# Research Paper
# Analytics for an Online Retailer: Demand Forecasting and Price Optimization

Krishna Kartik Darsipudi

June 29, 2018

## 1 Introduction

This paper written by Kris Johnson Ferreira, Bin Hong Alex Lee and David Simchi-Levi provides a novel method for an online retailer to use it's wealth of data and predict the demand as well as the optimum price of products which may or may not have been sold previously. The retailer that David et.al. have partnered with belongs to the online fashion sample sales industry and is called Rue La La.

Rue La La offers extremely limited-time discounts(”flash sales”) on designer apparel and fashion accessories. Rue La La has approximately 14% market share in the online fashion sample sales industry and only sells products online. Rue La La's website has several events, each of which represent a collection of for-sale products that are similar. When a user clicks on the event that he is interested, he is taken to a page with all the products that are being sold under that event. Each product on this event page is called a style. Each style has a number of different sizes that it is available in. Each of these size specific product is called a SKU (Stock Keeping Unit). Rue La La is associated with a number of designers. There designers directly ship the items to Rue La La warehouses and hand it over to the company merchants. Depending on the number and type of styles they have, the merchants decide on what future events to conduct. During an event, when a customer places an order for a particular SKU, it is shipped from the Rue La La warehouse to the customers. If the event ends or the particular style goes out of stock, the customer can no longer place an order for that item. Also, if the inventory is not completely used up then they will be incorporated into a future event. Styles being sold for the first time are called ”First Exposure Styles” and majority of Rue La La's revenue comes from these kind of styles. Rue La La's main challenge and also the challenge that the authors are dealing with in this research paper is pricing and predicting demand for these First Exposure Styles.

According to the data procured by the authors, more than half the First Exposure Styles are sold out before the end of the event. This motivates the idea to increase the prices of these First Exposure Styles as they will still be able to achieve high sell through. On the other hand, there still are a large number of styles that sell less than half the inventory by the end of the event. In this case, the motivation is contradictory to what was observed previously. These observations helped the authors understand the use of the prior data in order to construct a reliable pricing tool. The authors followed a twofold approach that begins with demand prediction model for the First Exposure Styles. The challenges that the authors faced wrt this were estimating lost sales due to stock-outs and predicting demand for items that have no historical data.This data is then used to maximize revenue by plugging them into the price optimization model. The base prediction model that the authors worked with are regression trees. This is because the non-parametric nature of regression trees doesn't assume the nature of the relationship among the various variables involved.

The challenges that the prize optimization model faced were that the prize of each style depended on the set of prices of the competing styles and hence each style could not be calculated by isolating it from the competition. Also, the non-parametric nature of the regression trees made the process of solving the optimization problem hard because there was no defined relationship between the predicted demand the other variable involved in its prediction.

# 2 Leagacy Pricing Process

Retailers usually base their prices on a combination of percentage markup on costs, competitors' pricing and intuition of the merchant about the best price for the product. Rue La La applies a fixed percentage markup on costs of all the styles and compares them to the competitors' pricing and chooses whichever is low. Fixed percentage markup is applied because identical styles are not normally found at the same time on competitor websites.

# 3 Demand Prediction Model

The authors have decided to aggregate the demand to the style level. This is because the prices of all sizes of a particular style are the same. Further rolling up the aggregation would eradicate the ability to compare the effect of the competing styles on the demand and prices. A few notations that have been followed throughout the paper are shown below

- $\mathscr{I}$ denotes the set of all styles that Rue La La sells.
- $\mathscr{S}(i)$ denotes all the sizes associated with the style $i$ with $s \in \mathscr{S}(i)$ being a specific size.

- The ordered pair $(i, s) \in \mathscr{I} \times \mathscr{S}(i)$ describes an unique SKU.
- The demand for a style without considering inventory constraints is given as $u_i$ and for a particular size is $u_{is}$.
- Inventory is held at the size level and is denoted by $C_{is}$.
- Total inventory for a size $i$ is $C_i = \sum_{s \in \mathscr{S}(i)} C_{is}$.
- Sales for each size of a style $i$ is $d_{is} = min\{u_{is}, C_{is}\}$
- Total sales for a particular style is $d_i = \sum_{s \in \mathscr{S}(i)} d_{is}$

Initially, all the normal features about the styles like quantity of SKU sold, price, event timings, product brand and size were already provided by the company. These features had a dearth of price related information, hence the authors decided to included a few price based features.

- discount off the manufacturer suggested retail price (MSRP)
- Relative price of competing styles. This metric was constructed by dividing the price of the style by the average price of the competing styles. Competing styles are all the other styles in the same event, basically all the other styles that the consumer can see in the same page as the style in question.

These features also show how the demand is dependant on the characteristics of the competing styles. The way in which relative price is constructed shows that the authors considered the average price of the competing styles as the reference for any style. This makes sense because consumers can see the prices of other styles on the webpage with ease. Difficulties in procuring external data for reference prices made the authors refrain from using them. Also, this is flash sale website and I believe that the reason stated previously for using a fixed percentage markup on costs can also be used to justify not using external prices. The styles that are sold in a particular event have very low chances of showing up elsewhere during the same time period and there's no guarantee that some other website would sell that style in a future time frame. So it is apposite to assume that the consumer would only look at the other styles on the same page competitors.

For items with stockouts, i.e. $d_{is} \leq u_{is}$ or in other words the sales underestimate the demand. Most of the approaches that deal with this involve a lot of sales data that unfortunately doesn't exist with Rue La La because they deal with minimum inventory. To tackle this the authors came up with a novel method. They used sales data from items that didn't stock out ($d_{is} = u_{is}$) to predict the demand for sales that did stock out. For every event hourly sales data from all the items that didn't stock out was aggregated to get the percent of sales that took place in that hour. This resulted in an empirical distribution of proportion of sales ($p$) for the first $X$ hours of the event. Each distribution was called a *demand curve*. This compensated for the modicum inventory by producing around 1000 demand curves. For more interpretablity, these curves were hierarchically clustered using Ward's minimum variance method. In general agglomerative hierarchical clustering the criterion to choosethe pair of clusters to merge is based on the optimal value of an objective function. Wards minimum variance method uses the square of the euclidean distance between the two cluster centers as the objective function. Whichever pair of clusters minimizes this objective function are chosen to merge. Hierarchical clustering was chosen to decide how many difference clusters the demand curves would have to be classified into. The number of clusters were observed to be 4. These clusters turned out to be based on the start times of the event and the day of the week. on further scrutinizing the demand curves it was found that there is a steep increase in percent sales at the beginning of the event as well as at 11AM the following day. This is the time when most of the events are started. The authors concluded from this that the shape of the curve is heavily dependant on the traffic pattern of the website. Now to estimate the demand of products that stocked out, the time at which the item stocked out was noted and the appropriate demand curve was referred, to find the proportion of sales that normally happened in that time. The demand is found as follows

$$Actual\ demand\ of\ the\ style\ (to\ be\ found) = u_i$$
$$Items\ sold\ in\ X\ hours\ of\ the\ event = d_i \leq u_i$$
$$From\ the\ demand\ curve, proportion\ of\ items\ sold\ in\ X\ hours = p = \frac{d_i}{\widehat{u_i}}$$
$$u_i \approx \widehat{u_i} = \frac{d_i}{p}$$

Demand and sales for new styles are predicted by constructing a regression model with the features and estimates of $u_i$ from historical data. This model was used to predict the demand of future first exposure styles, $\widehat{u_i}$. To do this the authors had access to size curves to estimate the proportion of demand of a style $i$ that had to be attributed to each of its size. Similar styles with similar sizes are said to be a *product type*. Some new notations introduced are as follows

- $\mathscr{T}$ denotes the set of product types.
- $t(i) \in \mathscr{T}$ denotes the product type associated with style $i$.
- $\mathscr{S}(t(i))$ denotes the set of all possible sizes associated with the product type $t(i)$
- $q_{(t(i),s)}$ denotes the percent of $\widehat{u_i}$ to allocate to size $s \in \mathscr{S}(t(i))$

Given $\widehat{u_i}$, $\widehat{u_{is}} = \widehat{u_i} * q_{(t(i),s)}$. So now, $\widehat{d_{is}} = min\{C_{is}, \widehat{u_{is}}\}$ and $d_i = \sum_{s\in\mathscr{S}(i)} d_{is}$. This demand and sales prediction for various regression models were compared on the basis of $R^2$, Median Absolute Error (MEDAE), Median Absolute Percent

Error(MEDAPE), Median Absolute Sell-through Error (MEDASTE). Regression trees with bagging consistently trumped its competition. The overfitting thats caused by allowing the trees to grow too large was overcome by 3 ways

- A branch was split only if the $R^2$ of the whole model increased by more than a *complexity parameter*.
- The minimum number of observations in the terminal node was put a lower bound by cross-validation.
- Bootstrap aggregation was used to build 100 regression trees to reduce variance. Average of all the tress was taken as the final demand.