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 competition metric as an input when building a predictive model for retail store assortment and locational 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 develop and validate a competitor index factoring all significant attributes, to aid client in planning plausible new store locations.

Literature Review

Name of the Paper	Initiative and Findings	Insights Obtained
Benchmarking Performance in Retail Chains: An Integrated Approach	Uses exogenous determinants of store demands and various other factors such as ambience, trading area etc. Maximum likelihood estimation is used to estimate parameters.	Resulting model improved the accuracy and R squared to 0.77 as opposed to prior model.
Formulating Retail Location Strategy in a Changing Environment	Explores the factors that guides retail stores to select the locations. The approach was to create a multiplicative competitive interaction model.	The resulting model gives the best site for locating a new store and the corresponding predicted profit as well.
Forecasting Market Share Using A Flexible Logistic Model	This research compares three models i.e. standard logit, inverse power transformation, and log version of it to generate forecast of market share of US farm-based catfish fillet.	Flexible logit models do have various advantage over standard logits. Log- IPT shows better performance over the other two models in predicting the market share trend.
Market Share Prediction: A New Model with rating-based conjoint analysis	Conjoint analysis is used to support product development, pricing and positioning and market share predictions. Author has also proposed RFC-BOLSE, Randomized First Choice and ordinary least square estimates.	RFC-BOLSE generates share of preference that incorporate reliability of various OLS based estimation and similarity of competing profiles. However, this model is basically confined to share of preference estimation and cannot be directly applied to estimate market share.

Table 1: Summary of literature

Methodology

Figure 2 outlines our study design, starting from data collection, cleaning, pre-processing, feature creation, model selection, and cross-validation design.

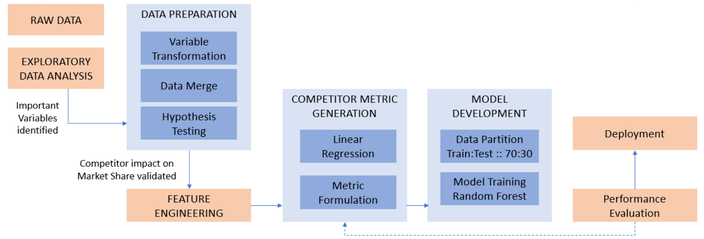


Figure 2. Study Design

Data

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.

Data Pre-Processing and Model Design

We created dummy variables and dropped correlated predictors to obtain stable parameter estimates. We identified feature vectors that could be expressed as linear combinations of other vectors and removed them. We normalized the quantitative variables to remove bias introduced by extreme values. Then, we partitioned the pre-processed data into training and test sets to evaluate model performance. The important drivers we identified included:

- *Competitor Count, Co-ordinates, Presence of Competitor "A" within 5 miles, Presence of Competitor "B" within 5 miles, Market Type, and Cluster.*

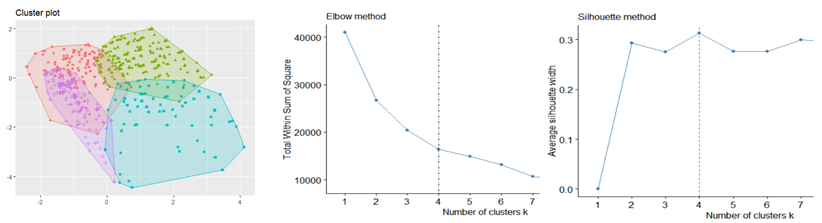


Figure 3. Cluster Information

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 to 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

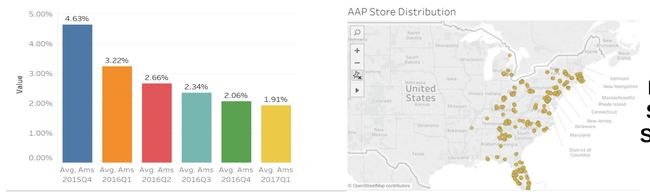


Figure 3. Market Share Trend and Store Distribution in the US

Results

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

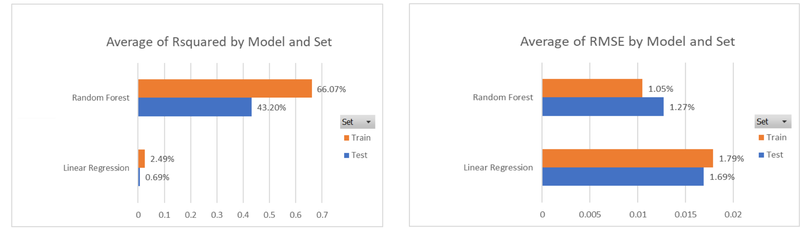
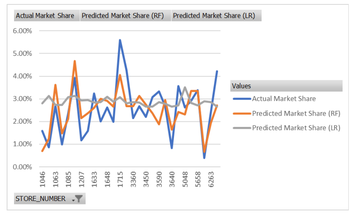


Figure 5. Model Evaluation

The Random forest produces the better 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.



Feature	Score
LATITUDE	100
LONGITUDE	82.683
COMPETITOR_COUNT	53.291
MARKET_TYPE3	28.948
MARKET_TYPE6	17.184
MARKET_TYPE5	14.608
MARKET_TYPE4	12.89
AZO_WITHIN_50	6.976
ORL_WITHIN_51	5.542
MARKET_TYPE2	0

Figure 6. Actual vs Predicted market Shares, Feature Importance Scores (RF)

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 asses 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.