5233
Vol 13, Issue 1 (2023)
http://eelet.org.uk

Table 2. The forecast error analysis on Product A and B of 7-11 channel.

Product A of 7-11 channel								
	Peak season (Mar. – Sept.)				Low season (Oct. – Feb.)			
Zone	R^2	Average of MAD	Forecast error baskets	Percentage of forecast	R^2	Average of MAD	Forecast error baskets	Percentage of forecast
1	0.551	512.86	15	11.86%	0.73	268.18	8	12.14%
2	0.494	333.26	10	12.09%	0.758	206.16	6	15.87%
3	0.432	133.63	4	11.21%	0.727	94.61	3	14.21%
4	0.418	188.87	5	12.18%	0.725	125.11	4	12.58%
5	0.442	136.00	4	11.58%	0.606	97.44	3	13.07%
Product B of 7-11 channel								
	Peak season (Mar. – Sept.)				Low season (Oct. – Feb.)			
Zone	R^2	Average of MAD	Forecast error baskets	Percentage of forecast	R^2	Average of MAD	Forecast error baskets	Percentage of forecast
1	0.406	464.27	13	14.23%	0.808	267.61	8	16.16%
2	0.301	347.69	10	13.60%	0.813	175.00	5	12.24%
3	0.298	151.56	4	18.69%	0.82	63.70	2	13.83%
4	0.325	189.26	5	19.56%	0.797	76.97	2	13.96%
5	0.359	135.45	4	20.20%	0.752	55.83	2	13.94%

Table 3. The forecast error analysis on Product A and B of Family Mart channel.

Product A of Family Mart channel								
	Peak season (Mar. – Sept.)				Low season (Oct. – Feb.)			
Zone	R^2	Average of MAD	Forecast error baskets	Percentage of forecast	R^2	Average of MAD	Forecast error baskets	Percentage of forecast
1	0.769	315.49	9	18%	0.743	134.35	4	12%
2	0.76	96.23	3	13.98%	0.783	48.92	1	11.45%
3	0.726	91.08	3	13.28%	0.766	54.55	2	12.17%
4	0.61	70.80	2	12.39%	0.769	35.97	1	9.35%
5	0.744	30.37	1	18.96%	0.418	21.37	1	18.88%
Product B of Family Mart channel								
	Peak season (Mar. – Sept.)				Low season (Oct. – Feb.)			
Zone	R^2	Average of MAD	Forecast error baskets	Percentage of forecast	R^2	Average of MAD	Forecast error baskets	Percentage of forecast
1	0.63	179.48	5	9.28%	0.795	103.56	3	9.40%
2	0.485	104.22	3	11.80%	0.792	50.20	1	9.27%
3	0.467	58.75	2	9.93%	0.6955	37.1	1	10.01%
4	0.3995	34.76	1	8.69%	0.601	26.15	1	9.41%
5	0.4647	17.18	0	11.24%	0.4305	15.4	0	15.68%

4. Forest Model Application

Using the constructed forecasting model, make a sales forecast for the new year (March to September, 2011), and perform error analysis. From Table 4 of the error analysis on Product A of 7-11 channel, the percentage of forecast error increases year by year within the acceptable range of the case company. Thus, its forecast model can be used as a reference for production scheduling.

Table 4. Table of forecast error analysis on Product A of 7-11 channel.

Product A of 7-11 channel			
Zone	Average of MAD	Forecast error baskets	Percentage of forecast
1	495.3	0	14.57%
2	328.8	9	15.38%
3	173.8	5	20.75%
4	208.4	6	15.28%
5	141.2	4	16.64%

In the future, the forecast model will be integrated with the existing data system of the case company. Through the link of online information, real-time access to Yahoo's weather forecast data for the next three days to provide forecast models performing sales prediction and assist in arranging onsite production scheduling and channel inventory scheduling, and to collect sales data from various channels regularly by updating the forecast model year by year and the future information system operation model.

V. Conclusion

In this study, the actual sales data, from the case company, of the two main cold drinks sold in the two major channels, 7-11 and Family Mart, were collected, and the prediction model was established through SPSS linear regression analysis. With the prediction error analysis including the MAD average absolute error value, the percentage of forecasting error, and the number of forecasting error baskets, etc, the average error of MAD is ±13%, which is in line with the requirements of individual companies. The forecasting model demonstrated for empirical evidence by using the data in March 2011. The value of forecasting error is also within the ideal range of the case company. Therefore, the actual sales value predicted by this study indeed has reference value. For the future study, it will be combined with the existing information system, using online Web-service to directly obtain the weather forecasting information for the further three days, and make sales forecasts to facilitate subsequent production arrangements and inventory scheduling, so as to provide production and sale personnel to make decisions for production and sales in order to reduce the overall operating costs.

REFERENCES

[1] D. Agrawal, C. Schorling, "Market share forecasting: an empirical comparison of artificial neural networks and multinomial logit model," Journal of Retailing, Vol.72 (4), pp. 383-407, 1996.
[2] T.M. Simatupang, and R. Ridharan, "The Collaborative supply chain," International Journal of Logistics Management, Vol. 13(1), pp. 15-30, 2002.
[3] U. Fayyad, G. Piatesky-Shapiro, and P. Smyth, "The KDD Process for Extracting Useful Knowledge from Volumes of Data," Communications of the ACM, Vol. 39(11), pp. 27-34, 1996.
[4] M. J.A. Berry and G. Linoff, "Data Mining Technique for Marketing: Sale,and Customer Support," New York, John Wiley & Sons, Inc, 1997.
[5] P. R. Peacock, "Data mining in marketing: Part I," Marketing Management, Vol. 6(4), pp. 8-18, 1998.
[6] P. Cabena, P. Hadjinian, R. Stadler, J. Verhees, and A. Zanasi," Discovering Data Mining from Concept to Implementation", Upper Saddle River, NJ, Pretice Hall PTR, 1998.
[7] S. Y. Hsu and D. Y. Sha, 2005b,"Due Date Assignment and Management Based on Date Mining," Proceeding of the 2005 CIIE Annual Conference, 374.
[8] B. M. Beamon and T. M. Ware, "A process quality model for analysis improvement and control of supply chain system," Logistics Information Management, Vol. 103(1/2), pp. 105-113, 1998.
[9] L. Aburto, R. Weber, "Improved supply chain management based on hybrid demand forecasts," Applied Soft Computing, Vol. 7, pp.126-144, 2007.
[10] P. Doganis, A. Alexandrids, P. Patrinos and H. Sarimveis, 2006, Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. Journal of Food Engineering, 75, 196-204
[11] S. Thomassey and A. Fiordaliso, "A hybrid sales forecasting system based on clustering and decision trees," Decision Support System, Vol. 42, pp. 408-421, 2006.
[12] S. T. Huang, N. H. Chiu, L. W. Chen, "Integration of grey relational analysis with genetic algorithm for software effort estimation," European Journal of Operational Research, Vol. 188, pp. 898-909, 2008.
[13] J. G. Zhang, X. Y. Ye and Y. Gong, "Forecast research on trend of rural e-commerce industry cluster based on grey neural network combination model," Journal of Commercial Economics, Vol. 1, pp. 192-194, 2017.
[14] N. Wang, "OWHA-based logistics demand combination improvement and forecasting model construction,"Journal of Commercial Economics, Vol. 1, pp. 192-194, 2017.