Analyzing Out Of Stock Patterns for a Consumer Goods Company

By Xu (Tony) Li
Advisor: Jim Rice & Sergio Caballero

Topic Areas: Inventory, Distribution, Metrics

Summary:

Tony Li has managed logistics planning, execution, compliance and process improvement for various industries, such as medical device, chemical and retail, in both Asia and North America. He received a Bachelor's degree in Business Administration from Nanjing Normal University.

KEY INSIGHTS

1. By normalizing dataset, same or similar shapes of patterns can be identified and aggregated. This method could possibly be scaled to all 432 GTINs to aggregate patterns from 4 million transactions, with an output of small sets of repeated patterns.

2. Steep drops seem to be infrequent events. As a result, out of stock seems to be unpredictable solely based on DC data.

3. Using the identified Pattern I dataset, store data could be further explored to draw links between the POS and OOS.

Introduction

The sponsor company is a global consumer packaged goods manufacturer (CPGM), headquartered in North America. The CPGM manufactures products, stores finished goods mainly in mixing centers, and ships to large retailers across the country. Shipments typically go to retailers' distribution centers (DCs), and then products are distributed to each store from there.

The particular problem presented is that the CPGM's retailer customers have been facing frequent out of stock (OOS) events at their DCs. Typically, each GTIN (Global Trade Item Number, a universally accepted unique identifier for each product being sold) is associated with 42 DCs.

Stock outs could be more frequent at some DCs during certain time periods, such as holidays. The CPMG is concerned about OOS events because they can lead to lost sales.

The sponsor company believes that there is a pattern for inventory drops. The objective of this research is to identify whether there actually is an identifiable pattern, either sudden or gradual, especially in the last two days prior to the OOS event. If there is a sudden inventory drop in the two days prior to an OOS, then actions could be taken to replenish inventory to the DCs in order to minimize the impact of an OOS.

Data samples

For this project, one product line was studied. Due to the complexity of the data structure and data volume, 20 representative GTINs were selected and further analyzed. An example is shown in Figure 1. Each GTIN contains inventory on hand information for all 42 DCs, with each DC containing one year's worth of time-series data points. It would be nearly impossible to manually go through each DC for each GTIN over one year's worth of data to detect a series of potential patterns, not to mention that there are 432 unique GTINs in the given product family. The range of inventory value is between 0 and 600, which makes pattern recognition even harder.
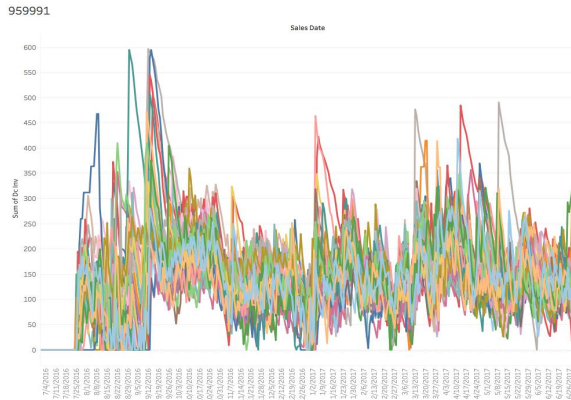
959991

Figure 1: Example of Time Series for one GTIN

Literature review

This section gives an overview of prior published research that has considered this problem. The Matrix Profile analysis method described by Yeh et al. (2016) of the Computer Science and Engineering department of the University of California Riverside is a time-series data-mining algorithm for pattern recognition. The reason to study a Matrix Profile is that it normalizes a dataset and converts each data point to an index (or Euclidean distance). Because the range of inventory values for the product studied greatly varies for different DCs and GTINs, it would be very difficult to search similar patterns using absolute inventory numbers. Furthermore, shapes, not exact value, are concerned in aggregating various patterns. In this research, a similar method was used to normalize datasets and then use the resulting indices to search similar patterns in new datasets.

Methodology

With inventory-level data provided by the CPGM, eight types of potential stock-out patterns were predefined and tested. Further study was based on three types of high-frequency patterns, including one in particular: inventory gradually decreasing to an OOS event within a certain number of days (7, 6, 5, and 4 days).

In order to identify a list of potential patterns, this study uses both index and similarity search methods. To avoid certain values dominating the whole dataset, the data is first normalized by converting each individual inventory value to an index. The next step is to use a predefined range of indices to search for similar patterns in the new dataset. For a specific pattern, both indices and absolute inventory values are used to identify when inventory is gradually declining without any replenishment shipment being received during the same time period.

There are five steps in order to search similar patterns:

1. Calculate the average inventory level within each subset of time series (length of subsets = 7, 6, 5 and 4 days);
2. Divide each inventory level by the average inventory level in order to obtain the index for each row;
3. Compare each index to the interval of the predefined index range: If each index is within the lower bound and upper bound of the predefined index range, then a pattern is identified, indicated, and recorded in the new dataset;
4. Slide the subset until the end of the time series in the same DC data;
5. Repeat the same steps for all 42 DCs' data.

Essentially, the same steps are repeated for all 20 representative GTINs. Furthermore, Python is used for this specific calculation. It takes about 20 minutes to calculate one GTIN and export files with indicators of similar patterns.

Reducing eight patterns to three patterns

In this research, eight different types of patterns were preselected based on the initial observations and feedback from the sponsor company. From a high level, five patterns describe how long inventory can last until the next stock out and whether replenishments could be associated with a single or a series of stock outs, whereas the other three patterns focus on behaviors between stock outs and replenishment orders.

The similarity search is conducted using ranges of index value. This search method is tested with five unique GTINs for all 42 DCs. Initial observations indicate that certain patterns happen more frequently than others. Pattern I, II and III are relatively high-frequency patterns among all eight patterns. As a result, they were selected and further analyzed. Pattern I represents behavior which inventory decreases before OOS, without any replenishment order during the same time period. Patterns II and III show the potential impacts of the number of OOS days when the inventory lasts for three days and then is out of stock.
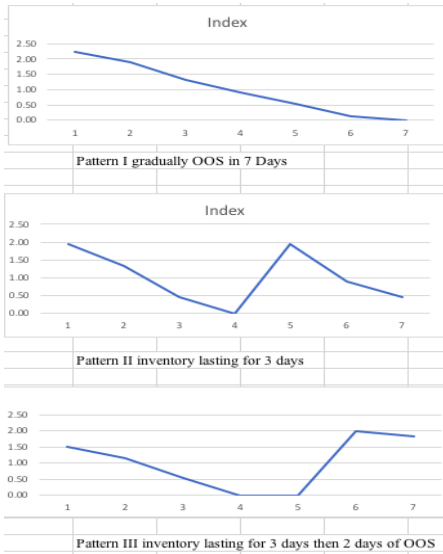
Figure 2: Three Patterns

7 Days		6 Days		5 days	
Bin	Frequency	Bin	Frequency	Bin	Frequency
0%	0	0%	0	0%	0
20%	8	20%	6	20%	5
40%	14	40%	17	40%	19
60%	4	60%	9	60%	21
80%	5	80%	6	80%	8
100%	0	100%	1	100%	5
More	0	More	0	More	0
	31		39		58
Average drop		Average drop		Average drop	
Per day	17%	Per day	20%	Per day	25%
For the last 2 days	33%	For the last 2 days	40%	For the last 2 days	50%
>70%	1	>70%	4	>80%	5
>70%	3%	>70%	10%	>80%	9%

7 Days		6 Days		5 days	
Bin	Frequency	Bin	Frequency	Bin	Frequency
0%	0	0%	0	0%	0
20%	4	20%	3	20%	4
40%	8	40%	7	40%	9
60%	1	60%	10	60%	14
80%	1	80%	2	80%	4
100%	0	100%	0	100%	0
More	0	More	0	More	0
	14		22		31
Average drop		Average drop		Average drop	
Per day	14%	Per day	17%	Per day	20%
For the last 2 days	29%	For the last 2 days	33%	For the last 2 days	40%
>70%	0	>70%	1	>80%	1
>70%	0%	>70%	5%	>80%	3%

Figure 3: Histogram for two GTINs

Pattern I and steep drops

Pattern I (gradual OOS) is selected, since Pattern I will not change the same shape when the m value becomes smaller, whereas both Pattern II and Pattern III will change their shapes when the m value becomes smaller. The inventory level gradually decreases towards zero (within 7, 6, 5, and 4 days). There is no replenishment order within the defined time interval. A gradually decreasing Pattern I is captured by two criteria: index and absolute value.

As mentioned earlier, the hypothesis of this research is that if there is a pattern for a sudden inventory drop, then the OOS could potentially be predicted. In this study, the inventory drop of the last two days is compared to the inventory level's starting point. For instance, for a particular Pattern I starting with 142 units of inventory, at Day 5 only 53 units are left in that DC, meaning that 37% of the total inventory dropped on Day 5 and Day 6 prior to the OOS event. With the calculated percentage, a histogram is further developed (see Figure 3). For the first GTIN, the inventory drop that is between 0 and 20% occurs 8 times; the inventory drops between 20% and 40% occurs 14 times.

A steep drop is defined as: for the 7- and 6-days patterns, the inventory reduces by more than 70% for the last two days prior to OOS; for the 5-days pattern, the inventory reduces by more than 80% for the last two days prior to OOS. Based on all 20 GTINs (Figure 3 show an example of two GTIN), steep drops are infrequent events, since the frequency is less than 10%.

During this study, store-level data was also analyzed and compared to Pattern I. One of the interesting findings is that POS data could be either higher or lower than the inventory starting level for the 7-days pattern. Meanwhile, the total store inventory level is still higher than the POS number, which implies that the inventory was not sufficiently distributed to the right stores. Accordingly, safety stock could be further communicated and discussed between the CPGM and retailers.

The potential insights of this study are: first, stock outs don't seem to be predictable based solely on DC data, since steep inventory drops appear to be infrequent events; and second, the sponsor company may be able to use the identified patterns to connect point-of-sale (POS) data with OOS events, and then identify the drivers of out of stocks.

Conclusions

Using the index and similarity search methods, a series of OOS patterns can be identified and aggregated in a large scale. Instead of manually going through each time series, a pattern can be generalized and described using both the index and the absolute inventory value. By studying a specific pattern, gradually OOS, it was found that steep drops seem to be infrequent events. The findings further indicate that stock outs don't seem to be predictable based solely on the DC data. Using the identified Pattern I dataset, store data could be further explored to draw links between the POS and OOS. Essentially, the firm could possibly use the index and similarity search methods to test more potential patterns and GTINs, and then aggregate by the GTIN-DC combinations. This method could possibly be scaled to all 432 GTINs to aggregate

patterns from 4 million transactions, with an output of small sets of repeated patterns. Based on the smaller dataset, the firm may be able to identify the drivers of out of stocks by connecting the POS data with OOS events.