

Rethinking Inventory Forecasting Problems in E-commerce: Exploring the effect of integrating forecasting and inventory decisions

Hernan Ceferino Vazquez, Virginia Dal Lago

MercadoLibre Inc.

Applied Machine Learning Research

hernan.vazquez@mercadolibre.com, virginia.dallago@mercadolibre.com

Abstract

Forecasting for inventory control is the process of calculating the inventory needs to fulfill future consumer demand. In general, this process is divided into two sub-processes. The first sub-process receives the current inventory information and forecasts future information, e.g. forecasts future demand from the demand information in the past. The second sub-process uses the forecast information as input to make inventory decisions, e.g. use a product demand forecast to decide how many units of this product to buy. Recent works highlight the importance of integrating forecasting with final inventory decisions, however, there is very little empirical evidence to support that integrating the decision is the best solution. In this work, we propose to explore the effect of integrating the inventory decision into the forecasting problem and compare it with the state-of-the-art approaches. For this, we evaluated the approaches in different operational tasks belonging to our business. Our preliminary findings show that predicting operative decisions instead of demand information could be better and the benefit can be capitalized even in low data scenarios.

Introduction

In recent years, e-commerce has developed rapidly. Consumers' buying behavior has changed, more people prefer to buy and sell in digital stores from the comfort of their homes. This has led to the emergence of new platforms that transformed the B2B and B2C paradigms, enabling consumers to participate as sellers in a C2C model (Gupta 2014). This brings new challenges for operations, among which is being able to carry out an efficient control of the inventory (Patil and Divekar 2014).

Inventory control is about all the operational decisions that are made to satisfy consumer demand and to be efficient in optimizing operations related to the management of warehouses, shipping, among others (Barwa 2015). In general, inventory control is carried out at the SKUs (stock-keeping units) level. These inventories are usually stored in warehouses ready to meet customer demand. The warehouses can also be strategically located in different geographical areas to be able to satisfy the demand at a required service level and/or a budget. In this way, in e-commerce businesses, efficiency in operational decisions is tied to being able to know

the demand for the products offered in it. However, this demand is generally unknown when making operational decisions and is therefore forecast.

In this context, forecasting is usually defined as the action of predicting how much demand is going to be for a particular SKU (often using sales as a proxy) at some future point or period time (Thomopoulos 2015). However, this look at inventory forecasting could be a short-sighted vision of the problem (Kourentzes, Trapero, and Barrow 2020). The argument is that inventory forecasting needs to be any predictive decision that makes e-commerce operations more efficient rather than just information for others' decision-making processes. In this paper, we support this idea, adding that predictive performance can also be improved if decisions are predicted instead of demand information. For this purpose, we explore the results of different approaches and techniques to quantify the difference over several inventory forecasting tasks. Our findings show that predicting decisions instead of information is better in almost all explored tasks and that the improvement is also evident even in scenarios where the amount of data is low.

The rest of the paper is structured as follows. Section 2 presents the background concepts used in this work. Section 3 describes the experimental setting we conducted. Section 4 presents the result analysis and highlights the benefits of predicting inventory decisions. Section 5 concludes and outlines our future work.

Background

Inventory forecasting can be viewed as a complete process in which the following inventory decisions are obtained from current inventory information (Goltsos et al. 2021). In practice, it is very common to find this problem divided into two sub-processes: The first sub-process receives the current inventory information and forecasts future information, e.g. forecasts future demand from the demand information in the past. The second sub-process uses the forecast information as input to make inventory decisions, e.g. use a product demand forecast to decide how many units of this product to buy.

As a forecasting problem, inventory forecasting is a bit different from other forecasting problems. Some of the main characteristics that describe an inventory forecasting problem are the following:

- It is a multivariate problem, since each product in the inventory can be described by more than one time series, for example, the time series of units available for sale, the time series of sales, the time series of the price, etc. . .
- The product time series can be correlated with other product time series, for example, the sales of a product can be highly correlated with the sales of another product within the same domain or category
- Time series have different degrees of intermittence, for example, certain products may have consistent unit amounts over time and other products may have units from time to time because they are seasonal.

The state-of-the-art approach to solve inventory forecasting problems is to divide the process, forecast product demand, and then use that information to make inventory decisions. It has the main advantage of producing a general forecast that can be used in different scenarios. Some of the techniques to solve inventory forecasting into this framing were well described in the past (Spiliotis et al. 2020). These techniques can be divided into two main categories: single-learning and cross-learning (Semenoglou et al. 2021).

Single-learning, are forecasting models that are trained in a series-by-series fashion, both statistical and machine learning ones, are the most used in forecasting. One reason for this can be attributed to many different factors related to the nature and the historical success of these models, as well as to the availability of data and computer resources over the years (Makridakis, Spiliotis, and Assimakopoulos 2021).

Cross-learning techniques are forecasting models that use information from multiple series when training forecasting models. In particular, some of these models can be effectively applied for forecasting numerous, mostly unrelated series, such as those found in inventory forecasting problems. Recently, based on the findings in M competencies (Makridakis, Spiliotis, and Assimakopoulos 2021), the effectiveness of these types of techniques where used to forecast demand has been proven.

Unfortunately little is known about the effectiveness of this approach in consequent inventory decisions. Recent work highlights the importance of integrating forecasting with inventory decisions, however, there is no empirical evidence to support that integrating the decision is the best solution.

In this work, we propose to explore the effect of integrating the inventory decision into the forecasting problem and compare it with the state-of-the-art approaches. For this, we will evaluate the forecast in different downstream tasks belonging to our marketplace and we will evaluate them with different operational efficiency metrics.

Experiment setting

Our objective is to investigate the effect of directly predicting inventory decisions rather than predicting demand and then making decisions based on it. To accomplish this objective we consider the following inventory forecasting tasks.

Inventory forecasting tasks

To evaluate our approach, we take three inventory forecasting tasks from our company. For each task, we attach an operational related metrics to understand what is the real value of the forecast. Two of these tasks have an associated asymmetric cost. For example, it is usually more expensive to lose a sale than to store one more unit. We reflect this asymmetry on an ad-hoc basis, making the error of not having one unit in the warehouse twice that of having one more. We called this error, asymmetric inventory error (AIE), and we calculated for each item as follow:

$$AIE = \sum_{day=1}^H 2 * deficit_{day} + surplus_{day} \quad (1)$$

where H is the forecasting horizon, $deficit_{day}$ is the negative difference between sales and the available units of the item in a particular day, and the $surplus_{day}$ is the positive difference between them. We model AIE as a single linear function for simplicity, however in real-world warehouses, different products can be modeled with different functions.

Below we describe the tasks and the metrics we use to evaluate them. For each item, we take as the number of perfect units (ground truth) the number of units that were finally sold on the expected horizon.

Inbound Restriction E-commerce sites that offer a service to store inventory to sellers face the problem of having to manage warehouse space. Moreover, it is very common for sellers to take more units than they need to the warehouse to try to do it as few times as possible. An operational decision to improve efficiency is to restrict the entry of units to be stored when the seller arrives at the warehouse. We call this task Inbound Restriction.

Accordingly, given an inbound intention (int), what we are trying to predict is how many units to restrict access to. After applying the restriction ($restr$) we measure the success of the task using the AIE.

$$\sum_{item=1}^N AIE((int_{item} - restr_{item}) - ground_truth_{item}) \quad (2)$$

Warehouse Replenishment Warehouses are places where the available units of each item are stored. These warehouses have limited space, so understanding which items to fill them with is important for operational management. Consequently, given a total capacity of space, simplified as the number of spaces available to store items, the task consists of determining the quantity of each item that minimizes the AIE.

$$\sum_{item=1}^N AIE(quantity_to_store_{item} - ground_truth_{item}) \quad (3)$$

In this case, the sum of the units to be stored cannot exceed the total capacity of the warehouse.

Out Of Stock Alert Another common task in e-commerce inventory management is to alert sellers when they are running out of stock. To do this, we consider the number of units available for an item, and what we want to predict is whether in an H number of days the seller will run out of units of that item. To evaluate this task, the difference between the units available and those sold is considered. If the difference is negative, we will send an alert. If the difference is positive, no. Since it is a binary problem we compute the accuracy of the decision to send an alert.

$$Accuracy(\text{remaining_units}_{item} < \text{ground_truth}_{item}) \quad (4)$$

Models

Additionally, we want to do this in the simplest way possible in terms of techniques and data complexity. To accomplish this objective we consider the following scenarios:

- Traditional series-by-series models to predict demand
- Single-learning ML models to predict demand
- Cross-learning ML models to predict demand
- Single-learning ML models to predict decisions
- Cross-learning ML models to predict inventory decisions

Since what we are pursuing is evaluating the framing of the problem (and not the models), we will use for all machine learning experiments the same class of models, Histogram-Based Gradient Boosting Machines, in particular the LGBM implementation. This family of models has proven to be one of the best in solving demand forecasting problems, as seen in the M5 competition. In addition, as traditional models, we consider AutoARIMA, Exponential Smoothing, and Prophet.

Data

The data set used for the inventory forecasting task focuses on one mid-size country site only. Given the large number of SKUs that we have per site, we only take into account SKUs that have had at least one sale per week in the established time range. We took six months between January 1 and July 1 of 2021, about 5000 SKUs. Each one of these is described day by day with the attributes shown in table 1. For simplicity, we do not consider data referring to the title and description of the item, nor referring to the domain or category.

Data preparation

Each task was split into train, test, and validation subsets. The validation was used for hyperparameter tuning and report results. For traditional models, we only consider the sold quantity series to predict demand. For machine learning models we transform the problem in a supervised fashion for each task. To do this we choose to use a "last window" strategy, with a window size equal to the horizon, in our case 28 days. A training example of the supervised dataset will contain, for each attribute (i.e. sold quantity, current inventory, etc.), the last 28 days concatenated with several summary features (mean, std, min, max, median, var, skewness, and

Table 1: Set of product attributes that are recorded for each day.

Attributes	Description
SKU	indicates the SKU to which the record belongs
sold quantity	number of units of the corresponding SKU that were sold on that particular date.
current price	point in time correct SKU's price.
listing type	relate to the exposure the items have and the fee charged to the seller as a sales commission (e.g. classic, premium).
shipping logistic type	type of shipping method the SKU offered (e.g. fulfillment, cross-docking, and drop-off)
shipping payment	whether the shipping for the offered SKU at that particular date was free or paid, from the buyer's perspective.
minutes active	number of minutes the SKU was available for purchase on that particular date.

kurtosis). In addition, we add to these features the mean and std related to the objective of the task (i.e. the inbound intention, the remaining stock, etc.). To train a model in the form of cross-learning, we use a supervised dataset with all SKUs, otherwise, the supervised dataset is divided by SKU and trained only with the data of the SKU to be predicted.

To see the effect of cross-learning, we run experiments using different amounts of SKUs for each task (10, 100, 500, 1000, 2500, 5000). We also run 5 experiments per combination of tasks, SKU amounts, and model, using different random samples and initialization of the data. We report the mean and standard deviation of our results.

Results

Table 2 shows the metric values for all the scenarios proposed. The values in bold in the columns correspond to the best value for the metric along that column. As it can be observed, *CL-decision-LGBM* shows the best results for all the scenarios. Cross-learning shows to be better than single-learning for all the tasks. At the same time, single-learning techniques are better than traditional ones except for the inbound restriction task in the lowest data scenario where AutoARIMA obtains the best result. Something important to observe from the results is that, in general, predicting decisions instead of demand is better within the same group of techniques (i.e. single or cross-learning). However, for the cross-learning setting, the difference tends to decrease when more data is used. Perhaps one of the reasons for the small difference in the scenarios with more data is due to the model selection. Another possible reason could be not using the operational metric (e.g. AIE) in the cost function

Inventory Task	Train Size Model	10	100	500	1000	2500	All Data
Inbound Restriction	AutoARIMA	0.37±0.21	0.30±0.06	0.28±0.05	0.25±0.01	0.26±0.01	0.27±0.01
	ExponentialSmoothing	0.28±0.04	0.29±0.07	0.27±0.05	0.25±0.01	0.25±0.01	0.26±0.01
	Prophet	0.27±0.10	0.38±0.06	0.38±0.06	0.34±0.02	0.34±0.01	0.36±0.01
	SL-demand-LGBM	0.39±0.14	0.38±0.10	0.40±0.08	0.40±0.07	0.41±0.07	0.41±0.06
	SL-decision-LGBM	0.48±0.04	0.49±0.04	0.51±0.04	0.52±0.04	0.52±0.03	0.52±0.03
	CL-demand-LGBM	0.77±0.18	0.79±0.13	0.80±0.11	0.81±0.10	0.82±0.09	0.82±0.08
	CL-decision-LGBM	0.81±0.07	0.83±0.06	0.83±0.05	0.83±0.05	0.83±0.04	0.84±0.04
Out Of Stock Alert	AutoARIMA	0.56±0.16	0.45±0.04	0.48±0.02	0.47±0.01	0.47±0.01	0.48±0.01
	ExponentialSmoothing	0.52±0.24	0.45±0.05	0.47±0.02	0.46±0.02	0.47±0.01	0.47±0.01
	Prophet	0.50±0.14	0.41±0.05	0.45±0.01	0.44±0.01	0.45±0.01	0.45±0.01
	SL-demand-LGBM	0.48±0.19	0.48±0.14	0.49±0.12	0.50±0.10	0.50±0.09	0.50±0.08
	SL-decision-LGBM	0.50±0.10	0.55±0.09	0.56±0.08	0.57±0.07	0.57±0.06	0.57±0.06
	CL-demand-LGBM	0.55±0.15	0.56±0.11	0.56±0.09	0.56±0.08	0.56±0.07	0.55±0.06
	CL-decision-LGBM	0.68±0.17	0.69±0.13	0.72±0.12	0.73±0.10	0.74±0.10	0.75±0.09
Warehouse Distribution	AutoARIMA	0.31±0.10	0.28±0.04	0.22±0.04	0.21±0.01	0.22±0.01	0.21±0.01
	ExponentialSmoothing	0.25±0.09	0.27±0.05	0.21±0.04	0.19±0.01	0.20±0.01	0.20±0.01
	Prophet	0.23±0.12	0.24±0.04	0.19±0.02	0.18±0.01	0.18±0.01	0.19±0.01
	SL-demand-LGBM	0.46±0.09	0.44±0.07	0.45±0.06	0.45±0.06	0.45±0.05	0.46±0.05
	SL-decision-LGBM	0.50±0.10	0.47±0.08	0.47±0.07	0.47±0.06	0.47±0.06	0.47±0.05
	CL-demand-LGBM	0.56±0.09	0.55±0.07	0.56±0.07	0.57±0.06	0.58±0.06	0.59±0.06
	CL-decision-LGBM	0.64±0.01	0.62±0.03	0.60±0.04	0.59±0.04	0.59±0.04	0.60±0.03

Table 2: Forecasting performance across 5,000 SKU daily series and different training sizes of the various forecasting approaches considered in this study. For AIE based task the score in the table is calculated as $1 - (\text{AIE} / \max(\text{AIE}))$. For all tasks, the mean and standard deviation of the results are shown using different random seeds and samples.

and directly optimizing it. In that case, we believe that the comparison would not have been entirely fair for the techniques that predict demand. Beyond that, we believe that in practice it would be a good way to boost the performance of the models that predict the operational decision.

Conclusions

In this article, we propose to explore the effect of integrating inventory decisions into the forecasting problem by comparing various techniques and scenarios. We assess the performance in different inventory tasks belonging to our marketplace. Consequently, an initial evaluation showed satisfactory results when predicting decisions in all the studied tasks. Additionally, cross-learning scenarios are better than single-learning and traditional techniques. Nevertheless, the difference tends to decrease as the data increases. One of the reasons for this may be due to the models used or not having directly optimized the operational metrics. In future work, we will focus on cross-learning methods and we will evaluate the use of different modeling techniques instead of only LGBM. In addition, we will explore different cost functions to optimize directly the operational metric. To do so AIE still needs further validation to more faithfully represent the objectives of each operational problem. Finally, it would be interesting to add new tasks and datasets to our evaluation to extend the scope of our findings.

References

Barwa, T. M. 2015. Inventory control as an effective decision-making model and implementations for company's

growth. *International journal of economics, finance and management sciences* 3(5):465.

Goltsos, T. E.; Syntetos, A. A.; Glock, C. H.; and Ioannou, G. 2021. Inventory-forecasting: Mind the gap. *European Journal of Operational Research*.

Gupta, A. 2014. E-commerce: Role of e-commerce in today's business. *International Journal of Computing and Corporate Research* 4(1):1-8.

Kourentzes, N.; Trapero, J. R.; and Barrow, D. K. 2020. Optimising forecasting models for inventory planning. *International Journal of Production Economics* 225:107597.

Makridakis, S.; Spiliotis, E.; and Assimakopoulos, V. 2021. Predicting/hypothesizing the findings of the m5 competition. *International Journal of Forecasting*.

Patil, H., and Divekar, B. R. 2014. Inventory management challenges for b2c e-commerce retailers. *Procedia Economics and Finance* 11:561-571.

Semenoglou, A.-A.; Spiliotis, E.; Makridakis, S.; and Assimakopoulos, V. 2021. Investigating the accuracy of cross-learning time series forecasting methods. *International Journal of Forecasting* 37(3):1072-1084.

Spiliotis, E.; Makridakis, S.; Semenoglou, A.-A.; and Assimakopoulos, V. 2020. Comparison of statistical and machine learning methods for daily sku demand forecasting. *Operational Research* 1-25.

Thomopoulos, N. T. 2015. Demand forecasting for inventory control. In *Demand forecasting for inventory control*. Springer. 1-10.