Recognizing campaigns that cause cannibalization (presenting the results)

Petra Huttunen
11.06.2018

Instructor: Tuomas Viitanen, TkT, Relex Solutions
Supervisor: Fabricio Oliveira
Cannibalization

A phenomenon where one product diverts sales from a substitute product, causing the sales of the substitute product to drop.

(Copulsky, 1976)
Research questions

• Which promotions cause cannibalization?
• Can we predict before the product is promoted, whether that promotion will cause cannibalization or not?
Findings from literature

• Products with similar attributes attract similar customers and can therefore divert sales from each other. (Mason and Milne, 1994)

• If one product is made more appealable to the customer because of a promotion, the customer can choose that product over another and cannibalization will occur. (Dawes, 2012)

• Promotions have different effects on product sales based on marketing, how good the offer is and how much the price is discounted. (González-Benito et al, 2010)
Method

• Dataset contained information about the campaign and cannibalizing product, cannibalization event and cannibalization relationship

• The goal was to fit a regression model to the data
  – Dependent variable: additional sales of cannibalization event

• Data was real customer data from 2013 to 2016
  – 2013-2015 was used as analysis period
  – Data from 2016 was predicted by using the model fitted to analysis data

• Product groups used in analysis were beef, chicken, coca cola and frozen potatoes
Regression models in general

- A regression model consists of three parts
  - Dependent variable $y$
  - Systematic part of the model $f(x; \beta)$
    - Function of the independent variable(s) $x$
  - Residuals $\varepsilon$ of the model
- The goal is to choose parameter $\beta$ so that the residuals are as small as possible
  - One option is to use ordinary least squares method

\[
    y = f(x; \beta) + \varepsilon
\]

\[
    \min_{\beta_0 \ldots \beta_k} \sum_{i=1}^{n} \varepsilon_i = \sum_{i=1}^{n} (y - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2
\]
Independent variables

- Product group
- Campaign types
  - Category, type, subtype
- Sales metrics
  - Sales quantity, baseline sales, additional sales
- Statistical increase of campaign (=relative sales increase)
- Campaign duration
- Price metrics
  - Price discount, price during campaign, price before campaign
- Information about the relationship
  - Regressor coefficient, correlation, p-value…
Correlations between the dependent and independent variables

- Statistical increase
- Campaign duration
- Price during campaign
- Regressor coefficient
- Sales quantity during campaign
- Price discount
- Correlation
- Number of observations
- p-value
Categorical variables

• Can we distinguish differences between campaigns in terms of cannibalization?
  – Campaigns seem to act in a very similar way
Categorical variables

• How do the product groups differ in terms of sales and prices?
  – Chicken seems to have biggest variation
Fitting the regression model

• Model was first complied with all of the variables
  – Variables were eliminated based on significance and multicollinearity
• Linear variables and non-linear variables
• Log-model
• Normalised vs. real data
## Fitted regression model

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.25</td>
<td>0.0034</td>
</tr>
<tr>
<td>Category: offshelf</td>
<td>0.065</td>
<td>0.13</td>
</tr>
<tr>
<td>Category: onshelf</td>
<td>0.20</td>
<td>3.13e-5</td>
</tr>
<tr>
<td>Statistical increase</td>
<td>0.02</td>
<td>1.80e-6</td>
</tr>
<tr>
<td>Campaign duration</td>
<td>-0.0058</td>
<td>0.0010</td>
</tr>
<tr>
<td>Correlation</td>
<td>1.1</td>
<td>9.72e-7</td>
</tr>
<tr>
<td>Number of observations</td>
<td>-0.00029</td>
<td>2.44e-9</td>
</tr>
<tr>
<td>P-value</td>
<td>-3.4</td>
<td>0.036</td>
</tr>
</tbody>
</table>
Goodness of the model

- Adjusted R-squared 15.63%
- All VIF-values were under 2
- No significant outliers (based on Cook’s distance)
Prediction of the model

- The model was used to predict the normalised additional sales of training data.
- The results were compared with training data.
- Dataset contained a lot of noise.
Issues with dataset

• Data contained noise and exceptions
• Exceptions were mostly caused by
  – Holiday events
  – Overlapping campaigns
  – Accuracy of baseline sales
    • Back-to-back campaigns
• Not all exceptions could be cleaned from the data
Conclusions

• The regression model had a low coefficient of determination
  – The model build based on analysis data didn’t explain the training data
• Confirmation that certain attributes do have an affect on cannibalization
  – Campaign categories
  – Strength of the cannibalization relationship
  – Statistical increase and campaign duration were interesting
• Other significant variables in various models
  – Price during campaign
  – Sales quantity
  – Subgroups
Future prospects

• Missing some possibly important information
  – Marketing plans for campaigns
  – Brand and quality of products

• Cannibalization was recognized only within product groups
References


