Forecasting in the context of retail

Martin BORDES

May 2023

Abstract

In many areas, forecasting has become one of the most important concepts in recent decades. This is especially true for the retail industry. This field consists of very large companies in many countries. It also involves special issues that make forecasting even more important, such as shelf life and promotions.

The goal of this theoretical work is to review the main considerations of a sales forecasting practitioner in the context of retail. We'll first look at some specificities to have in mind when dealing with this task. Then we'll review the main forecasting metrics for this problem. Afterwards, we'll focus on a major aspect of sales forecasting, uncertainty. Finally, we'll look at state-of-the-art methods thanks to an insightful challenge.

Contents

1 Introduction

In the retail industry, forecasting is used in a lot of different activities. On one hand, aggregated predictions, such as sales for all products, for all products in one category or for all stores in one region, can be used for marketing, accounting or even finance. On the other hand, more precise forecasts are an essential component of pricing and replenishment. In this work, I'll talk mainly about this second type of forecast. While this task was done with very simple methods or even intuition for a long time, the rise of machine learning has led to major improvements in the area of forecasting in the last decades and this has directly impacted the retail industry. A retail forecasting practitioner has to deal with a lot of different issues, some that are pretty general and others directly related to the retail context.

For any retailer, answering the problem of sales forecasting allows to anticipate the future in order to take the best decisions. In this way, it is possible to avoid the two main problems in this area: understocking and overstocking. The former can lead to empty shelves and therefore prevents the shop from meeting customer demand. This creates a potential loss of turnover as well as customer dissatisfaction. Overstocking causes even greater losses and waste if it leads to the non-sale of certain products. It is therefore essential to remedy this problem.

2 Specificities of the retail context

2.1 Explainability

When building a sales prediction model, the most important thing to keep in mind throughout the work is the future use that will be made of it. The set of people who will use the results of the model in the future is ususally called the business. It is essential to understand what their needs are in order to adapt the model to them. It is not enough to just provide the most accurate prediction possible. For example, let's imagine a product that can only be delivered in packages of 20. In this case, it is clear that providing a very precise prediction will not necessarily be useful. Whether you predict 82 or 89 sales, you will only be able to deliver 80, 100 or 120 units anyway. Even in the general case, the manager who has to decide what his store needs will generally not follow the model's predictions exactly. He is under no obligation to use the results. It is therefore essential to allow the business to have confidence in the model.

One of the possibilities for this is explainability. This means being able to explain the results of the model to the business and to answer questions such as: why do we have a high prediction today?

Explainability is a problem that needs to be asked at the beginning of the process because it guides the choice of the type of model. For example, using deep learning could improve the results but would greatly complicate explainability. Another possible example is delivery times. Before asking the question of modeling, one must, as we have seen, know the subject of prediction well. It would be totally inappropriate to predict weekly sales if deliveries are made on a daily basis and, in the opposite direction, inefficient.

2.2 Adapting to the product

However, all this is not enough. In addition to taking into account the needs of the business, it is necessary to adapt to the products. Again, predicting accurately is not enough because it is not necessarily what the business is interested in, and it will depend on the type of product. One of the main points will be the possibility of storage. Some products can be kept for a long time on the shelves or in stock. Others, on the other hand, are limited by their expiration date or the amount of space they take up in a warehouse. For some of these products, over-prediction is a serious problem. For others, under-prediction leading to stock-outs will be the only problem, or at least the biggest one.

2.3 The difference between sales and demand

Another question to ask when modeling is: what are we really trying to forecast? There are two possible answers to that: demand or actual sales. In the first case, we try to determine the quantity that customers would like to buy. In the second case, we take into account other limiting factors such as a limited stock. The best solution will often be to predict the demand. The forecast can always be adapted to reality later on. A concrete example of this problem is the stockout. Let's imagine a product that was out of stock last week. It has therefore made zero sales each day of it. In the next prediction, it will be better to provide the actual demand prediction for this product, whether there is also a shortage this week or not. So we don't want the model to anticipate the stockout. If there is indeed a stockout, there will be a strong over-prediction. However, it will have no impact since no delivery has taken place. All this reasoning applies if what we are interested in is mainly inventory management. On the contrary, in the case of a pricing problem for example, it would be much more appropriate to predict the actual sales.

3 Review of forecasting metrics

What we have seen so far has highlighted the fact that the most important thing to keep in mind when building a sales forecasting model is the problem you want to address. The most direct way to adjust the model to what you really want, whether it is to fit your business needs or to fit the specificities of your products, is the choice of the metric. First of all, the metric is a notion that is bound to evolve as a project progresses. For a private company, the ideal metric would be the net gains generated by the proposed solution. However, this is of course very difficult to obtain, even only to estimate, and impossible at the beginning of the project. Therefore, one should start by using a much simpler metric. We can classify metrics into three main categories: those based on error, those based on bias and those based on volatility. The first are the most common. They consist in looking at the difference between the forecast and the actual number of sales. The most basic metric is the MAE (Mean Absolute Error):

$$MAE = \sum_{i=1}^{n} |P_i - V_i| \tag{1}$$

where P and V are the prediction and the actual number of sales respectively.
A common adjustment to error metrics is scaling. It is easy to see that predicting 5 sales instead of 20 is much worse than predicting 245 sales instead of 260. To remedy this, one can divide by the scale, which is the number of actual sales. If we apply this to the MAE, we create the MAPE (Mean Absolute Percentage Error), one of the most widely used metrics:

$$MAPE = \sum_{i=1}^{n} \frac{|P_i - V_i|}{V_i} \tag{2}$$

Depending on the context, it is also sometimes wise to give more importance to high values. For this, we can perform another transformation to the MAPE to obtain the WMAPE (Weighted Mean Absolute Percentage Error):

$$WMAPE = \frac{\sum_{i=1}^{n} |P_i - V_i|}{\sum_{i=1}^{n} V_i} \tag{3}$$

Retail is a context in which it is interesting to give more importance to products with the highest sales volumes. The most suitable metric here will therefore often be WMAPE.

However, error metrics are not the only ones that exist. We must not forget the bias metrics and the volatility metrics. The former allow us to prevent over-prediction or under-prediction. The bias is then :

$$B = \sum_{i=1}^{n} P_i - V_i \tag{4}$$

Volatility allows us to quantify the extent to which the prediction evolves during the generation of a new forecast. A simple measure at time t is :

$$V^t = \sum_{i=1}^{n} |P_i^{t+1,t} - P_i^{t+1,t-1}| \tag{5}$$

where $P_i^{t+1,t}$ is the prediction of step t+1 made at step t.

In practice, an error metric will often be used as the main metric, allowing to judge the quality of a model at first sight and sometimes having also an importance in the optimization phase.

4 Uncertainty

Very often for this subject, another problem can be very interesting to dig into: that of uncertainty. Indeed, it is not enough to give the business only a raw number without giving it a glimpse of the confidence we have in this result. Let's imagine that we predict a total number of sales of 100 for a certain product. It would be totally possible that there is a 10% probability that the number of actual sales will be higher than 110, just as it would be possible that this number will be 150. However, the manager's response in both cases would be totally different. For a product that can be easily stored in large quantities, it will be preferable to plan for more volume in the second case. It is therefore important, if not necessary, to complete the result of the simple forecast, which is also called a point forecast, with information on the uncertainty of this result.

This information can take many forms. However, two solutions are more commonly used: the probabilistic forecast and the confidence interval. The first consists in providing a probability distribution associated with each result. This is obviously only applicable if you are working with integers. We then obtain a result similar to this one:

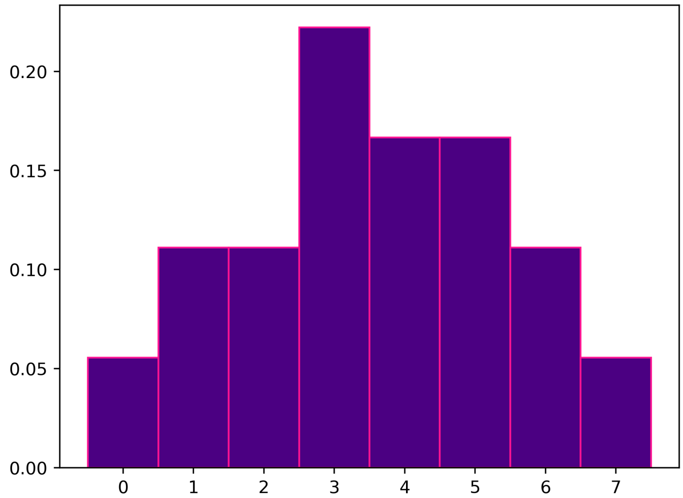

Figure 1: Probabilistic forecast example

This is the representation that gives the most information. It covers all possible scenarios. However, it has some drawbacks. On the graphical example, there are only 8 different possible values, which gives relatively high probabilities. In a case where the volumes would be more important, there are too many possible values and thus probabilities are too low and lose their meaning. It is then necessary to group the values. In the context of mass distribution, the volume of sales varies greatly depending on the product. It is therefore necessary to adapt the grouping to each product. This creates significant complications at the time of modeling.

With the confidence interval, part of the information is lost, but this makes it easier to read and simpler to use. For each output, in addition to the point forecast, two values corresponding to two quantiles of the probability distribution of the result are provided. These two quantiles form a confidence interval:

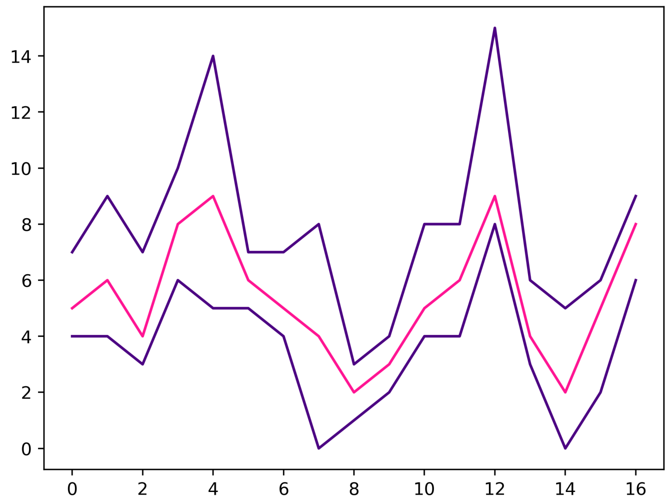

Figure 2: Confidence interval example

An important question then arises: which interval to choose? There is no right answer to this. The higher the interval, the more certain the decision will be. The smaller the interval, the more information it will provide. To make this decision, it is often appropriate to discuss with the business. The solution can sometimes be to provide several intervals. In this case, we sacrifice a little bit of readability for more information.

5 State-of-the-art methods : the M5 competition

After having discussed the important considerations for a retail forecasting practitioner, it could be interesting to talk about the methods that provide the best results for this task. One way to do so is to look at challenges. It allows us to see good approaches and interesting findings. There is one, very important in retail, that provides a lot of useful insights to practitioners, the Madridakis Challenge.

The Madridakis Open Forecasting Center (MOFC) is a forecasting research center based at the University of Nicosia in Cyprus. This institution is particularly known for its competitions, the first one having been held in 1982. Each new edition allows to see the evolution of forecasting techniques and to compare the best. The last one, the fifth edition, took place in 2020 and therefore gives an interesting insight into the current performances. This competition was held on Kaggle in partnership with Walmart, the largest retailer in the US. This edition actually had two separate competitions. The first one was called M5 Accuracy (Makridakis et al. (2022a)) and consisted in making a point forecast of the sales of different products, at different mesh sizes, for the next 28 days. The data included explanatory variables such as prices, promotions, day of the week and special events. The second one was called M5 Uncertainty (Makridakis et al. (2022b)). It was conducted on the same perimeter and with the same data as the M5 Accuracy, but the expected results were an estimate of the quantiles of the probability distribution of sales.

Regarding the results, a first thing to highlight is the significant progress made since the last competition, the M4, which however only dates from 2018. To realize this, we compare the best results to reference models. This shows how forecasting is a field that evolves at a very high speed. Secondly, the results clearly show that the ranking of the models according to their performance evolves significantly with the level of aggregation. Therefore, there is no magic solution to all the problems. It is necessary to adapt. The most important point to take away from these competitions is the following: Machine Learning methods proved to be far superior to the others. While M4 had already marked a revolution by having in the first two places two hybrid methods including partly

Machine Learning, M5 marks the advent of this type of models for forecasting problems. Indeed, all the best performing methods are exclusively based on Machine Learning. LightGBM is a Gradient boosting algorithm that has been observed in many of the best solutions. We will come back to Gradient boosting and LightGBM in more detail later. We have also seen a significant number of methods including Deep Learning. This is also accompanied by the highlighting of the superiority of cross-learning. The 50 best solutions of the M5 Accuracy used all the information present in the dataset, and not a trained "series-to-series" approach such as time series methods for example. The latter are sufficient to highlight the level, trend and seasonality but do not take into account the exogenous explanatory variables. Moreover, retailing is an area in which these variables have a preponderant importance, in particular because of promotions or price changes. The difference between the best solutions, using Machine Learning, and the reference models is even higher for the M5 Uncertainty. This can be explained by the fact that less research has been done in the uncertainty domain, which leads to less good reference models. This competition also confirmed the importance of model combinations, even simple ones with equal weights. Indeed, the top three in the M5 Accuracy used a simple arithmetic mean of several models to obtain their final prediction.

It may be interesting to detail the solution of the winner of the M5 Accuracy, which represents well the lessons of this competition. As stated earlier, his solution is based on an arithmetic average of different models. They were actually all LightGBM models. He grouped the data by store. There were 10 for this competition so that created 10 different models. He did the same at the store-product category aggregation level (30 models) and at the store-product department level (70 models). In addition, for each model there was a recursive and a non-recursive approach. The recursive approach means that when we want to predict at horizon n, we take into account the forecasts for horizons n-1, n-2,...etc which are in the future. This makes a total of 220 models and each forecast is based on the average of 6 models. He then performed an optimization based on the Tweedie distribution, specifically adapted to data sets with a high mass in zero, which is the case here and often in retailing. Indeed, at the daily grid, many products are not sold.

6 Conclusion

Sales forecasting in the retail industry is a problem with many specificities. The most important aspect to keep in mind before and during the construction of a model is the objective that is targeted. This includes thinking about who is going to use the future forecasting tool and what the needs of these people are, adapting to the different types of products on which the work will be done or differentiating sales from real demand. Still in this logic of determining the objective, an essential step is the choice of the metric. We have detailed the characteristics of the most commonly used metrics. In order to be able to fully anticipate future customer behavior and respond as effectively as possible, a point forecast is not enough. It is essential to introduce the notion of uncertainty into a solution. The two main categories, probabilistic forecasting and confidence intervals, each have their qualities and drawbacks.

Finally, the analysis of the M5 competition, specialized in the field of mass retailing, allowed us to highlight the most effective modeling solutions to this problem. One in particular emerged from the competition, the use of Gradient boosting, mainly through the LightGBM library.

References

Fildes, R., S. Ma, and S. Kolassa (2022): "Retail forecasting: Research and practice," *International Journal of Forecasting*, 38, 1283–1318, special Issue: M5 competition.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos (2022a): "M5 accuracy competition: Results, findings, and conclusions," *International Journal of Forecasting*, 38, 1346–1364, special Issue: M5 competition.

Makridakis, S., E. Spiliotis, V. Assimakopoulos, Z. Chen, A. Gaba, I. Tsetlin, and R. L. Winkler (2022b): "The M5 uncertainty competition: Results, findings and conclusions," *International Journal of Forecasting*, 38, 1365–1385, special Issue: M5 competition.

Seaman, B. (2018): "Considerations of a retail forecasting practitioner," *International Journal of Forecasting*, 34, 822–829.