ISSN: 2688-9323

Forecasting using the ARIMA Model and Frequency-used Machine Learning Models

Song Pu[1,2]

[1]School of Economics and Management, University of Electronic Science and Technology of China, Chengdu, 611731, China

[2]School of Economics and Management, Chengdu Technological University, Chengdu, 610031, China

Abstract

The main challenges of platelets supply chain management is uncertainty in demand due to its very short shelf live. A accurate forecast for the platelet demands can help make a better decision on required products, address supply shortages, and reduce wastage and costs. This paper uses the ARIMA model and Frequency-used machine learning models for the platelet demand forecasting to test the performance of each model. According to a real-word data from the Chengdu blood bank, none of machine learning models have the advantage of the ARIMA.

Keywords

Forecasting; Platelet Demands; ARIMA; Machine Learning Model.

1. Introduction

Platelets are blood cells used by the body in the process of blood clotting, which can be used to treat patients undergoing chemotherapy, AIDS, hypersplenism, bone marrow transplant, organ transplant, etc. In recent years, the demand of platelets is increasing greatly. Statistically, from 2008 to 2011, the production of platelets in American increased by 13% (about 4.277 million doses). In Canada, the collection of platelets increased by 9.6% between 2008 and 2009. In addition, the amount of platelets transported to hospitals increased from 91,600 doses in 2004 to 119,528 doses in 2012. platelets have a shelf live of 5 days, which is shorter than other blood products (35 days for red blood cells , 1 year for frozen blood and Plasma).. Based on statistics, about 8.1% of blood in hospitals and blood banks was expired in United States. Especially, more than 22% of platelets are outdated. Most literature focused on the blood inventory management in hospital or bank, which can reduce blood wastage effectively. Nevertheless, there is still a huge challenge for the platelets on account of its short shelf live [1]. This research predicts the platelet demand by using various models such as ARIMA and many Frequency-used machine learning models to determine whether the machine learning models are better than the ARIMA for the platelet demand forecasting. Besides, a real-world dataset is tested in this study, which provide the baseline performance for future on this research stream.

The rest of this paper is organized as follows. Section 2 reviews related literature. In Section 3, the methodology description is given. Section 4 presents the empirical studies. And the final section gives some conclusions.

2. Literature

Platelet demands may vary from day to day. Because they are used for patients who may be related many uncertain factors. For example, patients suffered from accidents and trauma incidents arises randomly [1]. Most research proposed inventory management of platelets to satisfy demand and reduce wastage in hospitals or blood banks (blood centers)[1, 2]. However, the perishable products still lead to a great challenge to the inventory management in virtue of its short shelf live. Especially, the inventory management for platelets is now still in the exploratory phase at home and abroad. The research about platelet demand forecasting is extremely scarce, while only a few studies are related to blood demand forecasting [3]. The ARIMA model (Box-Jenkins model) is a classic model for time series, which is also commonly used for blood demand forecasting. The ARIMA model is also used for short-term forecast of blood demand in some research. Recently, many machine learning methods have been developed rapidly and widely used for demand forecasting including short-term traffic flow, passenger flow and blood products. Fanoodi et al. demonstrated that the ANN models can significantly improve the prediction of uncertainties in platelet demand [4]. The recurrent neural network (RNN) is a expansion of ANN, which can fully capture characteristics of time-series data. However, RNN cannot well capture long time dependencies in time-series data due to the problems of vanishing and exploding gradients. Lately, the long short-term memory neural network (LSTM) which is a expansion and update of RNN was proposed to overcome the limitations associated with RNN [5].

3. Methodology

Several frequency models including ARIMA, RNN, LSTM, SVR, GBDT and RF are employed for the platelet demand forecasting:

• ARIMA: it is s a classic statistical model designed based on Gaussian white noise.

• RNN: it is an expansion of ANN, which shows better ability in capturing the characteristics of time-series data when compared with the ANN.

• LSTM: it is an update of RNN, which has the advantage in learning the time series with long time span and overcomes the problem of vanishing and exploding gradients.

• SVM: it is a machine learning model based on the statistical learning theory and the structural risk minimization principle. Numerous recent studies suggested that the SVM can achieve high prediction performance for the nonlinear time series prediction.

• GBDT and RF: they are recently developed machine learning models based on decision trees and have produced promising results in creating predictions for large, complex data sets.

4. Experiment

4.1. Data

The proposed forecasting model is applied to the data collected by Chengdu blood center, which supplies the blood service for about 120 hospitals and blood banks in Chengdu. The daily platelet data are collected from January 1st, 2014 to July 2nd, 2017 (see Figure 1). Only the O type platelets are tested since they are universal platelets. However, our methodology can also easily be applied to other types. The original dataset should be divided into two subsets: training dataset and testing dataset. Generally, 80% of the original data are used as training dataset and the others are used as testing dataset [6].

ISSN: 2688-9323

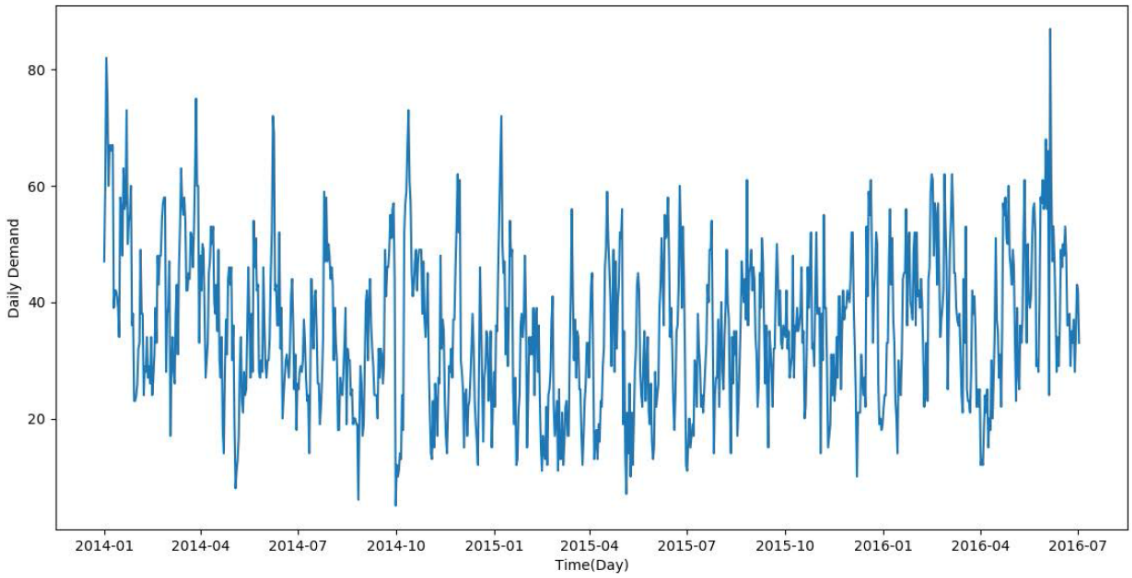


Figure 1. Daily demand from January 1st, 2014 to July 2nd, 2017 in Chengdu blood center

The parameters of ARIMA is determined according to the autocorrelation function (ACF) and minimum Akaike information criterion (AIC) values. And the parameters of RNN, LSTM, SVR including the number of training epochs, batch size, number of nodes, dropout rate and kernel function are determined by the the sensitivity analysis [7].

Two widely used performance indexes have been applied in the literature to evaluate the performance of the blood demand forecast model such as root mean square error (RMSE) and mean absolute percentage error (MAPE) [3, 7]. They are calculated by the following equations, respectively.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - \hat{Y})^2}{n}} \tag{1}$$

$$MAPE = \frac{1}{n}\frac{\sum_{t=1}^{n}|Y_t - \hat{Y}_t|}{Y_t} \times 100\% \tag{2}$$

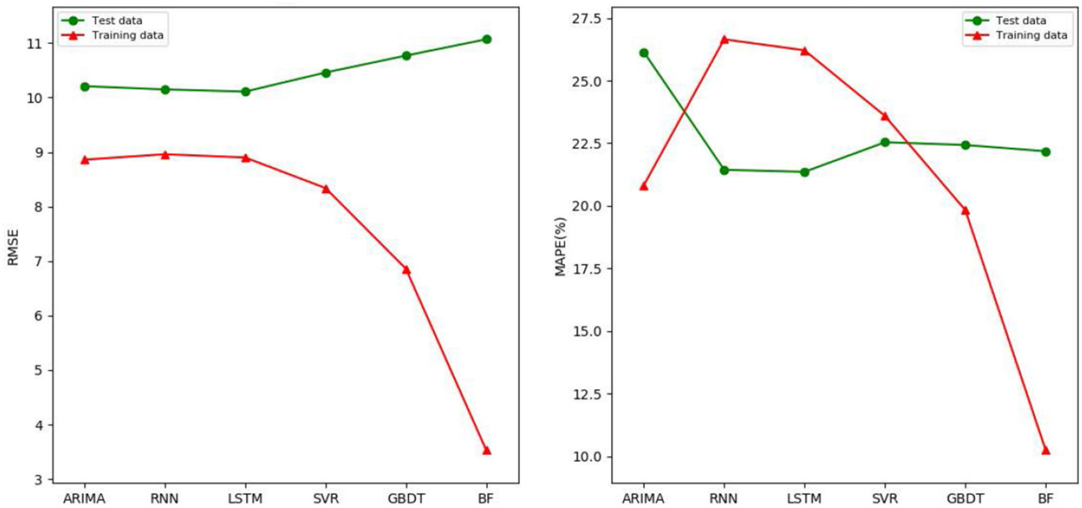
Where $Y_t, \hat{Y}_t$ are the measured data and predicted data, respectively.

4.2. Results and Discussion

The results of different models are summarized in Table 1. As seen from the Figure 2, none of the machine learning models are more accurate than the ARIMA model. Thus, the ARIMA model is suitable for the platelet demand forecasting, which is also one of the proposed models for the platelet demand forecasting by Fanoodi [4]. Besides, the BF seems to be over-fitting, because it has a great performance on the training data with less accuracy on the test data compared with other models.

Table 1. The results of all hybrid models

Methods	Test data		Training Data	
	RMSE	MAPE(%)	RMSE	MAPE(%)
ARIMA	10.21	26.14	8.86	20.82
RNN	10.15	21.44	8.96	26.65
LSTM	10.11	21.36	8.90	26.21
SVR	10.46	22.54	8.34	23.59
GBDT	10.77	22.43	6.86	19.83
BF	11.07	22.18	3.53	10.25



(a) (b)

Figure 2. The RMSE and MAPE for both training data and test data,respectively

5. Conclusion

This study proposes the ARIMA model and Frequently-used machine learning models to predict the platelet demand. According to experiments, none of the machine learning models are more accurate than the ARIMA model. Therefore, the ARIMA model is still suitable for the platelet demand forecasting.

Acknowledgments

Education project of Chengdu Technological University (no. JG2017B13).

References

[1] J.Beliën, H.Forcé (2012) Supply chain management of blood products: A literature review. European Journal of Operational Research, Vol. 1(2012) No.117, p.1–16.

[2] D. Zhou, L.C. Leung, W.P. Pierskalla: (2011) Inventory management of platelets in hospitals: Optimal inventory policy for perishable products with regular and optional expedited replenishments. Manufacturing& Service Operations Management, Vol. 13(2011) No.4, p. 420–438.

[3] S.M.Fortsch, E.A. Khapalova: Reducing uncertainty in demand for blood. Operations Research for Health Care, Vol.9 (2016), p.16–28.

[4] B. Fanoodi, B. Malmir, F.F. Jahantigh: Reducing demand uncertainty in platelet supply chain through artificial neural networks and arima models. Computers in Biology and Medicine, Vol.113 (2019), p.1–10.

ISSN: 2688-9323

[5] Z. Zhao, W. Chen, X. Wu: (2017) Lstm network: a deep learning approach for short-term traffic forecast. IET Intelligent Transport Systems, Vol. 11 (2017) No.2, p. 68–75.

[6] M.Kraus, S.Feuerriegel: Decision support from financial disclosures with deep neural networks and transfer learning. Decision Support Systems, Vol. 104 (2017), p.38–48.

[7] C. Xu, J. Ji, P. Liu: The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. Transportation research part C: emerging technologies, Vol. 95(2018), p.47–60.