





A Hybrid Forecasting Framework with Neural Network and Time-Series Method for Intermittent Demand in Semiconductor Supply Chain

Wenhan Fu ^(✉) , Chen-Fu Chien , and Zih-Hao Lin

Department of Industrial Engineering and Engineering Management,
National Tsing Hua University, Hsinchu, Taiwan
s103034467@m103.nthu.edu.tw

Abstract. As the primary prerequisite of capacity planning, inventory control and order management, demand forecast is a critical issue in semiconductor supply chain. A great quantity of stock keeping units (SKUs) with intermittent demand patterns and distinctive lead-times need specific prediction respectively. It is difficult for companies in semiconductor supply chain to manage intricate inventory systems with the changeable nature of intermittent (lumpy) demand. This study aims to propose an integrated forecasting approach with recurrent neural network and parametric method for intermittent demand problems to support flexible decisions in inventory management, as a critical role in intelligent supply chain. An empirical study was conducted with product time series in a semiconductor company in Taiwan to validate the practicality of proposed model. The results suggest that the proposed hybrid model can improve forecast accuracy in demand management of semiconductor supply chain.

Keywords: Demand forecasting · Intermittent demand · Combining forecasts
Neural network · Semiconductor supply chain

1 Introduction

Semiconductor industry is a capital-intensive industry that demand fulfilment and capacity utilization significantly affect the revenue and profit of semiconductor companies [1, 2]. With the development of the technology, the semiconductor industry has become highly vertically integrated. The segments of “Vertical Disintegration” in semiconductor supply chain consist of IC design, wafer fabrication, and packaging and testing. In semiconductor supply chain management, one of the biggest challenges is to forecast the demand pattern of a great number of products. For various industries such as electronics, automotive and aircraft, stock keeping units with intermittent demand accounts about 60% of the total stock value [3]. There are several categories of semiconductor products based on physical design such as ICs, op-amp, capacitor, transistors, resistor, diodes etc. Numerous sub-categories exist in each specific category with great number of products per sub-category.

Semiconductor companies need to make specific demand forecasting product by product as critical input of inventory control and ordering strategy. Each customer from downstream asks for types of component products at the same time. There are some gaps in semiconductor product demand forecasting: insufficient downstream information, low trade-off in market analysis and intermittent demand occurrence.

This research aims to propose a hybrid demand forecasting approach based on historical sales information to capture the intermittent pattern of semiconductor products demands as a part of to support flexible decisions in intelligent supply chain. In particular, an empirical case was employed to validate the model's effectiveness and compared with other classic methods. The proposed decision framework and process will assist decision makers in demand forecasting faced with historical data deficiency and demand uncertainty.

The remainder of this study is organized as follows. Section 1 introduces the background, significance and motivation of this research. Section 2 reviews demand categorization and intermittent demand forecasting. Section 3 proposes the research framework to make classification and forecasting of semiconductor products. Section 4 presents an empirical study in a leading semiconductor distributor in Taiwan to validate the proposed approach. Section 5 concludes with discussions and points out future research directions.

2 Literature Review

2.1 Demand Pattern Categorization

In semiconductor components management, thousands of different components have highly diverse usage individually [4]. Various demand patterns of the spare part causes a great segment in the inventory cost semiconductor components distributor [5]. Whereas, it's not practical to do a deep research to fit a demand pattern for each component, for there are too many components and that would cause a highly computational cost which is not efficiency. Thus, some categorization methods are built up to group the products, and they are invented to identify the appropriate forecasting method for the classification result.

Eaves and Kingsman [6] proposed a demand classification method, considering transaction variability, demand size variability and lead-time variability. However, there is not quantified method to classify the demand pattern. Syntetos et al. [7] used some mathematical proof to quantify the classification matrix. They proposed two cutoff value for calculate different demand pattern. One is the inter-demand interval ($ADI = 1.32$) It's used to identify the demand intervals. The other one is coefficient of demand variation ($CV2 = 0.49$). It can be used to identify whether the demand various is highly. The demand pattern can be classified into four types using this two-cut value. Lolli et al. [8] proposed a multi-criteria framework based on analytical hierarchy process (AHP) to classify and control inventory with intermittent demand.

2.2 Intermittent Demand Forecasting

Intermittent demand refers to demand pattern that appears randomly with high percentage of zero values between non-zero demand occurrences. The fluctuant nature of intermittent demand time-series render typical forecasting methods challenging to operate. Croston [9] presented an exponential smoothing based intermittent demand forecasting method by estimating separate extrapolation of non-zero demand size and demand interval between successive occurrences. Syntetos and Boylan [10] claimed that the original Croston's method was biased and made a modification with approximately unbiased demand/period estimates. Teunter et al. [11] developed a new decomposition method that updates the non-zero demand size and the demand incident probability separately.

Some artificial intelligence approaches were applied in intermittent forecasting problem. Gutierrez et al. [12] proposed a neural network modelling in lumpy demand forecasting and compares the performance of neural networks to traditional methods. Kourentzes [13] proposed dynamic demand rate forecasts with neural networks to capture potential interactions between the non-zero demand occurrences and zero demand intervals. Lolli et al. [14] applied single-hidden layer neural network trained back-propagation and extreme learning machines for forecasting intermittent demand and test it on industrial time series.

Though artificial intelligence method performs well in demand forecasting, neural network modelling and forecasting needs large amount of training data with high computing cost, which is rare for industrial company with off-line operation. Thus, how to balance the trade-off between traditional forecasting method with small sample data and neural network approaches worked with big data is a crucial issue. With effective evaluation and categorization of demand patterns, it will be supposed to improve forecasting accuracy with relatively low computing cost for real-time decision-making.

3 Methodology

To handle the deficient downstream information and intermittent demand occurrence for semiconductor products, we designed a hybrid demand forecasting framework for products/items in semiconductor supply chain which is shown in Fig. 1. First, we define the problem that semiconductor products with different demand patterns need to be forecasted in the mechanism. As two representatives of artificial intelligent and time series model for intermittent demand forecasting, recurrent neural network and SBA are applied in forecasting method construction stage. Then we use RFQV (Recency, Frequency, Quantity and Variance) analysis based on RFM (Recency, Frequency and Monetary Value) model [15] as the feature to classify products by demand feature. Each product is categorized into one subset and it is changeable with time moving. Finally, three accuracy matrices are performed for result evaluation and interpretation.

3.1 Problem Definition

Semiconductor companies usually maintain demand planning and inventory control weekly for thousands of products (with lead-time from 6–12 weeks) to keep customer service levels and combine predictions into orders for the component vendor.

The present problem is to establish an effective demand forecasting mechanism for all products in the light of their different characteristics since none of existing model can handle the diversity features of product demands.

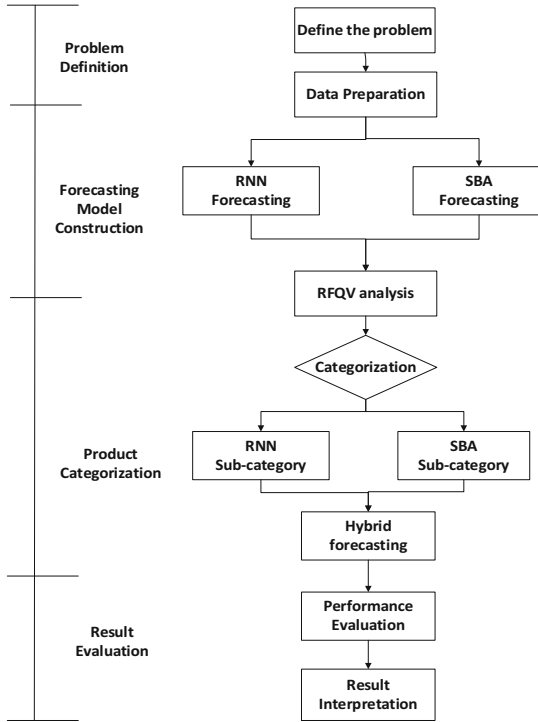


Fig. 1. Hybrid intermittent demand forecasting framework

3.2 Forecasting Model Construction

Two original forecasting models were occupied used at forecasting model construction stage in our framework. First is Recurrent Neural Network (RNN) for time series demand forecasting. The other is Syntetos-Boylan Approximation (SBA) [10], a typical representative of parametric time series method.

Suppose the demand series of product is $X = (x_1, x_2, \dots, x_n)$, the intermittent demand series of product (X) is split into two variables: non-zero demand $C = (c_1, c_2, \dots, c_n)$, and demand interval $I = (i_1, i_2, \dots, i_n)$, where n means sales time points. The forecasting model construction process are illustrated as follows:

Step 1: select specific product number $C = (c_1, c_2, \dots, c_n)$, $I = (i_1, i_2, \dots, i_n)$.

Step 2: set sample length: number of times that sample point is taken, e.g. sample length = 4, training sample $z_1 = (z_1, z_2, z_3, z_4)$.

Step 3: set lead time for predicted time interval (e.g. lead = 3, $c_7 = (i_2, i_3, i_4, i_5)$ to forecast i_8).

Step 4: Move the interval length and demand before and after the moving average ($z_1 = \frac{c_1 + c_2 + c_3}{3}$).

Step 5: Divide the interval length and demand by the maximum period (scaled to 0–1)

Step 6: The pre-processed data is divided into C and I. Each model is trained for adaptive parameter settings constants.

Step 7: Combining predicted demand into single values based on prediction intervals.

In particular, for adaptive parameter settings constants of SBA and RNN model. For SBA model, the mathematical formulation to update prediction are:

$$\begin{cases} c_{t+1} = \alpha x_t + (1 - \alpha) + c_t \\ i_{t+1} = \alpha q + (1 - \alpha) + c_t \text{ if } x_t \neq 0 \\ y_t = \frac{c_{t+1}}{i_{t+1}} \left(1 - \frac{\alpha}{2}\right) \end{cases} \quad (1)$$

$$\begin{cases} c_{t+1} = c_t \\ i_{t+1} = i_t \text{ if } x_t = 0 \\ y_{t+1} = y_t \end{cases} \quad (2)$$

where q means the non-zero interval still time t , α is a smoothing parameter which can be adjusted between 0 and 1, y_t is the predicted demand in time t .

For RNN model, we added the structure of the neural network to three layers and used LSTM to replace the general recurrent neurons as shown in Fig. 2. The LSTM has a special structure in the recursive layer. The memory area consists of memory cells and gates. The purpose of the gate is to control the transmission of messages.

LSTM computes the input $X = x_1, x_2, \dots, x_n$ according to the following formula to the output $Y = (y_1, y_2, \dots, y_n)$ with $t = 1, 2, \dots, n$:

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{cm}m_{t-1} + b_c) \quad (5)$$

$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_t + b_o) \quad (6)$$

$$m_t = o_t \odot h(c_t) \quad (7)$$

$$y_t = \emptyset(W_{ym}m_t + b_y) \quad (8)$$

where W represents a weight matrix (e.g., W_{ix} is a matrix from the input gate to the input), b is a vector of errors (e.g. b_i is a vector of errors in input layers), σ is a logic

and transfer function, i , f , o , c , m represent the input layer vector, the forgetting layer vector, the output layer vector, the cell activation vector, and the output activation vector respectively. The above vectors all have the same dimension, while \odot denotes the element-to-element multiplication. g is a hyperbolic tangent function as the activation functions for cell input and cell output. Finally, \emptyset is the network output activation function.

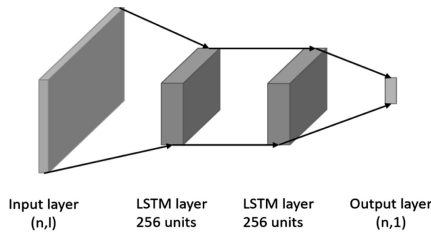


Fig. 2. Recurrent neural network structure

3.3 Product Categorization

It's crucial to select appropriate demand features for demand categorization of semiconductor products from historical sales recording dataset. Semiconductor products, especially component products like ICs, were made as part of electronic end-products like smart phone. The physical characters of such products have no influence in its demand sales.

For product value analysis, recency and frequency are favorable for inventory manager while quantity is more meaning than monetary for low unit price in semiconductor industry. Considering inter-demand interval and coefficient of demand variation are widely used in spare parts classification [7], we adjust RFM to RFQV (recency, frequency, quantity and coefficient of variation) for analyzing and evaluating product demand potentiality in the future. We separated the products by demand pattern through decision tree classification with the RFQV features. The response objective of demand pattern classification is the individual accuracy of SBA and RNN respectively. To customize the forecasting methods to fit the two categories, SBA and RNN with their rules in RFQV features. Thus, different kind products embedded with their corresponded forecasting method and combined as hybrid forecasting framework with two or more models.

3.4 Result Evaluation

For evaluation and interpretation of demand forecasting model, three different accuracy metrics: mean square Error, absolute error and mean absolute scaled error were used for validation. Root mean square error (RMSE), measures the average of the squares of the errors or deviations, is the common criterion for the demand forecast accuracy. Since RMSE enlarge extreme point's error in calculation, we use mean absolute error (MAE) that relatively robust to extreme value to balance the performance. In addition,

mean absolute scaled error (MASE) [16] is a metric never undefined or infinite for non-trivial cases and is proper for evaluate intermittent demand series, which is common seen in semiconductor products.

4 Case Study

To demonstrate the effectiveness of the proposed model, an empirical study was conducted in a semiconductor company in Taiwan. The experiment data included weekly demand time series 78 products recording for 2 years. In data preparing stage, we separated dataset into three parts: training data for 1-year period, validation data for half-year and testing data for the last half-year.

Since the proposed model is a hybrid of RNN and SBA, we compared the forecasting results with these two approaches independently while the model used in empirical company- moving average and other two common-used methods (Croston and TSB) were added for performance evaluation. We devise data-driven adaptive parameter settings for constants in those models during the validation process to avoid overfitting.

The performance comparison includes three forecasting accuracy metrics on average for different methods on the testing period dataset are presented in Table 1. The results show that the proposed hybrid method performs better in all three accuracy criteria while the computing time cost is much less than RNN method. Thus, classification mechanism of product demand pattern can improve current prediction accuracy with limit computing cost.

Table 1. Performance comparison of different methods

Method	MASE	RMSE	MAE
Moving average	1.334	1374.7	814.6
Croston	1.114	1248.9	644.5
TSB	1.170	1241.0	648.7
SBA	1.077	1157.6	606.8
RNN	1.058	1161.9	600.6
Proposed	0.845	1035.8	588.1

5 Conclusions

In this study, a hybrid forecasting approach with recurrent neural network and time series method is proposed for irregular demand problems to support flexible decisions in inventory management, as a critical role in intelligent supply chain. The results in empirical study suggest that the proposed hybrid forecasting with product categorization can improve forecast accuracy in semiconductor product demand management. This study is also a typical case showing the combination decision solution embedded with both traditional model and artificial intelligence method, which can even better than AI only approach in some field. Future research can be done to enhance the

forecasting approach in different scenarios (such as sufficient/deficient historical data) or import new ensemble operator instead of classification in combining forecast stage.

Acknowledgements. This research is supported by Ministry of Science and Technology, Taiwan (MOST106-2622-8-007-002-TM1; MOST 105-2218-E-007-027).

References

1. Chien, C.-F., Chen, C.-H.: A novel timetabling algorithm for a furnace process for semiconductor fabrication with constrained waiting and frequency-based setups. *OR Spectr.* **29**(3), 391–419 (2007)
2. Chien, C.-F., Dou, R., Fu, W.: Strategic capacity planning for smart production: decision modeling under demand uncertainty. *Appl. Soft Comput.* **68**, 900–909 (2018)
3. Bacchetti, A., Saccani, N.: Spare parts classification and demand forecasting for stock control: investigating the gap between research and practice. *Omega* **40**(6), 722–737 (2012)
4. Mönch, L., Uzsoy, R., Fowler, J.W.: A survey of semiconductor supply chain models part III: master planning, production planning, and demand fulfilment. *Int. J. Prod. Res.* (2017, in press). <https://doi.org/10.1080/00207543.2017.1401234>
5. Chuang, H.H.-C., Chang, H.L., Hsu, P.-C.: Improving the performance of inventory control-taking W company as an example. In: *The 17th International Conference on Electronic Business*, Dubai, UAE, pp. 289–292 (2017)
6. Eaves, A.H., Kingsman, B.G.: Forecasting for the ordering and stock-holding of spare parts. *J. Oper. Res. Soc.* **55**(4), 431–437 (2004)
7. Syntetos, A.A., Boylan, J.E., Croston, J.D.: On the categorization of demand patterns. *J. Oper. Res. Soc.* **56**(5), 495–503 (2004)
8. Lolli, F., Ishizaka, A., Gamberini, R., Rimini, B.: A multicriteria framework for inventory classification and control with application to intermittent demand. *J. Multi-Criteria Decis. Anal.* **24**(5–6), 275–285 (2017)
9. Croston, J.D.: Forecasting and stock control for intermittent demands. *Oper. Res. Q.* **23**(3), 289–303 (1972)
10. Syntetos, A.A., Boylan, J.E.: The accuracy of intermittent demand estimates. *Int. J. Forecast.* **21**(2), 303–314 (2005)
11. Teunter, R.H., Syntetos, A.A., Zied Babai, M.: Intermittent demand: linking forecasting to inventory obsolescence. *Eur. J. Oper. Res.* **214**(3), 606–615 (2011)
12. Gutierrez, R.S., Solis, A.O., Mukhopadhyay, S.: Lumpy demand forecasting using neural networks. *Int. J. Prod. Econ.* **111**(2), 409–420 (2008)
13. Kourntzes, N.: Intermittent demand forecasts with neural networks. *Int. J. Prod. Econ.* **143**(1), 198–206 (2013)
14. Lolli, F., Gamberini, R., Regattieri, A., Balugani, E., Gatos, T., Gucci, S.: Single-hidden layer neural networks for forecasting intermittent demand. *Int. J. Prod. Econ.* **183**, 116–128 (2017)
15. Fader, P.S., Hardie, B.G., Lee, K.L.: RFM and CLV: Using iso-value curves for customer base analysis. *J. Market. Res.* **42**(4), 415–430 (2005)
16. Hyndman, R.J.: Another look at forecast-accuracy metrics for intermittent demand. *Foresight: Int. J. Appl. Forecast.* **4**(4), 43–46 (2006)