

Application of Big Data Analysis in Sales Forecasting

Qianqian Ma^{1, a, *}, Ailing Xie^{2, b}

¹College of Mechanical and Vehicle Engineering, Hunan University, Changsha, China

²Fisher College of Business, The Ohio State University

a. mqq202104060601@hnu.edu.cn, b. xie.1426@buckeyemail.osu.edu

*corresponding author

Abstract: As a matter of fact, with the rapid development of the computation ability and data delivering, the big data analysis becomes a common tool to realize the predictions. With this in mind, this paper focuses on the impact of big data analysis on sales forecasting, starting with an introduction to the present-day context and significance of sales forecasting and some of the forecasting scenarios implemented by previous generations based on machine learning. After introducing these backgrounds, we start to introduce big data analysis and its impact on sales forecasting. According to the analysis and the summary of the previous experience, we introduce the commonly used models, the parameters of the models, and the evaluation of the models as well as the application scenarios, and the specific application results. Finally, we demonstrate the limitations as well as future perspectives of big data analysis and sales forecasting. Overall, these results shed light on guiding further exploration of big data analysis in sales prediction.

Keywords: big data analysis, sales prediction, sales forecasting model

1. Introduction

With the growth of the internet, many traditional ways of selling have been replaced by mobile phone software. More customers and merchants are favoring the use of software on electronic products to conduct transactions. At the same time, due to the advancement of technology and the personalized needs of customers, modern industries are faced with various challenges to customer demands such as rapid response and shortening of product delivery cycle [1]. Since then, businesses have looked for a way to more accurately predict customer demand, even if they use a method of calculating the future sales focus and sales volume and based on this calculation to rationally allocate production resources, so as to achieve the highest efficiency and low-cost of a production and sales method to achieve a competitive advantage. At the same time, many enterprises will develop some suitable systems and databases for them to store a large amount of data.

We now formally introduce the concept of sales forecasting. Sales forecasting is a traditional sequential forecasting problem that relies on historical sales records to predict sales in the future. Sales forecasting technology allows sales forecasting to be automated without relying on human expertise. In general, sales forecasting corresponds to businesses that may be manufacturing (manufacturing), retail (retailing), marketing (marketing), or wholesale (wholesaling) [2]. Sales forecasting has emerged in the e-commerce retail and brick-and-mortar retail industry, and its use is very important in the current chain of drugstore sales forecasting research involves a number of fields,

including data mining, machine learning and artificial intelligence. Machine learning technology is also widely used in chain drugstore sales forecasting, but the field is facing the problem of data quality and the influence of multiple factors. Forecasting is very helpful for businesses to analyze demand, make rational economic decisions and inventory management, accurate business forecasting can allow enterprises to complete the inventory adjustment and production resource allocation earlier to reduce the loss of goods production and become more price advantage. The significance of this research is as follows. Firstly, the era of data fiction is developing rapidly, and artificial intelligence technology is emerging AI technologies such as face recognition, personalized recommendation and AlphaGo are widely applied and are profoundly changing the operation mode of the traditional industrial chain [3].

Nowadays, many traditional industries have begun to rapidly integrate with artificial intelligence technology, using machine learning and other technologies to assist high-level decision-making. Not only that, but many modern sales industries also pursue data-based sales, that is, the use of data to predict and solve problems, thereby reducing the cost of inventory planning. It is necessary to make a prediction through the analysis of scattered data. The study provides an algorithmic basis for the subsequent development of a sales forecasting system. The research is of great significance in that it allows companies to gain greater benefits from accurate, efficient, and stable sales forecasting results, and thus has a greater advantage.

Machine learning has evolved to achieve very impressive results in many areas, some prediction scenarios are cited where machine learning has become more mature. The first is face recognition, which is done by recognizing certain identifiable features on a person's face and tracking them. Face target detection is the basis of face recognition technology. Its task is to find out the location of all faces in the image for a given image. A number of faces in an image are framed with a number of rectangles, and the coordinates of these face rectangle locations are obtained [4]. Face recognition technology through machine learning has been developed for a long time and is relatively mature, and has been used in various aspects of our lives, such as access control systems, cell phone unlocking and so on. Then there's the smart car aspect, which will not only integrate the Internet of Things, but will also understand the owner and its surroundings. It will automatically adjust its internal settings to the driver's needs, such as temperature, stereo, seat position, and so on. It will also report malfunctions and even fix them on its own, it will drive itself, and it will provide real-time advice on traffic and road conditions. Autonomous driving requires a lot of knowledge in computer science and technology, such as real-time computing and analyzing big data such as road conditions and the state of the car itself. A system model based on the random forest algorithm in machine learning for intelligent decision-making judgment of automobiles is proposed, and Matlab/Tensorflow is used as the development platform for the development of intelligent automobile decision-making control system software [5]. For example, speech recognition, a broadly defined natural language processing technology, is used for smoother communication between people and people, and between people and machines. Speech recognition has been used in all aspects of life such as, Apple's siri, smart speaker assistant, Ali's T-mall Elf, KDDI a series of intelligent voice products and so on. In addition to the above mentioned prediction scenarios already available in machine learning, there are also areas such as fraud detection, product and service recommendation, medical analysis, customer analysis, online search, etc. The prediction scenarios of machine learning are closely related to our lives.

The aim of this research is to explore the current state of big data analysis in sales forecasting and to uncover the limitations of big data analysis in sales forecasting that can help in the advancement of this field as well as to promote the development of this field.

2. Basic Descriptions of Big Data Analysis

Big data analysis, simply put, is the process of organizing and analyzing large amounts of data so that it can be applied where it is needed. Big data analysis requires five main processes, which are collecting, processing, storing, analyzing and visualizing data. Through these five processes we can gain a deeper understanding of the information and trends in the data and come up with the results we need to help us make decisions. Big data analysis can be invaluable to a wide range of industries today, in areas such as e-commerce, insurance, marketing, healthcare, real estate, and more. For example, in the e-commerce space big data analysis can provide value to the digital commerce space in a number of ways.

With big data analysis, suppliers can gather extensive data about competitors who are "out of stock" of certain items, and then ensure that shipments of those items are dramatically increased in order to appeal to impatient and eager shoppers. Analyzing the current state of the market and going trends can be utilized for sales forecasting. By collecting sales data including sales, sales volume, sales channels, sales rankings, etc., we can find out which products are more in line with consumer expectations and which products need to be improved. After collecting the data, we use technology tools to mine and analyze the data, and in this process, we generally build models, including sales forecasting models, competitor analysis models and so on.

Through data analysis and modeling, companies can have a better understanding of market trends and customer needs, which allows them to better determine the target market, product positioning, sales channels, promotional activities, etc., so as to efficiently utilize human resources as well as material resources. Data analysis can further prevent errors and poor decision making in sales planning. Let's use an example to illustrate the role of big data analysis in sales forecasting. Before the use of big data analysis, according to previous statistics, the recovery rate of the questionnaire is about 5%. Hence, one wants 300 effective questionnaires, one needs to send out 6,000 questionnaires, which is extremely time-consuming and labor-intensive.

However, with big data analysis, the above can be accomplished very quickly within a few hours of big data analysis. First of all, we accurately selected 2% of VIP customers and sent out the corresponding number of questionnaires, more than 30% of the questionnaires were recovered within a few hours of sending out the questionnaires, and within a few days, we recovered about 90% of the target number of questionnaires. This is because we customize the questionnaires to our customers on a one-to-one basis through big data analysis, and we send the questionnaires to each customer at the time when he is most likely to use his email, which greatly increases the probability of filling out the questionnaires, and therefore increases the efficiency of the recovery.

3. Influencing Variables

Sales forecasting is critical for many companies today, so to explore the impact and growth of big data analysis on sales forecasting, it's important to first understand what all the factors that influence sales forecasting are. There are generally external and internal factors that affect sales forecasting, so let's start with the external factors. The first is that changes in market demand for a product can affect sales forecasting. Changes in market demand is the most important external factor that affects sales forecasting. Market demand can be influenced by many factors, such as fashion trends, lifestyle changes, the occurrence of big events, demographic changes, and so on. For example, mobile accounts for 30% of online sales in the U.S. and has an average consumer satisfaction rating of 83%. As this trend is slowly changing the traditional way of shopping, e-commerce companies should do their best to develop the mobile side of their online platforms to meet the new shopping needs of consumers [6]. Secondly, the external environment, and the general economic environment, such as the financial crisis, oil resources, economic policy, etc. For the general environment of the

change will have a certain degree of impact on the enterprise revenue-generating products. For example, since the second half of 2016, the government's environmental assessment control has been further tightened, and many manufacturers with substandard environmental assessment standards have been forced to stop work and rectify their operations and have been unable to ship their products on time, which led to a more serious loss of general-purpose products of M Infant & Child in the first half of 2017 due to out-of-stocking [7]. One more thing, industry competitors are also a major external factor that affects sales forecasting. Market share is fixed, competitors' sales will affect the development of their own business, so understand the industry competitors is also an important part of the sales forecast.

Speaking of internal factors, the first thing that will affect the sales forecast is the marketing activities policy. Changes in the company's internal strategy of products, prices, market positioning, advertising and promotion, and promotional activities will have an impact on the company's sales. If the marketing campaign is in line with the consumer's intention, it will have a very favorable effect on sales. On the other hand, there is production as well as service, and production must be guaranteed in the case of normal sales. The pre-sales and post-sales of sales will also have different degrees of impact on the sales results, good pre-sales and post-sales will also increase the sales volume, can significantly reduce customer turnover. In the actual working situation, internal factors control, usually in the enterprise will be based on external factors to do timely adjustment, in order to keep up with the market rhythm, so that the enterprise profit. Fig. 1 gives a mindmap for factors.



Figure 1: Influencing factors.

4. Models and Evaluations

Sales forecasting is a key component in business decision making. With the advancement of technology, there are multiple models available for sales forecasting from traditional statistical methods to modern machine learning and deep learning techniques. Here are some of the models that have been frequently used in studies since 2015 and beyond, their parameters and how they were evaluated.

As for Recurrent Neural Networks (RNN) and Long Short-Term Memory Networks (LSTM), they have parameters include number of neurons, learning rate, batch size, activation function, etc. The evaluation metrics include Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) [8]. For prophet, it has parameters including seasonality, holiday effects, cycle length, flexibility of trend changes, etc., which can also be evaluated by MSE, RMSE and MAE [9]. For XGBoost, it has parameters including learning rate, maximum depth of tree, number of trees, regularization factor, etc., which can also be evaluated by MSE, RMSE and MAE [10].

An RNN is a type of neural network suitable for processing and predicting sequential data. It is characterized by the presence of loops in the network, which allow information to "loop" through the

network. An RNN accepts an input and produces an output at each time step. However, unlike a normal neural network, an RNN also takes the hidden state of the previous time step as input to the current time step. This allows the RNN to maintain a "memory", where previous information can influence later outputs. Although RNNs are theoretically capable of handling sequences of any length, in practice they tend to suffer from the problem of gradient vanishing and gradient explosion. This makes it difficult for the network to learn and retain long-term dependencies.

In order to solve the problems of RNNs, LSTM was introduced. LSTM is a special type of RNN which consists of three important gate structures: input gates, forgetting gates and output gates. It has following components:

- Input gate: decides how much new information we will store in the cell state.
- Forget gate: determines how much information we will discard from the cell state.
- Output gate: decides what value to output based on the cell state.

Through the interaction of these gates, LSTM can learn long-term dependencies without being affected by the gradient vanishing or gradient explosion problems. LSTMs perform better than regular RNNs on many tasks that require learning long-term dependencies. They are especially becoming very popular in areas such as natural language processing and time series prediction. Parameters of LSTM networks include number of neurons, learning rate, batch size, sequence length, activation function, optimizer, etc. Sales data is usually time-series data where each point in time may be influenced by previous data. This sequential nature makes RNNs and LSTMs particularly well suited to handle such tasks. For example, if we want to predict a store's sales for the next month, LSTM can use data from previous months as input to account for seasonality, holiday effects, promotions, and other factors that may affect sales. In summary, RNNs and LSTMs provide powerful tools for sales prediction, especially when long-term dependencies on historical data need to be considered. The sketch of RNN and LSTM are shown in Fig. 2.

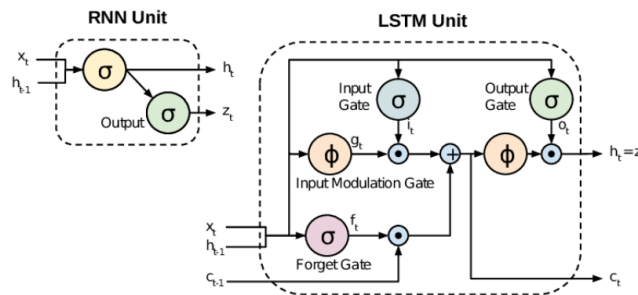


Figure 2: RNN and LSTM.

5. Application

In order to ensure adequate inventory on the shelves and avoid overstocking, retailers need to accurately forecast future sales. In addition, they can develop promotional strategies based on the predictions. By using deep learning models, some large retailers have improved the accuracy of their sales forecasts, resulting in reduced inventory costs and increased sales [11]. The retail industry is one of the largest applicants of sales forecasting. Forecasting is particularly important in this sector due to the diversity of products, customers and changing market trends. To ensure timely availability of products and avoid overstocking or stock-outs, retailers need to forecast future sales. An accurate sales forecast helps retailers decide when, how much, and where to buy. For example, large retailers such as Zara use advanced forecasting techniques to determine which styles should go into production and how many pieces should be produced.

Retailers often run sales promotions, such as discounts and buy-one-get-one-free. In order to determine which products should be discounted, when they should be discounted, and by how much, retailers need to forecast the likely impact of the promotions. Large retailers such as Walmart and Target use machine learning models to predict how specific promotions will affect sales of individual products and how to maximize profits. When introducing a new product, retailers need to predict how it will perform in the marketplace. This can help retailers decide whether to invest further in the product or how to market it. When Apple Stores introduces a new iPhone or other product, they use predictive modeling to estimate sales at initial launch to determine initial production and inventory strategy. By analyzing customer data, retailers can predict which customers are most likely to purchase a particular product and personalize marketing campaigns based on that information. Amazon uses its recommendation system to predict which products a customer is likely to be interested in and personalizes the shopping experience based on those predictions [12]. E-commerce platforms need to predict sales in order to optimize their advertising strategies, inventory management, and pricing strategies. Some e-commerce platforms using machine learning and time series analysis have achieved highly personalized product recommendations, resulting in improved conversion rates [13]. Pharmacies and healthcare organizations need to predict the demand for medications to ensure they always have enough inventory to meet patient needs. By using advanced predictive modeling, many pharmacies have been able to reduce drug shortages as well as waste from over-buying [14].

6. Limitations & Future Outlooks

Although big data analysis has been developed for a long time and has been fully applied to sales forecasting, big data analysis still has many limitations. First, there is the issue of information quality. In big data analysis, the quality of the analytical results relies heavily on the quality of the information collected. There may be a lack of precise criteria in the data collected, leading to the data collected having different meanings. Similarly, lack of clarity in the interpretation of the collected information resources may lead to multiple interpretations of the collected data very much affecting the accuracy of the data. Secondly, the collection of data cannot exclude duplication of collection, most of the existing data collection for data analysis cannot be detailed to the individual, so it leads to duplication of data collection. In addition to this there are many minor problems such as incomplete data recording, incorrect data recording, low timeliness and lack of precision. There is also the obvious limitation that big data analysis cannot do anything to completely protect personal privacy. When a company collects data, it often involves customers' personal privacy. If a company fails to protect customers' personal privacy, and it is utilized by criminal companies or people with ulterior motives, it will cause great harm to customers. Yahoo, for example, made its first public announcement about the data breach in December 2016 and said it happened in 2013. At the time, it was in the process of being acquired by Verizon, and it was estimated that more than 1 billion users' account information had been accessed by the hacker group. Less than a year later, Yahoo announced that the actual number of compromised user accounts was as high as 3 billion. Another example is the problem of data management due to data processing requires a large amount of data, however, today's enterprises for the management of data progress is not great, so the improper management of data is also an urgent problem to be solved, data management, including data storage, processing, transmission, and other aspects, but also need to consider the issue of data security and backup [15].

However, it is believed that soon, data will be deeply integrated with cloud computing, allowing big data marketing to play a greater role. At the same time, we believe that a new round of technological revolution is coming, data mining, machine learning, artificial intelligence will develop rapidly, when the big data analysis technology will also realize a qualitative leap. In the future, data science will become a specialized discipline. Colleges and universities will offer specialized data

science majors. The basic data platform will be complemented by a cross-domain data-sharing platform, which will also create a range of new jobs related to it. Eventually, data sharing will extend to the enterprise level and become the core of the industry of the future.

7. Conclusion

Sales forecasting is not only the core of strategic planning for an organization, but also a key tool for responding to market changes. With the development of big data and artificial intelligence, modern sales forecasting has shifted from traditional qualitative analysis to quantitative and data-driven methods. Using historical sales data, market trends and consumer behavior, companies can more accurately predict future sales and market share. These forecasts provide strategic guidance to help companies make informed decisions in supply chain management, inventory control, and marketing strategies. Through our research, we have found that sales forecasting in the retail industry has a critical impact on inventory management, promotional campaign strategies, new product launches, and customer analysis. Accurate sales forecasting can significantly improve supply chain efficiency, increase sales, reduce inventory costs, and improve customer satisfaction. Most successful retailers have integrated forecasting into their day-to-day decision making, achieving significant economic benefits. However, any forecasting model is subject to uncertainty and bias. As globalization and market dynamics increase, so does the complexity of forecasting. In the future, more research will be devoted to improving forecasting accuracy, fusing data from multiple sources, and accounting for unforeseen market changes. A deeper understanding of sales forecasting methods and strategies can help organizations better address market challenges and achieve sustainable growth in the long term.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

- [1] Xu, Y. (2022). *Vehicle Sales Forecasting Under a Combinatorial Model: An Empirical Study Based on Sales Data of a Car Brand*. Northwest University.
- [2] Wang, X. (2022). *Hierarchical Spatio-Temporal Neural Network for Sales Prediction*. University of Shandong
- [3] Yang, W. (2022). *Research on Sales Prediction of Chain Drug Stores Based on Ensemble Learning*. Yunnan University of Finance and Economics.
- [4] Li, L. (2022). *Research of Face Image Generation Method Based on Adversarial Machine Learning*. University of Nancang
- [5] Fu, Z., Zeng, Y., Gui, J., Wu, X. (2022). *Machine Learning Based Decision Making System for Intelligent Vehicles*. *Computer programming skills and maintenance*, 1, 126-128.
- [6] Hu, X. (2019). *A study of the company's sales forecast and sales strategy*. University of Ningbo.
- [7] Wang, H. (2020). *Research on sales forecasting of general products of m Baby Company*. Hebei University of Technology.
- [8] Chen, Q., and Wei, Y. (2016). *Deep learning for sales forecasting*. *arXiv preprint arXiv:1607.05636*.
- [9] Taylor, S. J., and Letham, B. (2018). *Forecasting at scale*. *The American Statistician*, 72(1), 37-45.
- [10] Chen, T., and Guestrin, C. (2016, August). *Xgboost: A scalable tree boosting system*. In *Proceedings of the 22nd acm sigkdd International Conference on knowledge discovery and data mining*, 785-794.
- [11] Wang, Y., Wang, S. and Ye, J. (2016). *Retail sales forecasting using deep learning*. *Journal of Retailing*, 92(3), 317-328.
- [12] Chen, K.Y., and Kuo, R.J. (2017). *Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry*. *Computers & Industrial Engineering*, 102, 488-499.
- [13] Zhang, D., and Wang, J. (2018). *Sales forecasting in e-commerce using convolutional neural network*. *International Journal of Database Theory and Application*, 11(1), 77-88.
- [14] Ma, S., Zheng, Y., and Jiang, H. (2017). *A deep learning approach for sales forecasting in retail drugstores*. *Computational Intelligence and Neuroscience*, 17, 13.
- [15] Wu, S., (2017). *Research on the Product Cultural Marketing Strategy for CF Group Co*. *Marketing*, 18, 22.