2019 the 4th IEEE International Conference on Big Data Analytics

A Selection of Advanced Technologies for Demand Forecasting in the Retail Industry

Jiaxing Wang
Department of Computer Science and Software Engineering
Xi'an Jiaotong-Liverpool University Suzhou P.R.China
e-mail: Jiaxing.Wang18@student.xjtlu.edu.c n

G. Q. Liu
International Business School Suzhou
Xi'an Jiaotong-Liverpool University Suzhou P.R.China
e-mail: Guoquan.Liu@xjtlu.edu.cn

Lu Liu
Department of Computer Science and Software Engineering
Xi'an Jiaotong-Liverpool University Suzhou P.R.China
e-mail: Lu.Liu18@student.xjtlu.edu.cn

Abstract—**Retail companies always attempt to find a forecasting method to balance their purchasing and sales, whereas the performances of various prediction techniques are still not reliable. Furthermore, it is a question that how to select the proper forecasting model for some specific type of products. In this research, the classical forecasting models and the latest developing forecasting technologies are compared together based on the perishable products and non-perishable items respectively. The process is designed to compare the performance of typical statistic methods with several machine learning methods based on the thousands of historical transaction record of a large grocery retailer. The criterion is also explored in this study include predictive performance, generalization ability, runtime, cost and convenience to evaluate the comprehensive performance of these models, thus companies can easily choose their most accepted model.**

Keywords- Demand Forecasting; Deep Learning; ARIMA; Recurrent neural networks (RNN); Retail Industry;

I. INTRODUCTION

In the retail industry, in order to stay competitive in a highly dynamic market, companies have the motivation to apply advanced technologies to forecast the demand for customers. And the accuracy of the prediction is extremely important. Retail companies attempt to maintain a minimum inventory to satisfy the demand of customers, thereby reduces distribution expenses and stocking costs and improve their profits [4]. However, it is common for product demand trends to change in a very short time, and the precision of forecasting will be very low [10]. If prediction over the demand, grocers are stuck with overstocked, perishable goods. On the contrary, popular items quickly sell out, leaving money on the table and customers fuming. Both situations lead to lose profit and competitive power. Many types of research explored deep learning methods try to get a higher accuracy. Furthermore, both the academic world and the industry put forward new technologies to forecast the demand. Nevertheless, it is very hard for retail companies to choose the most accepted model, because there does not exist a universal method to deal with such a question, and the accuracy is still not stable, especially in the complex case, as retailers may have various locations with unique needs, have seasonal products, and unpredictable good or bad news. Under the circumstances, it is necessary to put forward a selection criterion to help companies to find the most suitable model to implement. This research will distinguish the performance of typical statistic methods with several machine learning methods and establish a proper evaluation criterion.

II. LITERATURE REVIEW

Retail companies have traditionally adopted forecasting methodologies to instruct their production and operation. There are various technologies to match the demand for customers in the past. The traditional time series forecasting based on statistics has been an active research area for decades, linear statistical approach such as Auto-Regressive (AR)model, Moving Average (MA)model, Auto-Regressive Moving Average (ARMA) model, and autoregressive integrated moving average (ARIMA) model are applied to a multistage supply chain model [6]. Babu and Reddy compared the accuracy of Basic ARIMA, Trend based ARIMA and Wavelet-based ARIMA in temperature prediction area. They concluded Trend based ARIMA have a better performance through measuring Mean Absolute Percentage Error, Maximum Absolute Percentage Error and Mean Absolute Error of their outcomes[1]. Soder combined seasonal autoregressive integrated moving average (SARIMA) and discrete Markov processes to predict the power market price. The research considered the factor of seasons because of the pattern of power demand is highly related to weather [9]. Slow employed fractionally integrated autoregressive-moving average (FARIMA) model to forecast bandwidth requirements [7]. Hu et al. proposed an Autoregressive Integrated Moving Average with Generalized Autoregressive Conditional Heteroscedasticity (ARIMA- GARCH) model for traffic flow prediction [3].

These time series models provided an effective method to observe, measure and reduce the bullwhip effect. However, researchers claimed nonlinear models are able to have a better performance than their counterparts of linear forecasting in the real world [5]. The typical machine learning algorithm Support Vector Machine (SVM) and Support Vector Regression (SVR) is employed to predict the demand [11]. With the seasonal and promotional factors are a consideration in their model, they declared a better performance of SVM than traditional statistical models in a small-scale dataset. Whereas, it needs an unaffordable computing power at high dimensional dataset. Recurrent neural networks (RNN) serve as a powerful tool for a large-scale chaotic time series dataset [2]. His study compared various methods forecasting accuracy of the foundries test set and concluded RNN and SVM performance best. Furthermore, Long Short-Term Memory (LSTM) algorithms as a special RNN was designed to store information for long periods of time to overcome the shortcoming of the vanishing gradient in the previous network. Marino et al. compared the standard LSTM and LSTM-based Sequence to Sequence (S2S) to forecast the Building Energy Load [8].

III. Description of the Experiment

This experiment used Keras and scikit-learn packages to build the traditional statistic models and machine learning models to process the data of a retail company's transaction records. And the study compared the models in five dimensions.

A. *Introduction and Hypotheses*

The hypothesis in this study is that different types of products need different forecasting models, especially the demand forecasting of perishable items and non-perishable items require totally different methods.

B. *Materials*

As shown in Figure1, I use the cross-industry standard process for data mining (CRISP-DM) to instruct the analyses of the dataset from a large Ecuadorian-based grocery retailer. The data contains the sale information of their hundreds of supermarkets, with over 200,0 different products on their shelves. Business understanding is the first step of the research, external factors like Oil price, Holidays and natural disasters should be considered in the model. And from the business knowledge, the perishable goods should have a higher weight because of their characteristic. This project used Tableau to do the Exploratory Data Analysis (EDA) to show the structure of tables and find some products is highly correlated with the day of week. Then the datasets need to remove duplicates, modify the errors, check the consistency. A prepared dataset is the base of the next step. The following algorithms are used to compare in this pilot study: ARIMA, SVM, RNN and LSTM etc.

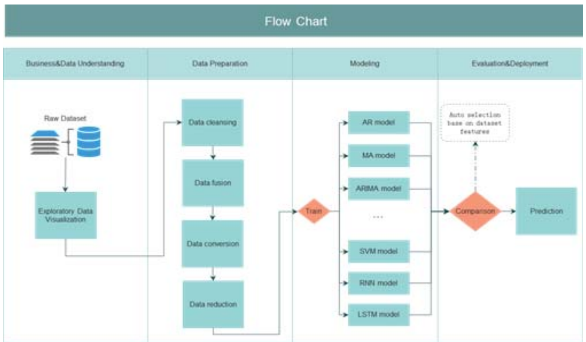

Figure 1. The Methodology of the whole study

C. *Methods*

ARIMA is a generalization of an ARMA model in time series analysis. The AR part of ARIMA indicates that the current value of the variable only dependent on the previous values of this variable. The MA part indicates that the current value of the variable only dependent on the external disturbance. The integrated part indicates that the current values have been influenced by the difference between the current value and previous value. These three parts make sure the model can fit the dataset. The ARIMA models are the following:

AR(p):

$$x_t = \delta + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + u_t \tag{1}$$

δ is constant term, ϕ is coefficient of Auto-Regressive term, u_t is residual, x_t is the value of x at time t. Let $L * x_t = x_{t-1}$, the characteristic equation of AR(p):

$$\Phi L = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p = 0 \tag{2}$$

MA(q):

$$x_t = \mu + \mu_1 + \theta_1 \mu_{t-1} + \theta_2 \mu_{t-2} + \dots + \theta_q \mu_{t-q} \tag{3}$$

The characteristic equation of MA(q):

$$x_t - \mu = (1 + \theta_1 L + \theta_2 L_2 + \dots + \theta_q L_q)\mu_t = \Theta(L)\mu_t = 0 \tag{4}$$

ARIMA(p,d,q): Let $w_t = \Delta^d x_t = (1 - L)^d x_t$ Then:

$$w_t = \varphi_1 w_{t-1} + \varphi_2 w_{t-2} + \dots + \varphi_p w_{t-p} + \delta + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q} \tag{5}$$

The characteristic equation of ARIMA(p,d,q):

$$\Theta(L)\Delta^d x_t = \delta + \Theta(L) u_t \tag{6}$$

SVR is a method using the same principle as the SVM for classification to deal with regression problems. This model seeks and optimize the generalization bounds of the given dataset. They calculate the loss function beyond the acceptable errors compared with the true value. The SVM training algorithm builds a model that predict the dependent variable through the trained data.

RNN is a class of artificial neural network using sequential information depended on the previous computation to train the weight between next layer. The connections between nodes have the ability to reflect the temporal dynamic behavior for a time sequence. The memory parts of RNNs allows they process sequences of inputs. This makes them applicable to predict the unknow items. As shown in the Figure 2.is the structure of a typical RNN:

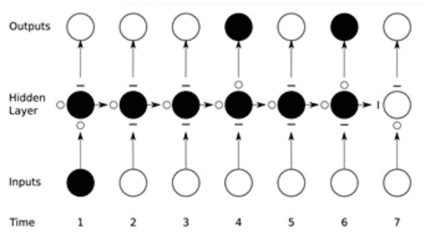

Figure 2. Structure of RNN

LSTM is a variant of RNN, which is special. Compared with other traditional RNN, LSTM has redesigned its structure of hidden layer. LSTM's hidden layer also called memory module, with the following structure: a cell, three gates: an input gate, an output gate and a forget gate separately. Every unit has four layers. All the components together control the read, write and reset of the cell. As shown in Figure 3. The red parts are different activation functions, the yellow parts represent vector's algebra, for instance, Cartesian product. The arrow shows the direction of propagation. Two solid lines' convergence means that the two parts of the vector are merged, and one solid line is divided into multiple pieces to indicate that the vector is copied and then flowed separately to different parts of the network.

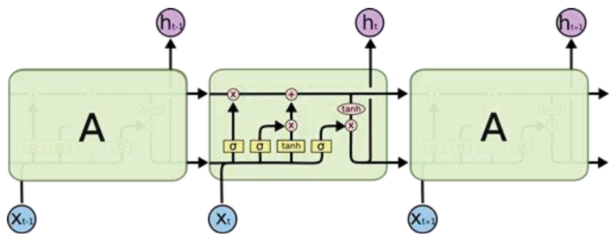

Figure 3. Structure of LSTM

IV. Results and Discussion

There are thousands of kinds of items in the dataset. This study divided them into perishable items and non-perishable items. Figure 4. shows the distribution of mean sales spread relatively evenly throughout the year except December because of the Christmas day.

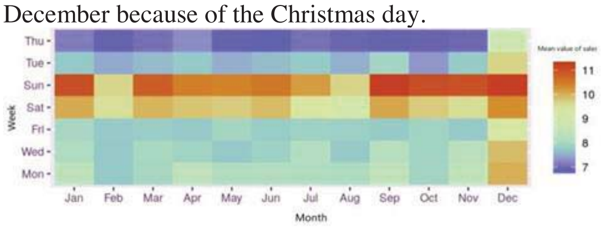

Figure 4. Exploratory analysis of Sales Data

Table 1. Shows the value of evaluation index in 5 dimensions. As shown in Figure 5, SVM, RNN and LSTM have a high predictive accuracy to deal with perishable products, whereas ARIMA is outstanding in the runtime and easy to use aspects. Non-linear machine learning technology is better than traditional forecasting method in many dimensionalities. And more complex framework of machine learning needs more computer power and maybe not have a better performance. At the same time, the characters of models experimented by non-perishable products are similar with perishable products. However, As shown in Figure 6, there still indicate some differences, the runtime of ARIMA is not as shorter as it shows in previous test. SVM is more suitable for perishable products because of the highest accuracy. LSTM is the most acceptable method for non-perishable items on account of its lower cost and excellent prediction performance.

TABLE I. The Value of Evaluation Index

Perishable(normalization)	ARIMA	SVM	RNN	LSTM
Runtime	0.2854	0.5730	0.7153	0.5007
Generalization Ability	0.3333	0.5167	0.5667	0.6667
Convenience	0.3636	0.4545	0.5455	0.6364
Predictive Accurancy	87.13%	99.10%	98.90%	96.80%
Cost	0.1818	0.7273	0.6364	0.5455
non-Perishable(normalization)	ARIMA	SVM	RNN	LSTM
Runtime	0.4433	0.4241	0.5951	0.4049
Generalization Ability	0.3333	0.5167	0.5667	0.6667
Convenience	0.3636	0.4545	0.5455	0.6364
Predictive Accurancy	86.50%	98.60%	95.90%	98.90%
Cost	0.1818	0.7273	0.6364	0.5455

(a) Perishable items. (b) non-Perishable items
Figure 5. Performance of models.

The limitations of this study are mainly from the rough classification of products, the items are divided into two part which may lose a lot of information from the characteristics of specific goods. And this project not contain all of the forecasting technologies, so it can be improved by compared more models in a more elaborate way. Furthermore, because of the dataset used to train the model are all from a single enterprise. Although the model gets a high performance, the network may still lack generalization ability. The model needs more data to check validation and accessibility. At the same time, there may exist overfitting problem which needs to be modified. The evaluation criterion may not applicable all categories such as fashion goods and luxury goods. In addition, the model ignores big events and the emotion of human beings, which may lead to a drastically change here. The study also can be improved by considering some economic factors such as GDP, Inflation rate, Unemployment etc.

V. Conclusion

This study explored the performance of ARIMA, SVM, RNN and LSTM in five dimensions include predictive performance, generalization ability, runtime, cost and convenience. And the experiment shows specific goods in retail companies required special technology to forecast the demand. Furthermore, it is deserved to explore the matching rule between models and products. Generally, SVM and LSTM have an excellent performance in the prediction tasks.

ACKNOWLEDGMENT

This research is supported by KSF-A-13. The authors extend their deep sense of gratitude to our families and peers who helped improve the literal and technical content of this paper.

REFERENCES

[1] Babu, C. N., and Reddy, B. E. Predictive data mining on average global temperature using variants of arima models. In International Conference on Advances in Engineering, Science and Management (2012), 256–260.

[2] Carbonneau, R., Laframboise, K., and Vahidov, R. J. E. J. o. O. R. Application of machine learning techniques for supply chain demand forecasting. 1140–1154.

[3] Chen, C., Hu, J., Qiang, M., and Yi, Z. Short-time traffic flow prediction with arima-garch model. In *Intelligent Vehicles Symposium* (2011), 607–612.

[4] Christopher, M. *Logistics supply chain management*. Pearson UK, 2016.

[5] Chu, C.-W., and Zhang, G. P. J. I. J. o. p. e. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. 217–231.

[6] Gilbert, K. J. M. S. An arima supply chain model. 305–310.

[7] Ilow, J. Forecasting network traffic using farima models with heavy tailed innovations. In *Acoustics, Speech, Signal Processing, on IEEE International Conference* (2000).

[8] Marino, D. L., Amarasinghe, K., and Manic, M. Building energy load forecasting using deep neural networks. In *Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE*, 7046–7051,

[9] Olsson, M., and Soder, L. Modeling real-time balancing power market prices using combined sarima and markov processes. *IEEE Transactions on Power Systems 23*, 2 (2008), 443–450.

[10] Singh, N., Olasky, S. J., Cluff, K. S., Jr, W. F. W., and Manugistics, I. Supply chain demand forecasting and planning.

[11] Yue, L., Yin, Y., Gao, J., and Tan, C. Demand forecasting by using support vector machine. 272–276.