Applying the new digital frontier. Optimisation of sales promotions using heuristic techniques and machine learning

Krzysztof Szlaski, University of Warsaw alumnus, Big Data engineer at P&G

At the verge of the 4th industrial revolution companies must adapt to a new digital environment that is mostly dependent on data and analytics. Companies protect their data more than ever as data are considered as the most valuable resource. This condition is due to the fact of creation of new data analytics tools which allow gaining meaningful business knowledge from data increase profit by several or over a dozen per cent. As a result, many companies and models which allow extraction of knowledge from data and also — advancement in computation powers of processing units which together allowed this progress. This type of innovation is essential to empowering companies to gain a market advantage in a resource-constrained world.

Due to that, the primary goal of this paper is to answer the following research question: Does the heuristic methods and out of the box machine learning can find empirical patterns and valuable business insight in the small dataset containing local sales data? Part of this approach is to make analyses that prove that the data provided can give some insight into customers and the product. In this article, however, we assume that this data is complete, and correct. The first section contains daily sales data for 12 different (Stock Keeping Unit) SKUs that were sold in 9 different stores. Also includes total beer sales and overall total sales for each store. In total, we have 18,043 observations in the coffee dataset and 50,922 observations in the beer dataset. The observations do not evenly represent the stores. The sales data for both datasets has the following characteristics: Shop number — determines which store the sale was made in; Date of sales — the date on which a specific product was sold in a specific store; SKUcategory — specifies the product name and EAN number. Value of sales — the sum of the daily sales of a specific product in each store. Number of the units sold — a unit sold each day in a specific store. Number of transactions — in how many transactions were total sales in each store.

For each dataset we have following information about the stores: Index number of the store. Area of the store — this metric is shown in square meters. The stores 1-3 are large-area stores and 4-9 are small area stores. Voidoship. Type of the town — this informs us on the segment of population of the town in which the town is located. We have 5 types of towns: city above 500,000 population, city between 100,000 and 500,000 population, city between 50,000 and 100,000 population, below 50,000 population and countryside.

The data was pre-processed for the search of missing data and outliers. For the purpose of the analysis most of the columns were dropped e.g.

Predicting the optimal period of the year in which the price promotion affects the sales value

1. We start the analysis and check which day of the month it is. We set a separate column in our dataset with those observations.
2. Then we use one-hot encoding to split our column in multiple features.
3. We create an algorithm which will mark our days as promotional. The algorithms are using a moving average of the value of sales. The default way we do it is per store but to make more meaningful conclusions in some situations, we might calculate a rolling average based on all stores. Moving average counts ten days before the day we study at that point. Furthermore, the average is computed using weights that are the sales values for each store to balance the metric. The weight is proportional to the sum amount of sales in one store for a specific product compared to the sales overall of all stores in one store. This procedure we call the promotional moving average, and we also have the value of sales for each day (observation).
4. At this point, we have in our dataset all required information. Each day of the month is flagged, whether it is promotional or not, and we also have the value of sales for each day (observation).
5. We create a list where each day of the month is the particular day of the promotion.
6. Now we use linear regression to check the impact of each-day of the month and we mean the day was significant or not of the value of sales. We will have one more feature: is_promotion, which only determines whether each day is a promotion or not.

The second approach is much simpler. We proceed as follows:
1. We aggregate the sales of the first ten days of each month.
2. Then we create a list of the first ten days of each month and we mean the day was significant or not of the value of sales.

The data preparation for our analysis consists of a few steps:
1. First, we mark our observations [days] as promotions with use of our promotional moving average.
2. Then we make multiple lists where each observation is marked as promotion.
3. The algorithm checks if the day (observation) was a promotion day or not. So, we add it to the list if it is. After that, we take the next first day of the promotion and creates a new list after finding it.
4. Then we take each day of the consecutive days to the promotion start from 1.
5. After that, we take each day of the consecutive days to the promotion start from 1.
6. Then we sort the list by the descending order.

We managed to answer our question stated in the introduction to the article positively. Using heuristic methods based on simple statistics and use of out of the box machine learning methods might bring valuable business insight and we might use this to improve the profitability of the company.

Identity of decision-making rules in the case of parallel application of promotion

We use our promotional rolling average to define when we have a promotion day (observation) or not. To calculate the base price, we choose prices from each segment (large-area, small-area). Doing that we can compare the price in each stores by using graphs. Actual price from each store as one graph and baseline price, which we calculated using all stores from the segment. We calculate when the price is below the baseline price, and we mark it as the promotional price.

We can picture that on the graph in which situation of the whole observed period described above occurs.

Suggesting how many days the promotion should cover to maximise the sales volume

1. We create a list of promotion campaigns and for each campaign we create a list of rolling average to flag the appropriate days and then add them to the list which we will later use for testing checking if the next day is also promotional.
2. The first day which is not promotional breaks the chain and the next promotional day starts a new promotion campaign.
3. We count how many promotion days we have each campaign and create a separate column for that.
4. Then we count a weighted average of sales per day for campaign-per-campaign.
5. Because we do not want to show promotions shorter than three days, we also reject promotions campaigns which are shorter than three days.
6. Then we count an average of days for the best and worst result for each store.

Suggesting what the highest prices should be in the promotion to make it useful

1. We use our promotional rolling average which is weighted by the number of sales the store has done in proportion to the all other stores during the observed period to define if each observation (observation) is promotion. We can call it average price or baseline.
2. Then we create a metric (1)

<table>
<thead>
<tr>
<th>Sales Value</th>
<th>Average price - promotional price</th>
</tr>
</thead>
</table>

| The promotional price is the price each day which we marked as a promotion and Sales Value is the sales which are the sum of sales of all units of product each day of the store. |
| What is important to mention: we want to maximise the whole metric. |

Based on this metric, we create a list ranked in descending order for each store.

The tools we use are:
1. Linear Regression Foundset Forrest, XGBoost, LSTM
2. All those tools were parameterized.
3. Data preparation for our analysis consists of a few steps:
4. Part of, based on the correlation and partial correlation we decide on adding lags to the data set. Based on this, we decide to choose our look-back period to be seven days, which means that sales correlate every week with the highest sales during the weekend.
5. Then we take each day of the consecutive days to the promotion start from 1.
6. We declare the MAE and RMSE scores to be calculated based on the predicted values.

Analysing the data using limited or small datasets is a challenging task as the knowledge which we get from these data needs to be completed by the subject matter expertise. This is, of course, the case in any data related matter. That might be an obvious truth, but as a data scientist, even the best machine learning or econometrics methods will not be sufficient to make a good analysis if we ignore the context of the data. The food product sales data always needs to be augmented with information about the seasonality of sales data on the small sample with some addition of knowledge of the fast-saying consumer goods might give as practical business intelligence and we might use this to improve the profit of the companies.

In this paper, we spotted the difference between the categories of products and the different behaviour of the customers regarding these products. The insights mentioned above were about the length of the promotions and best-selling day and the marketing of the price which is part of the question: "When, for how long and for how much?" each retailer planning promotion is asking.

We also concluded that beer sales are much more seasonal than coffee sales. The two primary data stamps drive sales of beer: weekends and the summer period from June to September. Also, coffee sales are much more dependent on the promotions. The tools used in this paper are showing how easily we can verify some sales hypothesis about the behaviours of the consumers.

All the fundamental questions about the promotion analytics can be answered with the use of the heuristic data mining techniques. As retailers collect more data each year, it becomes possible to make our algorithms more robust and accurate. It would also be suitable both for researchers and society if the large global retail chains would share some parts of the data which they collect. The wars for data between those companies cause increasing closure of access to open data. The possibility of using it for something else than for gaining an even more significant competitive advantage in an already very monopolized market decreased in recent years. As a last remark, we can say that, to overcome this situation, a small retailer should be using opportunities created for them by technology and collect their data to get left behind on the new digital frontier.