A non-parametric approach to the Multi-channel attribution problem

Meghanath Macha Yadagiri[1], Shiv Kumar Saini[2], and Ritwik Sinha[3]

Adobe Research, India
{mmacha[1],shsaini[2],risinha[3]}@adobe.com

Abstract. Multi-channel marketing attribution modeling is a two-stage process. First, the value of exposure from different marketing channel needs to be estimated. Next, the total surplus achieved needs to be assigned to individual marketing channels by using the exposure effects from the first stage. There has been limited work in exploring possible choices and effects of determining the value of exposure to different marketing channels in the first stage. This paper proposes novel non-parametric and semi-parametric approaches to estimate the value function and compares it with other natural choices. We build a simulation engine that captures important behavioral phenomenon known to affect a customer's purchase decision; and compare the performance of five attribution approaches in their ability to closely approximate the known ground truth . Our proposed method works well when marketing channels have high levels of synergy. We apply the proposed approaches on two real-world datasets and present the results.

Key words: Multi-channel attribution, Shapley Value, Simulation study

1 Introduction

Digital marketing has given a marketer greater access to customers as well as ability to observe each customer's interaction with different marketing channels. Access to the digital marketing data has allowed a marketer to assign each conversion event to various marketing channels. This is called the multi-channel marketing attribution (attribution) problem. The attribution modelling is a two stage process. In the first stage, the value of exposure from different marketing channel needs to be estimated. In the second stage, the total surplus achieved needs to be assigned to individual marketing channels by using the exposure effects from the first stage. There has been little work in trying to understand the possible choices and effects of determining the value of exposure to different marketing channels in the first stage. Similarly, there is a lack of literature on how to distribute the conversions that would have occurred without exposure to any marketing channel. The goal of this article is to fill these gaps in the literature on attribution modelling.

The data available for estimating an attribution model consists of a series of marketing touches that may or may not result in a conversion event. An attribution model first estimates the effect of exposure (exposure effect) to the marketing channel(s) on a conversion event using observational data. The exposure effect is captured by modelling the effect of a subset of the marketing channels on the likelihood of the conversion event. The exposure effect can be highly non-linear due to, among other reasons, the synergies between the marketing channel. The non-linearities in the exposure effect explored in [5],[7],[9] are captured by using the non-linear parametric models such as a logistic regression model. However, there is no justification for using a particular model. The answer to the attribution problem is highly dependent on the exposure effect model. We evaluate the parametric models used to estimate the exposure effect in the past using a simulation engine. The popular parametric models do not perform well, especially, when the level of synergies between the marketing channels are high. We propose a novel non-parametric approach to estimate the exposure effect. The proposed model performs well even when the synergies between the marketing channels are high.

The exposure effect model can be used in the second stage to divide the conversion events to each marketing channel. We use the Shapley Value approach to assign the conversion events that occurred due to the observed marketing touches. The Shapely Value approach used here is similar to the approach in [5]. This is called marginal attribution because only the increase in the conversions due to marketing touches are assigned. At times, the goal of attribution modelling is to assign all of the conversion events. The past literature on attribution modelling do not provide any answer other than the rule based approaches such as the first touch and last touch attribution models. We propose a Nash Bargaining [3] based approach to assign the left-over return to each marketing channel. The Nash Bargaining solution is a cooperative game theoretic concept to divide surplus in a bargaining games where players have outside options. We use the marginal attribution to each channel as the outside option of the channel. We apply the proposed models using two real-world data sets. The models generate consistent results across different real world datasets.

The rest of the paper is organized in the following manner. In Section 2, we present the related work in the field of Shapley Value and Algorithmic Attribution. In Section 3, we formally define the marketing attribution. Section 4 presents our proposed methodology and its advantages. In Section 5, we present our simulation study and compare the proposed models with the existing parametric models. In Section 6, we present our results on two datasets. We conclude the paper and discuss directions for future work in Section 7.

2 Related Work

As noted earlier, there is no past work in evaluating the models that estimate the exposure effect. In this section, we describe the work that touches upon the other aspects of the attribution model.

Algorithmic Attribution There have been several efforts to solve the problem of marketing attribution in an algorithmic fashion. One of the first models was proposed by [7]. In this paper, the authors propose two data-driven attribution models depending on estimated measure of channel relevance. The models lack a clear interpretation of fairness while assigning credit to the marketing channels. The authors in [1] model the customer journey as a funneling process leveraging the concepts of Hidden Markov Model to relate the stages of a customer to their conversion behavior. Different advertising campaigns may have varied notions of customer stages making such a model difficult to deploy in real-world scenarios. Similarly, [11] models the customer's journey as a Markov process and provides methods to estimate each channel's effective contribution and prediction of conversion rate. In [9], the authors present a model based on counterfactual analysis to solve the attribution problem and estimates the incremental effect of each marketing channel.

The only game-theoretic approach for marginal marketing attribution was proposed in [5]. In this paper, the authors initially recommend properties of a good attribution model. Further, the authors frame the problem of attribution as a causal estimation problem and then propose two approximate methods based on co-operative game theory. However, the proposed models are parametric and usually cannot be simplified. Further, there is no comparison of the two proposed models to evaluate which model performs better. Also, the paper does not highlight the difficulties of deploying a game-theoretic framework to real world settings. Additionally, the paper is restricted to marginal marketing attribution - assigning the surplus produced by customers to the marketing channels. The methods and evaluations presented in this paper are motivated by the limitations of the current literature.

3 Problem Definition

Problem Setting Let $\Upsilon = \{U_1, U_2, ..., U_n\}$ be n users targeted by the marketer using k marketing channels $\Omega = \{C_1, C_2, ..., C_k\}$ [1]. We define *return* from a user as a measure of response to marketing activity undertaken by the marketer. Some examples of return may include purchase, revenue, page views etc. The *left-over* for a user is defined as the *return* that would have occurred without exposure to any of the k marketing channels. Also, *surplus* for a user is defined as the total *return* minus the *left-over*. Hence, $return = surplus + left\text{-}over$. Given that the the marketer has generated a return R, surplus S and left-over L, we would like to algorithmically assign $\mathbf{\Psi} = (a_1, a_2, ..., a_k)$,to the k marketing channels, where $\sum_j a_j = S$ and $\mathbf{\Pi} = (l_1, l_2, ..., l_k)$ where $\sum_j l_j = L$ to each of the channels that the marketer has used. For this purpose, we model the problem at a customer level by considering various marketing channels through which each customer in Υ has been targeted.

[1] We use i as a counter for elements (users) in Υ as well as elements (users) in Υ depending on the context, j is also used in a similar fashion for the set Ω.

Customer level Definition Let $\mathbf{E_i} = (e_{i1}, e_{i2}, ..., e_{ik})$ be the binary vector of a user i who has been exposed to the k available marketing channels. Here, $e_{ij} = 1$ if and only if the user i has been exposed to a channel j and is 0 otherwise. Let $\tau = (b_1, b_2, ..., b_n)$, where $\sum_i b_i = R$ be the total $return$ generated by the marketer on all the users targeted by the k available marketing channels. For all the users in Υ, we would like to come up with $\rho = (s_1, s_2, ..., s_n)$ the surplus produced by each customer. We achieve this by computing s_i^j, the surplus produced by marketing to customer i due to channel j in the matrix $S_{att} = [\mathrm{s}]_{ij}$. In S_{att}, the row-sum is the total surplus produced by each customer i , $\sum_j s_i^j = s_i, s_i \in \rho$. The column-sum is the attributed surplus to a marketing channel j , $\sum_i s_i^j = a_j, a_j \in \mathbf{\Psi}$ and $\sum_j \sum_i s_i^j = S$ where S is the total surplus achieved by all marketing efforts. Further we use the customer surplus estimations to calculate $(l_1, l_2, ..., l_k)$, the $left - over$ vector $\mathbf{\Pi}$

4 Approach

To solve the problem defined in the above section, we first estimate the surplus of each customer. Next, we model the multi-touch attribution problem as a coalitional game [2] at a customer level and then apply Shapley Value to determine the surplus that needs to be assigned to each channel. In the general marketing attribution problem where multiple channels are involved, some of the channels may influence a user more than others, as a result, may possess different bargaining power, hence the concept of $fairness$ of distribution of the final gains is an essential property which Shapley Value captures. Also, the measure of channel importance in an attribution model is the channel's expected marginal impact on conversion, where the expectation is taken over the possible orderings of the channels in Ω which is well addressed by Shapley Value. Finally, we propose a method to assign the *left-over* return to the marketing channels.

Customer-level Approach: We model the attribution problem by defining coalitional games (f, Ω_i) at a customer level. Where Ω_i are the various channels a customer i has been exposed to, $\Omega_i \subseteq \Omega$, $|\Omega_i| = k_i$ and the value of the characteristic function f is defined to be an estimate of the surplus produced from the customer. Further, we use the concept of Shapley Value to distribute the gains to the all the channels in Ω_i that were collaboratively involved in influencing the customer to generate return. However, Shapley Value is not readily applied to a problem, one needs to define the value function that satisfy certain criterion [3] [8]. Also, the theory assumes that it is possible to assess the expected gain for every possible co-operation, that is, for all 2^{k_i} possible combinations. Hence, one would need to estimate the expected surplus of a user when he has been

[2] A coalitional game denoted by (f,N) is defined by a characteristic function f and total number of players in the game N, where f maps subsets of players to real numbers: $f : P(N) \rightarrow R$ with $f(\emptyset) = 0$, where $\emptyset$ denotes the empty set and $P(N)$ is the power set of the N players

[3] One key criterion is that the characteristic function f should satisfy $f(\emptyset) = 0$.

exposed to all the possible combinations of channels in Ω_i. A marketer may not necessarily have information of the behavior of the user when exposed to various combinations of marketing channels which makes the direct application of Shapley value to real-world highly impractical. To address the issues that stem from the definition of Shapley value, we initially provide non-parametric and semi-parametric approaches to estimate the expected surplus from a customer.

Existing Methods In [5], the authors propose to estimate the expected surplus from a customer using a fully parametric approach. The models proposed in the paper leverage popular binary classification models like 1) Logistic Regression 2) Elastic-Net Regularised Logistic Regression. Another parametric binary classification model one could use to estimate the surplus is Random Forest [4]. That brings to our first contribution in this paper, we propose using non-parametric and semi-parametric models to estimate the surplus for each customer.

Initial setting Firstly, for a given data set, if the total number of marketing channels is k, a maximum of 2^k possible combinations (excluding repetitions) of marketing channels could be used by the marketer to target the users. Hence, a maximum of 2^k channel combinations could be observed in the dataset. For each such combination of marketing channels s, we define $f(s)$ as

$$f(s) = \frac{Purchases(s)}{Purchases(s) + Non - purchases(s)} \tag{1}$$

This function is the frequentist estimate of the conditional probability of purchase given exposure to the channels in s. Typically, in a dataset, all the 2^k channel combinations are not observed. Also, $f(\emptyset)$, where $\emptyset$ is the null set may not be calculated by the above formulation since the marketer may not have information about the users who were not targeted through any of the marketing channels and have converted. Hence, we propose the below methods to estimate the $f(s)$ of unobserved channel combination including $f(\emptyset)$. The proposed estimates are further used to form a coalitional game at a customer level.

Notation	Existing estimation	Notation	Proposed estimation
E1	Logistic regression	P1	(1) , (2) and (3)
E2	Elastic net regularised logistic regression	P2	(1) and Logistic regression
E3	Random forests	P3	(1) and Random forest.

Table 1: Notation for Existing(E) and Proposed(P) methods

Non-parametric Model: Let P_Ω be the power set of k marketing channels in Ω and by using the initial setting, say, we have information about a subset of the all possible combinations. Let this set be $S_{obs} \subseteq P_\Omega$. We approximate the value of $f(s), \forall s \in P_\Omega \setminus S_{obs}$ in the following manner

a) For each $f(s)$ to be estimated, $s \in P_\Omega \setminus S_{obs}$, consider all the subsets of the

combination of channels s (P_s, the power set of combination s) that belong to S_{obs}. Let this set be T_{obs}. Hence $T_{obs} = P_s \cap S_{obs}$
b) The value of $f(s)$ is given by averaging over all the elements in T_{obs}

$$f(s) = \frac{1}{|T_{obs}|} \sum_{j \in T_{obs}} f(j), \forall s \in P_\Omega \setminus S_{obs} \tag{2}$$

We use the above equation to estimate $f(s)$, $\forall s \in P_\Omega \setminus S_{obs}$. Note that the above definition may also not always estimate $f(\emptyset)$ if a marketer does not have information about customers who have not been exposed to any channels and have converted. To tackle such instances, we provide a formulation for estimation of $f(\emptyset)$.

Estimation of $f(\emptyset)$: In our formulation of f, both the effect of channels and $\emptyset$ (no-channels) are inherently captured. We split $f(s) = f(s + \emptyset)$ to separate out both the effects. $f(s) = f(s+\emptyset) = t(s)+f(\emptyset)$. Now, we assume that t is linear. Let Ω^* be the set of all non-overlapping cover sets [4] of the set comprising elements of s. For each such covering set K in Ω^*, $f(s) = f(s + \emptyset) = \sum_{p \in K} t(p) + f(\emptyset)$. If |K| is the cardinality of the set K, then the equation could be re-written as

$$\begin{aligned} f(s) = f(s + \emptyset) &= \sum_{p \in K} t(p) + f(\emptyset) + |K| f(\emptyset) - |K| f(\emptyset) \\ &= \sum_{p \in K} f(p) - (|K| - 1) f(\emptyset) \\ \Rightarrow f(\emptyset) &= \frac{1}{|K| - 1} \left[\sum_{p \in K} f(p) - f(s) \right] \end{aligned} \tag{3}$$

We average the estimates of $f(\emptyset)$ for all covering sets K $\in \Omega^*$ to compute the final estimate of $f(\emptyset)$. The estimation steps (1), (2) and (3) are combined denoted as P1 throughout this paper.

Semi-parametric Model: Alternatively, to estimate the value of $f(s)$ $\forall s \in P_\Omega \setminus S_{obs}$, one could use probabilistic estimates from a binary classification algorithm. We train the classification algorithms in the following manner. a) *Predictive variables*: For each customer, the feature vector is equal to E_i, as defined in Section 3.1. b) *Response variable* : For each customer i, the response variable is assigned 1 if the user i produced return to the marketer and 0 otherwise. The probability estimates from the classification models are interpreted as the likelihood of a customer to provide some return to the marketer given the customer has been exposed to a particular set of marketing touches. This is exactly what we are estimating in a non-parametric fashion in equation (2) and hence could be replaced by the probability estimates from the binary classification models. Thus, unlike a fully parametric model as proposed in [5],[7],

[4] Non-overlapping cover set : Given a set of elements Θ = { 1,2,...,n }, Δ = { U_1,U_2,..,U_k } is a non-overlapping cover set of Θ if $U_1 \cup U_2 \cup .. \cup U_k = \Theta$ and $U_i \cap U_j = \emptyset$, $\forall$ i,j in Δ

we suggest to use a combination of equation (1) and the probabilistic estimates from the binary classification models. The two models used in this paper for the parametric estimates are Random forest [4] and a logistic regression [6]. Also, the probability estimate of a customer to produce return given that he has been exposed to no marketing channels is equated to be $f(\emptyset)$. In a logistic regression, this leads to the effect of the intercept term towards the conditional probability. For future use, we denote the estimation step (1) combined with logistic regression as $P2$ and (1) combined with Random forest as $P3$. Throught out this paper, we denote various attribution models using the notation defined in Table-1.

Attribution of Surplus: After calculating the surplus estimates for all the possible 2^k subsets of the marketing channels, the next step is to assign the surplus to the marketing channels. This is achieved in the following manner.

- Consider all the users Υ in the given data set who have made a purchase. Let this set of users be Λ, $\Lambda \subseteq \Upsilon$
- For each customer in Λ, consider the channels the customer has been exposed to. Let this set of channels be $E \subseteq \Omega$.
- Frame a coalitional game (g, E) , $g(s) = f(s) - f(\emptyset)$ where $f(s)$ $\forall s \in$ power set of E is calculated using one of the estimation models described above.

Observe that $g(\emptyset) = 0$ and g satisfies all the properties of the value function of Shapley value [8]. Hence, using the concept of Shapley, the total gain $g(E)$, interpreted to be the fractional contribution the marketing channels in s has made to the surplus generated by the marketer is distributed to all the marketing channels involved. The channel-level attributions (pay-offs) for each marketing channel in E for the customer in Λ according to Shapley is given by

$$s_i^j(g) = \sum_{T \subseteq E \setminus \{j\}} \frac{|T|!\,(|E| - |T| - 1)!}{|E|!} (g(T \cup \{i\}) - g(T)) \tag{4}$$

The s_i^j calculated here is plugged into the matrix S_{att} defined in section 3. For the users who belong to $\Upsilon \setminus \Lambda$, s_i^j is 0, since there was no *return* from these users due to the exposure of various marketing activities. The aggregated channel attribution for all the k channels is calculated by cumulating the customer-level channel attributions obtained from (3). If n be the total users under consideration, then the aggregated channel attributions for each channel in Ω is given by,

$$\sum_i s_i^j = a_j, j = 1, 2, ..., k \tag{5}$$

and $a_j \in \mathbf{\Psi}$ as defined in Section 3.
Observe that instead of forming coalitional games (g, E) at a customer level, we could group all the customers who have been exposed to a particular set of the marketing channels. Once the grouping is done, we could form coalitional games for each set rather than at an customer-level. Such a set-level formulation would

hugely reduce the number of games formed and is hence computationally more efficient.

Attribution of Left-over: After attributing the surplus, our aim now is to attribute the *left-over*. The total *left-over* to be attributed to the k channels is obtained by using a_j computed through equation (4) and the left-over is computed by $L = \sum_i b_i - \sum_j a_j = R - S$, b_i and a_j are defined in Section 3. Next, we assign channel-wise left-over attributions using a_j by $l_j = \frac{a_j}{\sum_j a_j} * L$.

This formulation is inspired by the concept of Nash bargaining solution. Nash bargaining is a bargaining problem modeled with an outside option. If the bargaining collapses (if there is no co-operation among the players) each players gets the outside option. In our case, the players, as above, are the k marketing channels and the outside option is the surplus attributed to each channel. Since we are interested in assigning the *left-over* return to the k marketing channels, it is assumed that the *left-over* return was generated due to no co-operation among the marketing channels. Once we have computed $\mathbf{\Psi} = (a_1, a_2, ..., a_k)$ and $\mathbf{\Pi} = (l_1, l_2, ..., l_k)$, the final attributions would be given by the sum $(a_1 + l_1, a_2 + l_2, ..., a_k + l_k)$

5 Simulation Framework

All attribution models aim to formalize an answer to the credit assignment problem, but a natural question to ask is which of these is better, or which of these gives a more correct answer. Here lies one of the biggest problems of all Attribution models; the inability to evaluate different attribution models. No academic work has thus far explored this question. Here, we try to formalize an approach to evaluating multiple attribution models. Our approach includes simulating data with many of the behavioral characteristics that marketers and customer display when interacting with each other.

Here are the major aspects of our simulation set-up. Given n customers, we first simulate the touches they experience. The touches can be from different marketing channels. Next, these marketing touches lead to some propensity of the customers to conduct a transaction with the marketing organization. Finally, the goal of the simulation framework is to compare an estimation method with the true values of the simulation. At the end, we describe the parameter values we used for the simulation set-up.

5.1 Parameters of Simulation

Number of Touches For each of n customer, we simulation the number of interactions they will experience with the marketer. This is simulated to be a Poisson random variable. Let's say that the ith customer has M_i number of interactions with the marketer, we assume that $M_i - 1 \sim Poisson(\lambda)$. In other words, $\mathrm{P}(\mathrm{M}_i = k + 1) = \lambda^k \frac{\exp -\lambda}{k!}, \quad k = 0, 1, 2, \cdots, \infty$. Note that since we are adding 1 to a Poisson random variable because we would not see a customer in a dataset unless the individual had at least one interaction with the marketer.

Type of Interaction Once an interaction has taken place, we need to assign the type of interaction. This is achieved by sampling the interaction from a all possible interactions with a Multinomial distribution. Assuming there are k types of interactions, let $(\pi_1, \cdots, \pi_k)$ denote the multinomial probabilities for each type of interaction ($\sum_1^k \pi_i = 1$). Define $(Z_{i1}, \cdots, Z_{ik})$, where Z_{ij} is the number of interactions of type j for the ith customer. Then, $\mathrm{P}(\mathrm{Z}_{i1} = z_1, \cdots, Z_{ik} = z_k) = \frac{M_i!}{z_1! \cdots z_k!} \pi^{z_1} \cdots \pi_k^{z_k}$, x'where $\sum z_j = M_i$. Since all the attribution models we study only consider whether or not a channel has been activated, and not the number of touches of a particular channel, we define $X_{ij} = 1_{Z_{ij}>0}$ (where 1 denotes the indicator function).

Probability of Event Given a set of touches of different types, we need to next simulate whether the interactions $(X_{i1}, \cdots, X_{ik})$ (denoted by $\mathbf{X_i}$) will lead to a conversion event resulting in returns for the marketer. Let Y_i denote the binary event of whether the ith customer converts. We model the distribution of Y_i as

$$\log\left(\frac{P(Y_i = 1|\mathbf{X}_i)}{P(Y_i = 0|\mathbf{X}_i)}\right) = \alpha + \sum_{j=1}^{k} \beta_j X_{ij} + \sum_{j=1}^{k}\sum_{l=1}^{k} \beta_{jl} X_{ij} \times X_{il}.$$

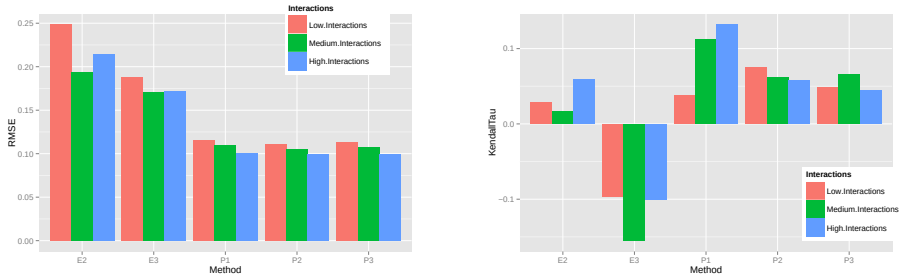
Here α is the intercept term, this term can be controlled to have different proportion of purchasers in the population. Note here that β_j is the main effect, that is, the increase or decrease in the log-odds ratio if the jth channel is added to the media-mix of an individual customer. The term β_{jl} is the interaction between the jth and l^{th} channels, in other words, the effect on the log-odds ratio when both channels are simultaneously active.

5.2 Evaluation of Truth and Comparison

For comparing the proposed models, we initially introduce the notion of *true* attributions of each marketing channel. This is done by equating $f(s)$, as defined in Section 4 to the sigmoid function with the assumed main channel effects β_j and interaction β_{jl}. Note that the same sigmoid function is used to generate the propensity of each customer in the simulated dataset. Hence, we believe a perfect attribution model would capture the marketing effects in a similar manner. Consider s be a set of marketing channels which belong to P_s, the power set of k marketing channels in Ω. If $s = (X_1, \cdots, X_k)$, $f(s)$ is given by

$$f(s) = \alpha + \sum_{j=1}^{k} \beta_j X_j + \sum_{j=1}^{k}\sum_{l=1}^{k} \beta_{jl} X_j \times X_l$$

We simulate a total of three datasets by varying the interaction effect β_{jl} between the channels - Low (interaction factor $\beta_{jl} = 0.01$), Medium ($\beta_{jl} = 0.05$) and High ($\beta_{jl} = 0.1$). We assume the number of marketing channels to be fixed to 7 in all our experiments. We provide a comparison of the proposed models



(a) RMSE (b) Kendall-Tau correlation

Fig. 1: Comparison of RMSE (Lower is better) and Kendall-tau (Higher is better) from true attributions - Proposed models vs Existing models. (Refer Table 1 for notation)

with the existing parametric models (defined in Table-1). We omit the model E1 (which leverages sigmoid function) from the comparison since we use sigmoid function to calculate the propensity of each customer in the simulated dataset. In each dataset, we generate 1 Million users. Next, we randomly sample 10,000 users and run the proposed models - iteratively for 100 times on each dataset. For each instance in the iteration, we compare the channel attributions of all the *five* models (Proposed and Existing omitting E1) with the true attributions. For comparison, we use two metrics to calculate the proximity of the models to the truth. Firstly, we calculate the Root Mean Square Error (RMSE) [2] from the true attributions for each model. This error gives a measure of closeness between the true and estimated channel level attribution.

Next, for each iteration, we compute the Kendall-tau rank correlation coefficient [10] between the estimated attribution values to channels from the models mentioned in Table-1 to the true attribution values based on the ground truth. For each dataset, we average the RMSE and Kendall-tau rank correlation over all the iterations. The results are presented in Figure 1. The two accuracy measures supplement each other. RMSE captures the big differences in the attributed values and may be insensitive to the smaller differences. Also, RMSE can be agnostic towards the ranking of the marketing channels. Whereas, Kendall-tau rank correlation captures the ranking of the marketing channels and hence is more sensitive towards the smaller differences in the attributed values supplementing RMSE.

Firstly, in the existing methods, we notice that Random Forests (E3) performs better than Elastic net regularized LR (E2)[5] in terms of RMSE, but are worse off in the Kendall-tau rank correlation. Hence, a marketer interested in strategizing the preference order of his marketing channel spends, would be better off if he does not use a parametric model E2 for the estimation phase. The proposed semi-parametric models (P2 ,P3) outperform the non-parametric models (Kendall-Tau) in case of lower interactions between the marketing channels, whereas, the non-parametric models outperform the semi-parametric models in

case of medium and high interactions. Elastic-net regularized LR performs well in the case of higher interactions between the marketing channels. Such a simulation study could be further extended to help the marketer decide the best attribution model amongst the various existing models. The analysis performed also introduces a notion of evaluation framework to compare the performance of different attribution models.

6 Empirical Results

Data Description We apply our approach to two web analytics datasets. The data was collected using Adobe Analytics, an industry-leading solution for collecting, organizing, analyzing and reporting customer activities across multiple web-connected platforms.

a) **Travel and Experience Organisation Dataset**: The data is from the months of September and October of 2013. The whole data amounts to about 2 Billion page views from 26 Million unique visitors. For a quicker examination of our proposed approach, we perform a stratified sampling of our data. We sample about 1.5 million unique users who have visited the web property during the last two weeks of the data window, of these visitors, about 300,000 of them have made a purchase in the assumed time-frame. The users could be targeted through one of the 9 marketing channels described in Table-2. For each user in the data, we have information about the various marketing channels the user has been exposed to and the purchase, revenue generated by the user as a result of these marketing interactions. In this dataset, we calculate both the order and revenue attributions of the channels. This dataset is referred to as Dataset-1 in figures.

b) **E-commerce Retailer Dataset** The next data set we considered belonged to a large e-commerce retailer. The data ranged over a 100 day period during the summer of 2013. The data contained about 54 Million interactions with 18 Million customers. For the purpose of evaluating our proposed framework, we performed a stratified sampling on our data. The sampled data comprised of 400,000 unique users of which about 200,000 have made a purchase. The users were targeted by the marketer through a variety of 10 marketing channels. We omit the tabular description for this dataset due to space constraints. In this dataset, for each user , we have information about the marketing channels the user has been exposed to and if a purchase has been made by the user, the platform the purchase has been made ("$Instore(I)$","$Online(E)$") and the type of product that has been purchased. Leveraging this information, we find channel attributions specific to product category and purchase medium. This dataset is referred to as Dataset-2 in figures.

Exploratory Data Analysis Before applying our approach, we conducted exploratory data analysis in both datasets. We perform our analysis on all the users (purchasers and non-purchasers) and specifically on the converting (purchasing) customers since these are the users who have produced return for the

Interaction	Definition	Total (%)	Orders (%)
direct	User directly navigating to site	33.8	27.5
display_ad	User clicking on a display ad	0.5	0.3
email	A click on an email from the advertizer	1.4	1.7
other_owned	A click from otherowned web properties	12.5	24
other_website	Clicks from other websites not owned by the advertizer	10	8.8
social_media	User navigates from a social media site	4.3	2
search	Clicks on organic search	31.5	28
search_ad	Clicks on search ad	5	5.2
travelagents	A visit from a travel agents site	1	2.5

Table 2: Dataset-1 : Definition of various marketing channels along with the frequency of their occurences in purchases as well as the whole data

marketer. Table 2 provides definition of each marketing channels and has details about the number of times each of the marketing channels was exposed to users (Dataset-1). We also analyzed the number of marketing touches the customers in our dataset were exposed to. Due to space constraints we only provide a detailed analysis of Dataset- 1. From Table-1, we observe that *direct* and *search* are the marketing channels that all the users and also specifically purchasers are most exposed to. Whereas *display_ad* ,*travelagents* and *social_media* are the least occurring marketing channels. An accurate attribution model should gauge the incremental and interactive effect of each marketing channel without biasing for the frequency of occurrence of the channels. Also, from the exploratory analysis, we observe that more than 80% of the non-purchasers and more than 50% of the purchasers have only one marketing interaction. Traditional attribution techniques such as First touch and Last touch would attribute all the generated return from the purchasers to these marketing channels without considering the users that have not made a purchase. An ideal attribution model should find the true attribution of each marketing channel by contrasting the purchases with non-purchases.

Estimation We calculate the surplus estimations using the proposed models in the Approach section. Our assumption of linearity of t and the approximation of the value of $f(\emptyset)$ was validated by noting that the parametric estimate of $f(\emptyset)$ using a logistic regression and Random forest was fairly consistent with the estimated $f(\emptyset)$ using equation (3). For the second dataset, along with the channel-level attributions for the whole dataset, we calculate channel attributions specific to each product category and purchase media and for this purpose, we train different models for each of the product categories and purchase medium. Further, we compare the channel-level attributions from our proposed models with the existing methods (Refer Table 1 for notations). After the estimation step, we calculate channel attributions using equation (4). We do not include the left-over attribution presented in equation (5) in our results since

we aim to compare the attribution results with existing marginal attribution models. In the case of Dataset-1, we calculate both order and revenue channel attributions. We present few of the scaled attribution results in Figure-3 and Figure -4. The results of all the three proposed models are consistent and cross-validate each other. The credit assignment to the marketing channel *travelagents* and *other-owned* channels in the order attributions of Dataset-1 are noticeable. The channels receive a higher credit than few other channels that occurred more frequently in the dataset satisfying a notion of $fairness$. In Dataset-2, we portray the extensibility of the proposed framework. Hence, we calculate channel-attributions at a more granular level, specific to the product-category using the proposed framework. We note that the channels *direct* and *display* have a higher order attribution specific to Product-1, while they have a lower order attribution in case of the online media. Such insights specific to products give the marketer an added advantage while planning his overall spend across different marketing channels

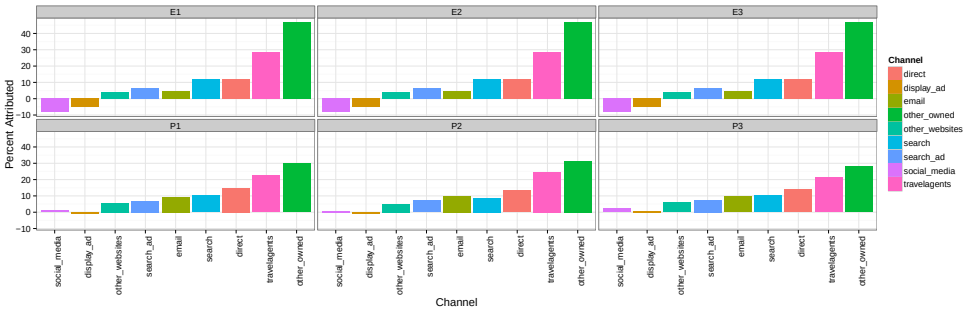


Fig. 2: Dataset-1, Channel-level order attributions compared across various attribution models defined in Table-1

Comparison We noted that the present state of the art data-driven attribution models were proposed by Shao et. al [7] and Delassandro [5]. In both the papers, the authors compare the proposed models with traditional attribution models like Last Touch Attribution (LTA) to prove the accuracy of their models. The exploratory data analysis we performed combined with the insights from the results made in the above section, it could be deduced that our models are capturing the true attribution of channels without being guided by the frequency of the marketing channels in the datasets. Hence we do not provide a comparison of the proposed approach with LTA. Instead, since our approach leverages Shapley value, we compare the proposed models (denoted by P1 , P2, P3 in the figure 3) with the existing parametric models defined in Table-1 (denoted by E1 , E2, E3 in the figure 3). We observe that the results from our models are fairly consistent with the parametric model in case of Dataset 1 but are slightly different in case

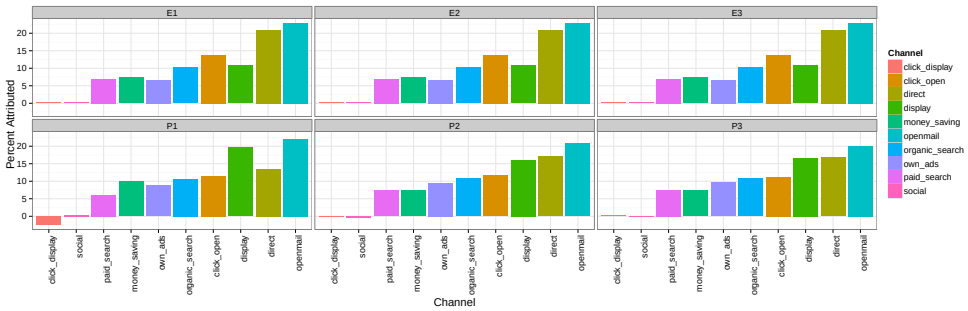


Fig. 3: Dataset-2, Product-1, Channel-level order attributions compared across various attribution models defined in Table-1

of Dataset-2 Product-1. We investigated this by observing the interaction terms β_{jl} for the marketing channels in Dataset-2 for model E1. The interaction terms were observed to be high. In the simulation study, we have established that the existing models do not perform well when there is high levels of synergy between the marketing interactions. Hence, we claim that the proposed models P1, P2, P3 would be closer to the true attributions in the case of Dataset-2 than the existing models.

7 Conclusions and Future work

The goal of this paper it to fill important gaps in the literature on attribution modelling. Past work on attribution has used parametric approaches to estimate the likelihood of conversion due to exposure to a set of marketing channels. We show that these approaches do not work well when there are high level of synergies between the marketing channels. This is shown using a simulation study. We propose a non-parametric approach to estimate the exposure effect. The approach performs well in presence of high level of synergies between the marketing channels. We use Shapley Value to assign the surplus that is derived from the exposure effect model. The Shapley Value approach is similar to [5]. We argue that it is important to distribute the left-over return to the marketing channels. Past work has not provided any solution to this problem. We propose an approach that uses the marginal attribution solution as the input to assign left-over return to the marketing channels. We apply the proposed attribution model on two real world datasets and present the results.

In future, we can extend this work in multiple directions to account for more realistic scenarios. The effect of advertisements is believed to decay over time, in this work, we have not attempted to capture this effect in the simulation model, nor do the different existing approaches capture these effects. This is an area which may be explored in future. The computation of Shapley value increases exponentially as the number of players in the game increases. While this may

be feasible when we model marketing channels (which may number between ten and twenty), it becomes infeasible when modelling marketing campaigns, which may run into hundreds. Hence, investigations to come up with closed form approximations for the proposed model could be another research direction. We provide a simulation engine that captures some of the important behavioural phenomenon known to be exhibited by customer when interacting with a marketing organisation, however, this may be further expanded by modelling more such characteristics.

References

1. Abhishek, V., Fader, P.S., Hosanagar, K.: Media exposure through the funnel: A model of multi-stage attribution. Tech. rep., Working paper, Carnegie Mellon University, Pittsburgh (2013)
2. Armstrong, J.S., Collopy, F.: Error measures for generalizing about forecasting methods: Empirical comparisons. International journal of forecasting 8(1), 69–80 (1992)
3. Binmore, K., Rubinstein, A., Wolinsky, A.: The nash bargaining solution in economic modelling. The RAND Journal of Economics pp. 176–188 (1986)
4. Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)
5. Dalessandro, B., Perlich, C., Stitelman, O., Provost, F.: Causally motivated attribution for online advertising. In: Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy. p. 7. ACM (2012)
6. Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., Tibshirani, R.: The elements of statistical learning, vol. 2. Springer (2009)
7. Shao, X., Li, L.: Data-driven multi-touch attribution models. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 258–264. ACM (2011)
8. Shapley, L.S.: A value for n-person games. Tech. rep., DTIC Document (1952)
9. Sinha, R., Saini, S., Anadhavelu, N.: Estimating the incremental effects of interactions for marketing attribution. In: Behavior, Economic and Social Computing (BESC), 2014 International Conference on. pp. 1–6. IEEE (2014)
10. Skidmore, A.K.: A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model. International Journal of Geographical Information System 3(4), 323–334 (1989)
11. Xu, L., Duan, J.A., Whinston, A.: Path to purchase: A mutually exciting point process model for online advertising and conversion. Management Science (2014)