Developing a Competitor Index for Market Share Prediction in Retail

Isha Bhalla, Udita Chakraborty, Rohit Kaul, Amratansh Sharma, Matthew A. Lanham
Purdue University Krannert School of Management
ibhalla@purdue.edu; uchakra@purdue.edu; rkaul@purdue.edu; sharm321@purdue.edu; lanhamm@purdue.edu

Abstract

In the retail industry, the nature of competition faced by any company exerts a strong influence on a firm’s market share. The aim of this research is to create a competitive metric as an input when building a predictive model for retail store assortment and location decisions. To this end, we engineered a new feature that accounts for the market class and competitor count and also included an element of time to account for the impact on sales before and after introduction of competitors. In collaboration with a national retailer, we developed a robust index to predict market share that uses Linear regression and Random Forests.

Introduction

Across any business or industry, when opening a new store, we try to create forecasts for sales and inventory based on similar stores that have already been established. When new stores are opened, many inputs are needed to create the original assortment and set reasonable stocking levels. Our industry partner, currently uses similar existing stores as a preliminary guide or benchmark. Identifying competitor strengths and understanding what drives customers to competitor stores can help companies navigate the treacherous competitive landscape by making sound strategic decisions. The motivation for our project is that different characteristics of the competitors’ stores exert varying influences on the client’s share of the market, and it is important to differentiate the significant characteristics from the non-significant ones. To this end, our project attempts to identify the strongest competitor characteristics and model those to develop a competitor index, that can in turn be used in conjunction with sales and inventory level information to predict market shares of individual stores.

The primary research questions that we aim to address are:
1) What impact do competitor stores have on the market share of our client’s stores?
2) Can we identify attributes of competitor stores that exert the most influence on the client’s market share?
3) Can we develop and validate a competitor index factoring all significant attributes, to aid client in planning feasible new store locations.

Literature Review

Competitor Index

Benchmarking Performance in Retail Chains: An Integrated Approach

Use any dissimetric determinants of store demand and various other factors such as ambience, trading area etc. Maximum likelihood estimation is used to estimate parameters.

Forecasting Retail Location Strategy in a Changing Environment

Explains the factors that guide retail stores to select the locations. The resulting model gives the best site for locating a new store and the corresponding predicted profit as well.

Competitor Market Share

Using A Flexible Logit-Logistic Model

This research compares three models i.e. standard high/lower power transformation, and log version of it to generate forecast of market share of U.S. farm-based cattle till.

Market Share Prediction A New Model with training based competitor analysis

Competitor analysis is used to support product development, pricing, store planning and market share prediction. Author has also proposed RFC-BOLSE, Reconciled First Choice and ordinary least squares estimates.

Methodology

We identified competitor parameters that have the maximum influence on market share through a regression model and generated a competitor metric that is used as a predictor in model training. Competitor score accounts for competitors, and distance from prime competitors. Finally, a random forest is run to predict the market shares.

Results

After incorporating the cluster variables, the performance of the two models are depicted in Figure 5.

Figure 2. Study Design

Store details, market share and competitor information are captured separately in different formats and at different granularities. We processed each dataset separately and rolled them up to a store level. We created features to indicate different competitor parameters and joined those together.

Figure 3. Cluster Information

Methodology

Figure 4. Model Evaluation

The random forest produces the lower test results – R-squared of 43.2% vs. an R-squared of 66.1% for the training set. The linear regression model on the other hand gives training set accuracy of 2.4% and test set accuracy of 0.7%. Thus, the random forest model was our chosen model owing to both lower test set RMSE and higher test R-squared.

Conclusions

By creating the new competitor index, we were able to build a more robust model which would enable us to better predict the market shares of new stores. Any new store addition in a pre-determined location would fall into one of the four clusters that we have identified. Our metric can also be used by the location planning team to determine the optimal store locations with good accuracy.

Limitations and Next Steps

We are exploring more models as the random forest model is a bit overfit (large train/test set R-squared difference). As next steps, we believe if we gather certain data on the stores, it can help us better assess the competitor influence. These factors can be value added services, size of competitor store vs client store, length of trading in each market etc. Also, we have averaged the time series data for each store. We can explore converting the data into a stationary time series and eliminate autocorrelations between market shares of same stores.

Acknowledgements

We would like to thank Professor Matthew Lanham for constant guidance on this project, as well as our industry partner for allowing us to help collaborate on a project that was very important to them.

Table 1: Summary of literature