THE BENEFITS OF BUDGET ALLOCATION

WITH AI-DRIVEN MARKETING MIX MODELS

THE LATEST IN AI TECHNOLOGIES REINVENT MEDIA AND MARKETING ANALYTICS @ ALLERGAN

H2O.ai

The Benefits of Budget Allocation with AI-driven Marketing Mix Models

The Latest in AI Technologies Reinvent Media and Marketing Analytics @ Allergan

Authors:
Akhil Sood, Associate Director @ Marketing Sciences, Allergan
Dr. Michael Proksch, Senior Director @ H2o.ai
Vijay Raghavan, Associate Vice President @ Marketing Sciences, Allergan

Introduction

The call for accountability in marketing has been growing over recent years to justify total investments in the area. Furthermore, to **allocate budgets across different marketing channels, managers have been trying to better understand their contribution to sales**. To model the impact of marketing on sales, time-series approaches have been proven to be successful. However, Allergan's experience and academic research shows that **marketing attribution models are challenging** as the complex relationships between variables need to be identified and model structures need to be defined. This is largely due to detrending for a definition of lag structures by Granger causality analysis, interaction terms for variable interactions, and possible polynomials for non-linear relationships and definition of endogenous variables. In our industry, we usually define GLM models with limited interaction effects between variables, to deny non-linear relationships via polynomials and to keep interpretability high, by excluding highly correlated endogenous variables, and apply ad-stock effects to the marketing mix drivers (product, price, place, promotion) or include a lag structure to them. However, this is a **very cumbersome, manual process, which takes weeks or even months** to accurately define even just one marketing attribution model for one brand~~,~~ given that companies usually have a broad product/brand portfolio or different customer segments to analyze. But there is good news! New AI algorithms and techniques cover those weaknesses, such as **Gradient Boosted Machines (GBMs) and Shapley values that make marketing mix modelling approaches more accurate and scalable** for a higher number of stores, brands, products, or customer segments.

Allergan Case – the promising journey of AI-driven marketing analytics

The Allergan team traditionally utilized GLM models for estimating the impact of promotion campaigns and to optimize the marketing mix of dozens of products across a variety of marketing channels. We started our AI-marketing journey two years ago together with H2O.ai to improve our marketing analytics. One of our recent challenges was to apply new machine learning algorithms to estimate the impact of a radio campaign for one of our products. For this project, both GLMs and H2O.ai's Driverless AI were utilized and compared. In addition to the radio campaign, we had multiple promotion campaigns on personal and digital channels that would interact in driving sales and we assumed non-linear response curves. We had data on customer characteristics like demographics, physician specialty, market potential, which were included as control variables. We were very excited to see how new machine learning algorithms perform as we all know how much effort it is to define traditional linear marketing mix models and that they had a lot of problems with complex real-world cases.

Traditional Linear Marketing Mix Models vs. new Algorithms

General Linear Models (GLMs) have been around since 1972 and have been used extensively in the area of marketing mix modelling as they allow for an additive approach (sales = baseline + TV spend + Radio spend + other channels). This, in combination with clear model specifications (e.g. VAR models), provides analysts with the opportunity to identify incremental contributions of marketing channels to sales. However, **new algorithms outperform linear models** not just in accuracy but also in explainability by accounting for non-linearity between a predictor and dependent variable and interactions between predictor variables. Tree-based models in particular score high on all those characteristics mentioned above, and among them Gradient Boosted Machines (GBMs). **Gradient boosting** is a machine learning technique, typically based on decision trees, that uses an ensemble of weak prediction models. Those models usually show a higher accuracy in out of sample predictions than GLM and are more robust than single decision trees. This difference can lead to a significant business impact compared to linear Marketing Mix Models.

The attribution of sales to marketing channels in GBMs can be done through a methodology called Shapley values. **Shapley values are named after Lloyd Shapley (1923 - 2016)**, a mathematician and economist and code breaker in WWII. Shapley introduced the theory in 1953 and won the **Nobel Prize in Economics for Attribution** in 2012. From the 1960s onward, Lloyd Shapley used what is known as Cooperative Game Theory to study different matching methods.

In a **cooperative game** in which the **payoff must be attributed to players** who have made unequal and interacting contributions, the **Shapley value determines the fairest distribution of payoffs**. For example, the Shapley value can be used to determine what each member of a group should pay in a restaurant when everyone shares their food. The theory of Shapley values are based on four axioms, which makes this method unique compared to other techniques:

- **Efficiency**: the feature contributions add up to the difference of prediction with the features and the average (expected value). This is highly important if you see it the other way around, it means that the average prediction is fairly distributed across all the features.
- **Symmetry**: the contributions of two feature values A and B should be the same if they contribute equally to the prediction.
- **Dummy:** a feature with Shapley value of 0 does not change the predicted value.
- **Additivity:** guarantees that for a feature A used in a model M that average the prediction of two other models M1 and M2, you can calculate the Shapley value for each model individually M1 and M2, average them, and get the Shapley value of A for model M.

Those axioms as requirements for model explanations sound logical but are currently only fulfilled by Shapley values. Research has also shown that Shapley value, compared to other explanations, **has much stronger overlap with human explanations** (Lundberg & Lee, 2017).

Allergan Case – the holy grail of marketing attribution analytics
The concept of Shapley values combined with new machine learning models is intriguing - it promises the holy grail of marketing analytics. It combines highly accurate out-of-the box machine learning models with high explainability leading to accurate marketing mix attribution models. Historically, we used linear models and knew their weaknesses from an accuracy perspective, but they are easily explainable and therefore provided a great value in marketing attribution analytics. Hence, we usually accepted all of their weaknesses in accuracy and adjusted the models in multiple iterations to ensure that model outputs align with our business understanding.

Overcoming the weaknesses of Linear Marketing Mix Models

Traditional linear time-series models have known weaknesses we are attempting to overcome with GBMs and Shapley values. First, the short-term impact of media on sales over several weeks can be low and is therefore very hard to estimate with linear models if the error of the model is higher than their impact. Further challenges include the problem of group aggregations and the time to calibrate a model correctly, non-linearity in market response functions, variable interactions, and asymmetries of market responses over time.

The Aggregation Problem

Linear Marketing Mix models need to be specified for each driver and interaction between variables. Besides marketing mix variables, control variables need to be identified (e.g. a trend variable, seasonal effects) and usually preliminary Granger-causality tests are used to limit the number of variables to be included in a final model and to define a lag structure accounting for the lag of drivers and possible ad-stock effects. However, **as the number of variables and groups (e.g. brands, stores, products) increases, the number of tests required to cover all possible combinations rapidly becomes excessive**. The same problem affects variable interactions (e.g. TV and online). Hence, it has become practice to prune the available data to consider only the top brands in a market, decompose price promotions, reduce product categories or apply the 80/20 rule to capture only relevant areas. However, that has a number of undesirable consequences, such as empirical generalizations (favoring the more successful brands or bigger stores), reduction of the comparability of findings, and the production of biased parameter estimates. This can all be overcome with GBMs. Parameterization of GBMs with an out-of-box sample is taking care of model calibration, the algorithm itself for variable selection and possible monotonic constraints can account for overfitting. As this is an automatic process, no models or lag structures need to be specified beforehand and can be applied to a big number of brands, stores or customer segments.

Non-Linearity of Market Response functions
Either linear models miss the ability to model non-linear relationships between predictor and dependent variable or polynomial functions have to be defined for each predictor variable, which can easily lead to an overfit of the model and non-logical variable response functions (e.g. that a higher marketing expenditure would reduce sales after a specific spending amount). But most real-world relationships and especially the ones in marketing are non-linear (e.g. $1000 investment in TV will have a different ROI compared to $1M). Figure 1 shows some examples of media response functions for different marketing channels. Each of them has specific inflection points or ranges that have been shown to be more efficient for media investment (here per GRP). But if non-linear relationships cannot be represented by a marketing mix model, **the definition of an inflection point to identify an optimal spending amount is not possible.** GBMs account for non-linearity and allow for more accurate market response functions. Monotonicity constraints can further improve estimations in case of estimating the impact of rare events.

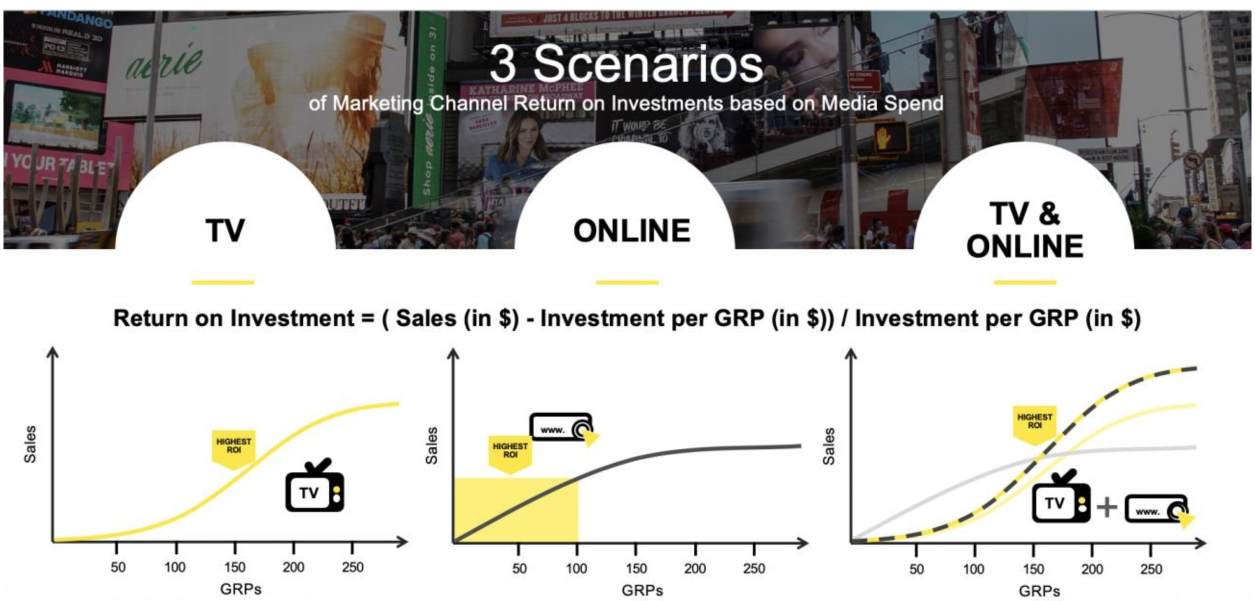

Figure 1: Exemplary Media Response Curve Scenarios

Variable Interactions
Although variable interactions (e.g. between TV and online invest) can be specified in linear models, they are very hard to identify in some cases as their impact can be small (e.g. seasonality and online invest), can lead to an overfitting of a model or the overall error is

overshadowing their effect. Figure 1 (at the right) is showing the combined impact of TV and online investment, which is a common phenomenon. Especially because their individual impact is non-linear, it will be hard for a linear model to pick-up any interaction impact. Furthermore, **it is very difficult to identify interactions between more than two variables.** However, there are a variety of cases where higher level interactions can play an important role (e.g. for media impact across different regions and brands). Estimating interaction effects with GBM is especially important for evaluating optimal budget allocations across media channels, to include the impact of seasonal or context effects (such as pricing or competitor actions).

Asymmetries of Market Responses over time

Market response models often assume a symmetric response pattern, which means that the sales impact of marketing spending over time increases or decreases the same in absolute value. However, **previous research has shown that the spending amount can influence its longer-term effects.** That means that the ad-stock could be longer for higher media spend for a channel such as TV, which is not covered by a geometric ad-stock decay function or a simple lag structure in a linear model but a tree-based approach would pick that up automatically. A behavioral explanation for such asymmetry could be that a higher spending is perceived as a signal of strong brand support. GBMs automatically pick up those time-oriented interactions of lag-variables and can provide further insights on ad-stock and carry-over effects.

Allergan Case – the comparison between both approaches

We wanted to compare the traditional linear marketing mix with the new machine learning based approach to experience the differences firsthand.

We started off with our traditional linear marketing mix approach. Over several weeks we developed a fixed effect linear model in SAS and ran multiple iterations to ensure that model output aligns with our business understanding e.g. coefficients of promotions are expected to be positive and specific consumer control variables are expected to be significant. During that process we explored and created multiple interactions between channels and promotions but had to keep the model manageable and were not able to include all interactions we created. Additionally, we tested different ad-stock values for each promotion channel until a final set of ad-stock values provided the best model fit. We put an extra effort in testing multiple functional forms for each promotion channel to account for non-linearity to measure for example saturation effects. We had to make a series of assumptions during that process.

Pretty confident about our non-linear marketing mix GLM, we were keen to see how the new approach would perform. We applied H2O.ai's Driverless AI to run a series of tree-based models (including GBMs) and extracted Shapley values via its Python client to calculate the contribution for each channel. We were able to do that within hours, not weeks. We were able to quickly get response curves for each channel based on Shapley values and did not have to test individual variables. This is especially helpful for complex marketing mix cases, where some functional form is typically "assumed" for some of the channels and models are strongly affected by variable interactions. Furthermore, no ad-stock testing was required – lagged promotion variables accounted for those effects. Additionally, we obtained promotion effects on individual customer level.

The time-dependence of Market Responses

Once an impact of a media channel has been identified via a linear model, it is usually assumed that this holds for the entire time sample, unless there are clearly identifiable structural breaks. However, even when there are no clear structural breaks, we know that market responsiveness may not be the same over time. For example, **advertising effectiveness might decline over the life cycle of a product, seasonal differences of a media channel's impact, the sales shock of an established brand after the introduction of a competing product or the own-price effects on media effectiveness**. As it is harder to identify and define interaction effects in linear models, the impact of extraneous/context variables on the marketing mix and the interaction between marketing mix variables are usually not accounted for. However, GBM models pick up those interactions. This is an important feature that can be used for temporal budget allocation and scenario simulations to improve the average ROI for a media channel (see Figure 2 below).

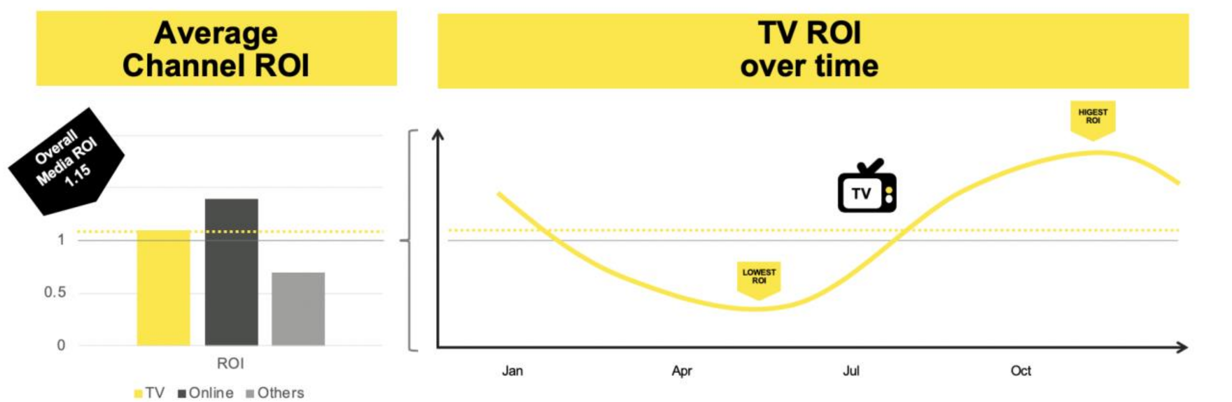

Figure 2: Media Channel ROI over time and Average Media Channel ROI

Optimal Budget Allocation

The final goal of marketing mix modelling is to allocate budgets across channels, products, customer segments and brands. A common approach to allocate marketing budgets among marketing practitioners has usually been upon the combination of insights from more or less static Linear Marketing Models and market statistics of media spending's of competitors to increase the share of voice for a company's media campaigns. Unfortunately, this will not lead to the expected long-term effects and will only show a short-term impact on sales. However, there might be other strategies that lead to higher ROIs over time. The three examples in Figure 3 illustrate three scenarios, a budget allocation based on market spending behavior, an optimized but static budget allocation scenario and an optimized temporal budget allocation. The temporal optimization has the highest impact on ROI and will allow the more optimal allocation of marketing budgets for specific time windows and real-world scenarios. This approach can also give you information for optimal budget allocations across brands, regions and customer segments. **A final simulation of input drivers will identify scenarios for your investments based on marketing mix interactions, market movements and other context effects.**

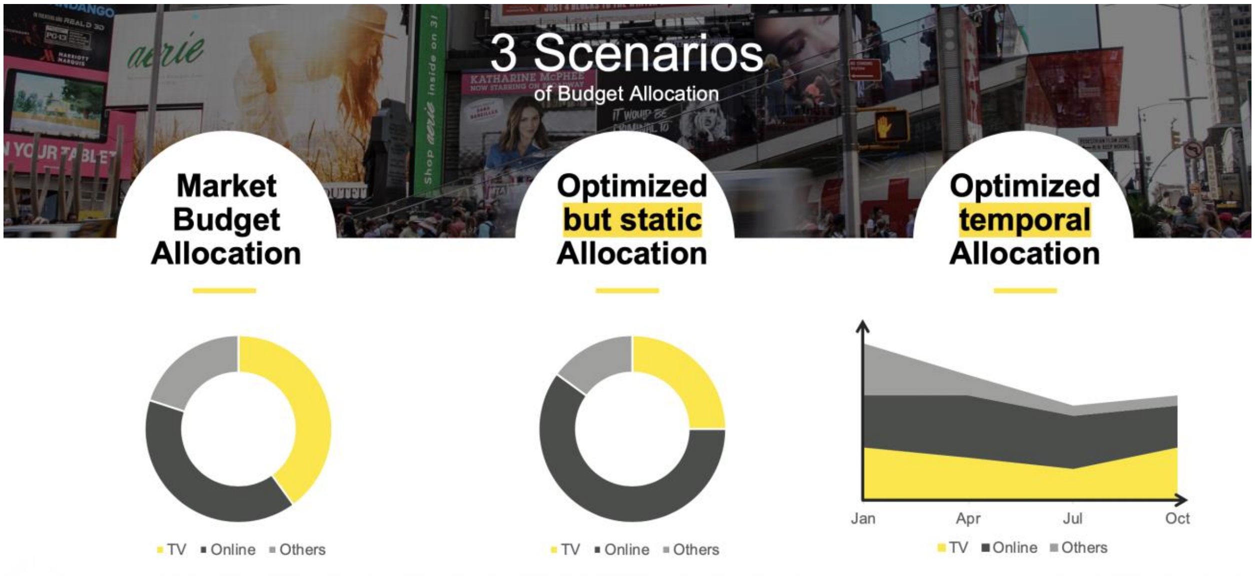

Figure 3: Three Media Invest Scenarios

Allergan Case – optimizing media budgets across customer segments
Compared to the traditional linear approach where we only got one single coefficient at national level for each promotion channel, we obtained the contribution of each channel to sales at individual customer level. This allowed us to segment customers on channel affinity, which proved to be very helpful to increase business impact.

Summary

Budget Allocation has always been a complicated topic in the field of Marketing as it is a **complex optimization problem** taking into account not just a variety of marketing effects (product, price, place, promotion), context effects (such as trend, seasonality), market actions (competitor actions, new product launches, competitive price changes) and groups that might show different market response patterns (different brands, regions, customer segments). **We hope that we could make a good case for the move to AI/ML by implementing GBMs and Shapley values in your marketing analytics to provide the unique opportunity to build more holistic marketing mix models.** Your final models provide a production ready environment for a marketing application that can help marketers better understand budget allocation and help them to improve the impact of marketing spend.

Allergan Case – closing on a successful proof of concept
Marketing mix/attribution analytics has always been a highly complicated topic and brought value only with a high effort in data preparation, testing of single variable impacts and variable interactions, functional forms for important response curves, defining ad-stock effects and applying business knowledge to get to logical results. With all that effort,

we could get to comparable results the new Machine Learning approach was providing in a shorter amount of time. Furthermore, we had to make several assumptions in our linear approach, which were not necessary with the new approach and we obtained unique and highly valuable outputs on customer level, which allowed us to optimize budgets across customer segments. We expect to see more benefits running more complex marketing mix simulations including a more marketing/media channels and estimating their impact for variety of real-world scenarios. We closed that proof of concept with a very positive result and with a high motivation to leverage that new concept to a variety of other marketing, sales and business development challenges.

This paper contains the views of the authors and is not the official position of Allergan or H2O.ai.
© 2020 H20.ai All rights reserved.
About H2O.ai
H2O.ai is an open source leader in AI with a mission is to democratize AI for everyone. H2O.ai is transforming the use of AI with software with its category-creating visionary open source machine learning platform, H2O. More than 18,000 companies use open-source H2O in mission-critical use cases for Finance, Insurance, Healthcare, Retail, Telco, Sales and Marketing. H2O Driverless AI uses AI to do AI in order to provide an easier, faster and cheaper means of implementing data science. H2O.ai partners with leading technology companies such as NVIDIA, IBM, Intel, AWS, Microsoft Azure and Google Cloud Platform and is proud of its growing customer base which includes Capital One, Progressive Insurance, Comcast, Walgreens and MarketAxess. For more information about how H2O.ai is driving an AI transformation, visit www.h2o.ai