ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/309416430

Toward a dynamic attribution model for marketing

Thesis · September 2016

DOI: 10.13140/RG.2.2.26999.21927

CITATIONS
0

READS
2,612

1 author:

Sami Sakly
Université Paris-Saclay

2 PUBLICATIONS **1** CITATION

SEE PROFILE

All content following this page was uploaded by Sami Sakly on 25 October 2016.

The user has requested enhancement of the downloaded file.

Stage de Master 2 Informatique
Apprentissage Information et Contenu

Toward a Dynamique Multitouch Attribution Model for Marketing

Auteur :
Sami Sakly

Responsable de stage :
Hanania Mickael Ouazan

Organisme d'accueil :
ARTEFACT

Secrétariat - tél : 01 69 15 81 58
courrier électronique : alexandre.verrecchia@u-psud.fr

Contents

List of Figures

Abstract

This work has been realised during my internship at Artefact as part of my Master 2 en Informatique at Université Paris Saclay under the Machine Learning for Information and Content specialization.

Digital advertising spending share is growing faster then ever compared to traditional advertisement [1]. Usually using multiple channel to reach the target, advertiser could not measure directly the contribution of each delivered advertisement and its impact on the conversion [2]. Multi-Channel attribution is an interesting managerial problem that attracted many marketing gurus but also researcher from the academia in the last few years. In this report, we will apply some of the latest marketing attribution techniques to analyze for the largest digital advertiser in France.

The results presented in this report give clear insights to marketing planning team how to allocate budget over the different channels optimally.

Mots clefs

online advertising, cross channel attribution modeling, algorithmic attribution, multivariate logistic regression, marketing.

1. Offline advertisement i.e. TV, Radio, printed ads..
2. In the marketing area, a conversion denote which ever action of engagement from the client side, going from either subscribing to the news letter to signing or upgrading a contract. It's the main goal of an advertisement campaign.

CHAPITRE 1

Internship Overview

1.1 Introduction

As part of my MSc curriculum at Université Paris Saclay, I have chosen to realize my end of study internship in a start-up over a research lab in order to have a taste of the business world and know learn how it's like to take responsibility in real business situation. I was not disappointed by my choice, my experience at Artefact was very enriching and insightful.

In this report I will describe the project I realized during this internship which is an analysis of attribution using different algorithmic marketing attribution model for the largest France online advertiser with over than 32M EUR for Media advertising for the year 2016. Also, I have done a lot of unstructured big data processing and optimization to prepare the correct format of data.

1.2 Acknowledgment

I am really thankful to all Artefact employee from whom I have learned a lot and with whom I share unforgettable memories. I won't forget to thank all the AIC Master teachers and staff for providing a solid curriculum and enriching our knowledge with all the courses and projects we have been through.

1.3 Hosting company

With a near constant growth of 20% of the advertisement spending in the last 5 year to reach 60 Billons $ in 2015[1], the Digital Marketing is among the hottest and most promising business area. The democratization of many new Advertisement/ Marketing tools led to a more accessible but complex market where specialized start up like Artefact provide huge added value by backing advertisers in every step in the design and the execution of their Digital Marketing strategies. Founded at the very beginning of 2016 as a result of a merger between two start up both founded in 2014 : Augusta Consulting a digital marketing consulting start up and Little Big Data a unique technology start up building marketing software using machine learning.

With a common goal : "Provide the consumer with the best advertisement experience using Data-Driven Storytelling" the merging of knowledge from both of the business consulting and technology world gave birth to a unique environment where business profile interact with the R&D engineers and Data Scientists to embody client's perception into the real world via different product such as Ads recommendation, very precise audience segmentation and a dashboard showing both aggregated and individual marketing KPIs.

Data-Driven Storytelling

With the rise of the real time bidding [1] and the democratization of the programmatic advertisement, the user online experience has become non organic and very unpleasant which resulted in an increasing of Adblock [2] usage costing more than 22 Billons $ as reported by PageFair an

1. Abbreviated as RTB, it's the technology simulating stock market for display add that appear on website. It gives the advertiser an automated way to buy media placement depending on its features i.e : size, site popularity, day time.. in less than a second

2. Adblock are software that blocks tracking the browsing history by 3rd party and showing media advertisement on visited websites.

anti-Adblocking solution. One proposed solution to attract these internet users back is by using Data-Driven story telling. It leverages the growing availability of data sets to analyze and uncover new angles on stories that will be used to adapt the marketer message to the consumer offering him exactly what he's looking for. It's the usage of data to complete the narrative for each audience [3].

Data Scientists at Artefact play an important role since they are helping consultant understanding business problem dealing with technical issue and provide the right way to use tools in order to satisfy client inquiry and also take part in product development by implementing machine learning models and big data processing workflows within the R&D department. I feel lucky taking part in such an experience where I have learned a lot about the digital advertisement business but also was exposed to a very important stack of technology I will go through later.

3. An audience is a group of people or identifier that share the same characteristic and business description. i.e. Millennial female living in Paris that have visited a specific product section of a cosmetic retail website is an example of audience

Chapitre 2

Problem Statement and existing solution review

'Half the money I spend on advertising is wasted; the trouble is, I don't know which half'.

*J*ohn Wanamaker

2.1 Introduction

Internet advertising has become an essential element of the marketing strategy and promotional mix of many industries. Advertisers today use in a daily basis a variety of online marketing channels [1] to reach potential customers, including display marketing and paid search, as well as e-mail, retargeted displays, affiliates and social media advertising. Customer can also come to advertiser website from non paid channels such as direct type-in or via non sponsored links appearing on search engine.

The usage of these multiple channels leads to multiple visits before concluding the conversion transaction. It's why now more than ever, 'J. Wanamaker' quote has a sense and point to one of greatest advertiser problem; what's the result of reaching a client over one of these multiple channels? On which channel should I put more money? What's the benefit of adding a new channel? This is exactly where **Attribution models** were introduced. They attempt to define how each interaction with advertisements along the customer's journey contributes to the customer's decision whether to realize the conversion action or not.

2.2 Customer journey

We have been using this term multiple time since the beginning of the report and since it's a key element to attribution element and for the marketing business in general, we thought of explaining it in this section in order to clarify this concept and introduce the assumption we are adopting in this study.

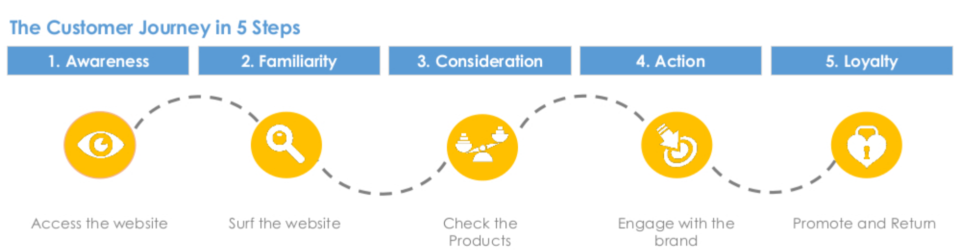

Figure 2.1 – Customer journey through the funnel framework illustration

A customer journey is all of its interaction with regards to the brand over one or multiple channels before or after making a conversion. We all have had customer journey before buying a products Usually, marketer use the funnel framework Figure 4.2 to model customer journey. It's a 5

1. A more detailed definition of each channel will be presented in the next chapter

steps framework where at with the first touch-point the future client is developing some kind of awareness to the brand which become familiarity with a little more touch-point. The next move forward in the funnel is very important, since the target is starting considering the product i.e. visiting products web pages, going to product comparing sites.. After that the customer will either take a negative action and thus leaves the funnel or make positive action and become a client which will allow him to stay in the funnel for the next step : Loyalty. At this moment, it's the company that will take branding, promotional actions to keep the client satisfied.

One thing we should keep in mind that at any time the client could leave the funnel for any reason and become cold with the brand.

2.3 Attribution models

Thanks to the development of Internet technologies, online advertiser can follow users exposure to advertisements using cookie data. A cookie is a small piece of data that contains information about the user's browsing habits, which can be transferred to a third party (advertiser) when the user visits a website[7]. This created the possibility to combine this information with customer behavior data across the different websites. Furthermore, the tracking can be extended to search engines. All this information is very important for marketing direction but such capabilities have raised privacy concerns. Combined, these pieces of information enable marketers to analyze customer's journeys to conversion thoroughly. This information can be used to recognize how different advertisements have contributed to the final purchase decision via attribution models. Given the proliferation of online channels and the complexity of customer journeys, measuring the degree to which each channel actually contributes to a company's success is demanding. Some of the reported benefits are :

- A better understanding of the customer, since it allows a global view of the ecosystem of touch-point and quantify the effect of each interaction with the brand all along the conversion journey.
- Knowing beforehand how many will probably convert and how much you will spend to acquire these customer allow a solid ROI analysis and prediction.
- Making marketing more effective by distinguishing the best path leading to conversion for each customer on an individual level

It has to be said that attribution are not in any way an innovation or a new invention. It has been around for few decades now. It arises in traditional (offline) advertising channels such as TV, radio and print. However, online channels offer a unique opportunity to address the attribution problem, as advertisers have dis-aggregated individual level data which were not previously available. Unfortunately, this was not available until very recent. Consequently, marketers had no choice but to use heuristics and rule based-techniques to estimate the attribution for multi-channel conversion path. Here below I will try to give a brief description of this different methods.

Single touch-point attribution model

The first and predominant set of heuristics adopted by marketer today attribute all the conversion value to one point of the customer journey. It's the basic level of attribution that yields approximated results but are very simple to use and implement. The most common one are :

- Last click attribution : last click model assigns all the credit to the last advertisement that was clicked just before the conversion
- First click attribution : first click model assigns all the credit to the first advertisement that was clicked just before the conversion
- First touch-point attribution : first touch-point attribution assign all the credit to what ever first touch-point that introduced the customer to the brand.

Of course, it has been pointed in many research that this models are flawed [4] since they only use a small part of the information.

Multi touch-point attribution model

Several multi-touch models have been proposed to account for the combined effects of advertisements of various channels. There are two different family of multi touch-point attribution models : Rule Based MTA and Algorithmic MTA.

Rule based multi touch-point attribution model

Similar to single touch-point these kind of attribution uses fixed rules to give credit to multiple touch-points instead of only one touch-point. We cite here some of them :

- Linear attribution : Linear attribution model assigns the same credit to all considered touch-points before conversion
- U-shaped attribution : U-shaped attribution model assigns more credit to the first and last touch-point and less credit to all intermediate touch-points before conversion
- Time-decay attribution : Time decay attribution model assigns decreasing credit to older touch-points before conversion.
- Raw attribution : Raw attribution model assigns all the credit to all touch-points before conversion.

These model are more sophisticated than the previous one, but still making very bold assumption which may corrupt the model output. The table in figure 2.2 shows an example of how each attribution model presented until now will credit a four touch-point customer journey.

TYPE	NAME	DESCRIPTION	What touchpoints would get credit			
			#1	#2	#3	#4
Single touch-point	Last click	Last click gets all the credit	0%	0%	0%	100%
	First click	First click gets all the credit	0%	100%	0%	0%
	First touchpoint	First touchpoint (click or ad impression) gets all the credit	100%	0%	0%	0%
Multi touch-point	Linear	All touchpoints get equal credit	25%	25%	25%	25%
	U-shaped	Touchpoints at the beginning and the end get more credit	40%	10%	10%	40%
	Time decay	Touchpoints at the end get more credit	10%	20%	30%	40%
	Raw	All touchpoints get all the credit	100%	100%	100%	100%

Figure 2.2 – Table showing example of attribution result for a 4 touch-point customer journey (source Criteo white paper)

Algorithmic multi touch-point attribution model

Previous visits may influence the users subsequent visits, such that the customer may return to a website through the same channel (**carryover effects**) or through different channels (**spillover effects**) as presented in [2]. This kind of hidden influence and connected interaction cannot be discovered by previous presented models. Being promoted as the marketing next thing, attribution attracted few researchers from academia who introduced algorithmic attribution models or what's also called dynamic attribution models, since each unique user journey receive a different credit allocation. In this section we will describe briefly these approaches.

Logistic regression attribution models :
Logistic regression was first used in marketing in [8] for modeling customers behavior on website. It's is generally used for classification problems. In a binary classification problem, we attempts to classify observations into two distinct classes, normally either to a positive (success) or to a negative (failure) class. The basic idea of logistic regression is to estimate the odds that an observation belongs to a certain class based on the information about the observation. Odds is the ratio of probability of an observation belonging to a class divided by the probability of not belonging to the class. Equation (2.1) show the formula of logistic regression and equation (2.2) shows the decision function.

$$p(C|\Phi) = \sigma(W^T\Phi) = \frac{1}{1+\exp(-W^T\Phi)} \tag{2.1}$$

Where C is the predicted class, Φ is a vector of input variables and W^T is a vector containing coefficients. If the probability is higher than a threshold value, then assign the observation to a positive class, otherwise to a negative class :

$$C(\Phi) = \begin{cases} 1 & if \quad p(C|\Phi) > Threshold \\ 0 & otherwise \end{cases} \tag{2.2}$$

The threshold value is a key parameter that is problem specific and that is between 0 and 1 since it's a measure of the probability that below we consider the observation belonging to the negative class and to the positive class otherwise.

Once the logistic regression is tuned, marketers will interpret its parameters (coefficients of predictors) for allocating budget over the different touch-points, or use their prediction ability to test the effect of interacting with only one channel at the time. One problem has been pointed out is the variability of this parameters. This has been solved in [11], by bagging logistic regression models and thus reduce the variability of the estimate and at the same time keep a high accuracy, i.e. prediction of conversion using touch-points.

Graph-based Markovian framework :
This method was presented in [5] followed two years later with another extension to the first paper where they provided results obtained by applying their approach in [6]. The author of the paper received EMAC McKinsey best marketing dissertation award for the year 2015 [2]. The usage of Graph-based Markov chain had an enormous success since in this approach we can identify structural correlations in individual level data ,i.e. we do not aggregate customer journey information (the chronological order of touch-point) as we do with logistic regression model.

2. https ://marketing-dissertation-award.eu/2015-winners

In this model, the authors represents customer journeys as chains in directed Markov graphs. A Markov graph $M = \langle S, W \rangle$ is defined by a set of states

$$S = \{s_1, s_2, .., s_n\} \tag{2.3}$$

and a transition matrix W with edge weights where W_{ij} is the transition probability from state i to state j.

$$W_{ij} = P(X_t = s_j | X_{t-1} = s_i) \tag{2.4}$$

$$0 \leq W_{ij} \leq 1, \sum_{j=1}^{N} W_{ij} = 1 \forall i$$

Applied to attribution, each vertex (state) of the Markov graph is a touch-point and each edge weight is the transition probability from one channel to another. This approach is very scalable since it's independent from the number of customer journeys, the graph will only contain states , the different considered touch-points. Another important benefit from using this model is the capacity of identifying structural correlation. Actually, the author of this paper were inspired by the work in [3] applied to search engine - display study. They adopt the notion of ***Removal Effect***, where Removal effect(s_i) is the change of probability of conversion if we have to delete the state s_i from the markov graph. Such a computation enables us to simulate counterfactual analysis on historical data and is therefore well suited for measuring the contribution of each channel or even channel sequences.

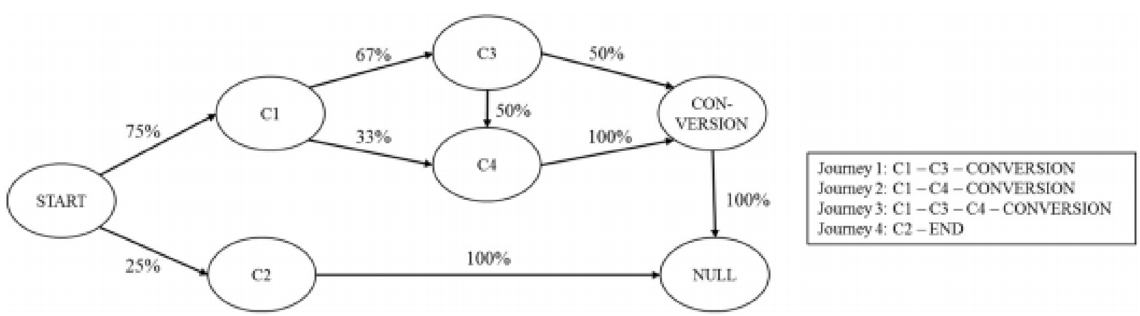

Figure 2.3 – Example Markov attribution graph

Figure 3.1 shows an example of a sample Markov chain graph that have 4 channels (C1,C2,C3,C4) to which we add :

- START state that expresses the starting point of a journey for (initial probability computation)
- CONVERSION state indicate if the journey ended with a conversion event
- NULL state indicate if the journey dit not ended with a conversion event.

Transition probability are printed over each edge the graph according the 4 journey indicated on the figure. We can find a computation of removal effect for this example in [3].

Other models :

We also had reviewed other proposed algorithmic models that has been introduced lately such the usage of HMM [3] in [2] where channel interaction are observation and the hidden states reflect the engagement of the customer through the funnel (disengaged, engaged, converted). In [9] the author introduce the usage of Multivariate Time-Series Models for studying the interaction between display and search ads.A multivariate time-series model consists of multiple individual time-series models, in which each time-series influences each other based on some principles.

We prefered to only give a description to method that we decided to implement for . Our choice was based on two main characteristic : The easiness of interpretation in order to convince stack holder to invest in the industrialisation of the the project and the easiness of implementation since we had a heavy preprocessing task and limited time to realize the project. Though curious reader can use pointed out element in the bibliography of this report to get a deeper look to other models. Since this was the first PoC for such a modeling task in the company, we made the choice of not going deeper in algorithmic method analyses since the data preparation task is too heavy.

2.4 Conclusion

In this section we presented the marketing attribution problem and some other concepts related to marketing that we need to know to understand the subject. We then presented some existing solution used for attribution with an emphasis to the logistic regression and the graph based Markov framework which we will use in our analysis.

3. Hidden Markov Model

CHAPITRE 3

Data and analysis methods description

3.1 Introduction

In this chapter we will first describe the tools we have used for this project, the characteristics of the data and the pre-processing pipeline. We will than describe the algorithmic approaches we implemented to analyze the attribution.

3.2 Tools and Software

Dataiku

At Artefact, Data Scientist uses Dataiku platform to analyze and build models over the available data. Dataiku is platform where we can plug our data whatever its type (synchronized from database, a collection of CSV file, or HDFS file..) and build preprocessing, modeling workflows using GUI [1]. The platform is deployed over a cluster of 12 machines with more than 40 CPU and 300 GB of RAM, where a Hortonwork Hadoop distribution is also installed. We used SPARK, Hive [2] for distributed computation and Python and R as main programming language to realize data manipulation and modeling without any job submission overhead, and thus win in productivity following the Sillicon Valley mantra "*Fail Fast, Fail often*". Indeed, nothing comes without cons, using an extra layer for analyzing the data added another layer of complexity to the stack that made some error debuging painful and required deeper understanding to choose the right configuration for some of the jobs.

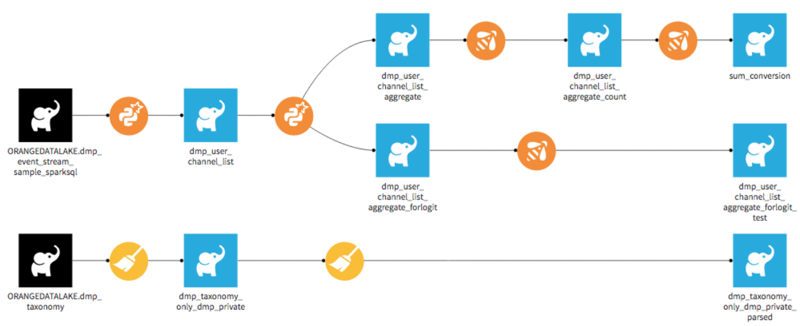

Figure 3.1 – A dataiku workflow example

Data Management Platform

Abbreviated as DMP, the data management platform is the hype in digital marketing in the last few years. Many technologies leaders (Oracle, Google, Adobe..) invested heavily in these kind of platform that is used to collect all user interaction using cookie and unique identifier on mobile. We distinguish two type of data :

— Navigational behavior through tags implemented in company website, partners website or mobile application.

1. Graphical User Interface
2. Hive is an SQL like language that works over distributed databases using Hadoop paradigm.

— Media interaction data : ad campaign exposure, click, paid search engine referral, social network referral..

It's also possible to enrich DMP users profile with non personal data from CRM such as demographic and geographic information, but also contract specific information which make the DMP a very powerful tool for new client acquisition and customer recognition.

DMP comes with a lot of other functionality such as advanced reporting, performance prediction, segment similarity and the most important advertising campaign activation through data transfer to other marketing technologies. This is beyond the scope of this project, but for curious reader we have been using Oracle DMP called "Bluekai" [3]

3.3 Data description

The most important element for a successful analysis is data. Luckily we have a lot of it at Artefact. Basically, since we host DMP data in our cluster to realize other type of analysis over it, we have all the recorded touch-point between a user and the brand. There are banners (display), social media and search engines. These are the most use internet channels for advertising.

Banner is an image or ad-content (as flash animation) that is delivered by an ad server. Website can show lots of display ads at the same time. Most of the time this become quickly annoying and push internet users to install ad-blockers. However, using Bayesian framework it was shown [10] that banners have a positive effect on the purchasing activity. It gets in line with the results from [9] that show the positive impact of banners in the conversion process.

One other important touch-point is search engines, such as Google, Yahoo, Bing. Unlike the banners, the search engine gives a response to the user's specific query. Therefore, one can expect to have a higher CTR for the search than for the banner since it's really user initiated. We can differentiate 2 sub-channels for search : SEM and SEO that are search engine marketing and search engine organic correspondingly. On Google search results one can observe the SEM results marked by a special label "Announcement". Announcements correspond to the keyword that was searched by the user. The advertiser should however pay the price that depend on several factors : popularity of the selected keyword, the maximum budget of the campaign and a quality score.

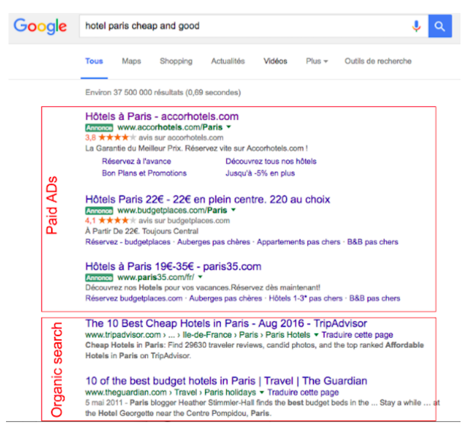

Figure 3.2 – Search results overview

.

3. https://www.oracle.com/marketingcloud/products/data-management-platform/index.html

Search engine organic result depends on the quality of the organization of the website. The search engine has it's internal mechanism, which ranks the pages and gives the result according to the keywords and the rank calculated.

The growth of the popularity of social networks brought also a channel that is actively used by the advertisers. Consider Facebook. There are two possible ways to make advertising - by email or by lookalike criteria. Email is straightforward : the advertiser gives an email of the target person and Facebook shows an advertisement to that person, if the email is the same as the one attached to the account. Lookalike is a method to show the advertisement to the people, who match the certain criteria.

When we extract the data from DMP and put it into Dataiku. It's very unstructured logs since it comes from different source : Adserver data, DMP navigational data, CRM data, campaign specific data and the list goes on.. Figure 3.3 show an example :

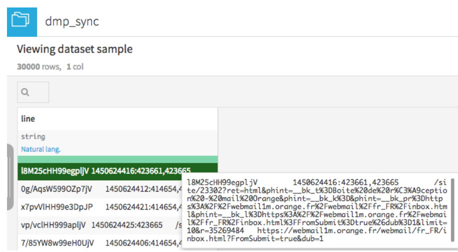

Figure 3.3 – Raw data in Dataiku

.

Here each line correspond to an event. Event should be parsed to extract every different information encapsulated in it. Among the most important field to find are the user ID and the timestamp which will be used to get later all user event in one place and order it by the time of occurring. The parsing means that we process every line to a columnar output mapping every different information to the correspondent column.

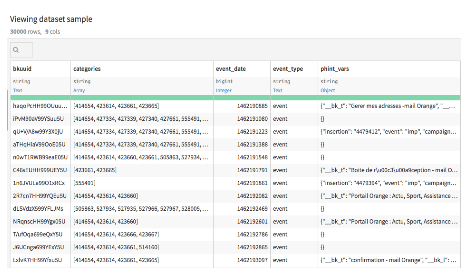

Figure 3.4 – Parsed data

.

Here each line represents an interaction with the unique user. In order to use the power and flexibility of Python we aggregate the last dataset by key, which is "bkuuid" in this case. We want to have all the history of the user in one line.

Next, we identify the channels that we want to find in this dataset. At the beginning of this section we have already covered the basic marketing channels as banners, search and social networks. A more explicit breakdown is given below :

bkuuid	event_stream	events
string Text	array Array	bigint Integer
JdRxR7Xs99ejfvhu	[{"tag_vars":"{\"bknms\": {\"lang\": \"a488ae48069...	22
9u1MsUHH99Y9Oo...	[{"tag_vars":"{\"bknms\": {\"lang\": \"a488ae48069...	10
Hyqu6LX599e/YHYJ	[{"tag_vars":"{\"dt\": \"0\", \"r\": \"774575174\", \"li...	5
fSJLjlWq999bsBJV	[{"tag_vars":"{}","referrer":"http://www.allocine.fr/...	4
CFTlzpP7999kjICV	[{"tag_vars":"{}","referrer":null,"bkuuid":"CFTlzpP7...	8

Figure 3.5 – Event stream

.

Tableau 3.1 – Digital channels

Online channel	Description	Channel type
Type in	Direct visit entering URL	Customer - initiated
SEA	Search engine advertising - paid search.	Customer - initiated
SEO	Search engine optimization - paid search.	Customer - initiated
Display	Banners and other graphical objects that show advertisements	Firm - initiated
Email	Send marketing messages using email	Firm - initiated
Retargeting	Personalised advertising based on the browsing history. The goal is to push the user towards purchase	Firm - initiated
Social media	Showing ads on Facebook, Twitter, Instagram, etc.	Firm - initiated
Affiliate	Commission based marketing towards advertiser. Could be a coupon or promotion on a third-party website that refer to the advertiser. The affiliate is then rewarded when the user has come the website	Customer - initiated / Firm - initiated

We have just taken the users, who appeared after 15 June 2016. The total dataset is more than 1 Billion events and more than 50 millions unique users. After anomalies cleaning (referencing and malicious bot visits) we obtained the following dataset :

Tableau 3.2 – General information

Type	Value
Total number of users	2,311,407
Total number of conversions	6403
Maximum number of events	50
Minimum number of events	1

You may think why we have not so many conversion ? We thought the same and we passed days looking for an answer. And here's our finding : the company we customer journey for the company providing data is not fully digitized. At some point you can ask to be called by a call center agent and you can accomplish the rest of your "conversion" offline. They also have physical store where you can buy their offering, so we see many people interacting with the brand website, but we do not have the ability to confirm their conversion. This one of the most complex problem in the digital marketing field : attribute offline conversion to online activities. This is beyond the scope of the prototype, so we will work our way around with only the conversion we are 100% sure about.

3.4 Modeling details

Our main goal of this analysis is that results be of practical use, thus the output of an attribution modeling approach must support decision making. Marketing activities,like any other business activities, are driven by the amount of budget at the disposal of the business unit manager. So the allocated fund is the primary source of constraints in marketing planning process. Therefore, the framework for analysis should be able to address financial matters directly, or at least yield directive for optimizing the budget allocation.

As expressed in Section 2.3 of the previous chapter, we have decided to use Logistic Regression and Markov chain model for this task since both yield the required result and both are considered state of the art in dynamic attribution modeling.

Logistic regression model

Logistic regression is a classification technique that yield the odds of belonging to one class and not the others using the input features. As in [11] and [7] we select the number of interaction on each channel to be features representation of the input since this satisfy the requirement that coefficients of the model should describe how each individual channel performs and that each describe the change in the customer's conversion probability after they have interacted with a particular channel.

Data processing : To prepare the data to the required format we had to start from the list of event for every unique id and provide the number of time each channel in Table 3.1 appear in that list.

We then partition the data into 80% training and 20% testing dataset in order to check that our models does not overfit. Of course, we keep original positive / negative class ratio for both datasets. The problem is really skewed and we can easily get erroneous results.

Two important decision we took after building the first models are, to only keep these 5 channels : **SEO**, **SEA**, **Type-In**, **Display** and **Retargeting** since only these channel appeared sufficiently on events to not be considered as outlier. We would have a 0 in prediction coefficient for Email for example. This make sense in some way, since we are dealing with new customer acquisition campaign, we do not have already users email to contact them. For the other channels, it was mainly tag related problems either no correctly deployed or it was blocked by another entity. The second decision was to coalesce consecutive interactions that occur within 10 minutes interval on the same channel as the same touch point. Thus we will reduce the length of the customer journey length without having any drawbacks on the model. It could be seen as a normalization of parameters.

Modeling : Once we had our training / testing data ready, we used the sklearn Logisitic Regression implementation to fit the model using 4-fold cross-validation. As for metrics selection, we used the accuracy as our first metric but since the problem is very skewed we also used AUC [4] too.This metric describes the probability of randomly chosen positively labeled observation to be classified to the positive class by the classifier.

In [11] author proposed to use an altered version of bagging to reduce the variance of the model and thus increase its robustness. Bagging is an ensemble method that is used to improve classification performance by combining multiple weak predictions to produce a single stronger prediction. The idea of bagging is to take N bootstrap samples of the training set and estimate N models using each dataset once. A bootstrap sample is constructed by randomly selecting M observations from the training set, with replacement, where M is the size of the training set. But in the case of [11] they also performed bootstrapping over the feature in order to delete any structural dependency between features. This is similar to what Random Forest algorithm do. We will also

4. Area Under Curve of the Receiver Operator Characteristic

report results of a similar implementation in the next chapter. We choose the same parameters for the bagging creation :

— maximum samples per estimator 5% from the total events (around 50K samples)
— maximum feature proportion 80%
— number of estimators = 100

Markov chain model

As explained in the previous chapter, the random walk in markov graph is the latest published attribution algorithm and the one that yielded the best result to the point that it attracted some folks from the open source community to implement it and made the package available for free. Focusing on having the largest possible benchmark for this PoC, we did not try to reinvent the wheel again. We used the available implementation [5] in R.

Since we do not specify explicitly the time of each interaction with a channel we only consider the succession of event, i.e. in our point of view every event are equidistant in term of time, the Markov chain graph is a deterministic discrete valued. It's also deterministic (since we have a finite number of channel. Thus the weights in Equation (2.3) can be easily estimated by the sample proportion as follows :

$$W_{ij} = \frac{n_{ij}}{\sum_{k=1}^{m} n_{ik}}, \forall i \tag{3.1}$$

Here n_{ij} is the number of times we observe a transition from state i to state within our input data and the same for n_{ik} and m is the total number of states. The weight are nothing other than the frequency of path showing in the input.

Other approach could be used to estimate the weight such as stochastic simulation such as Monte Carlo Markov Chain (MCMC).

In the previous chapter we said we will use removal effect to compute the real contribution of each channel. The core of Removal Effect is to remove each channel from the graph consecutively and measure how many conversions (or how much value) could be made (earned) without the removed one. The logic is the following : if we obtain N conversions without a certain channel/touchpoint compared to total conversions T of the complete model, that means the channel reflects the change in total conversions (or value). After all, channels/touchpoints are estimated : we have to weight them because the total sum of (T – Ni) would be bigger than T and normally it is.

Data processing : Of course using an existing package imposed some kind of data structuring. But nothing too fancy, since we already have the ordered list of channel for every user, we just created a list of all element with a ">" as separator and we also added "start" and "null" state , at the beginning of every touch-point list and at the end for the one do not ending with conversion.

Modeling : The usage of the package was quite straight forward. It comes already with extra functionalities such simple heuristic attribution model implementation (First-touch, Last-touch and linear) and K-order markov chain coefficient estimation. We will compare results yielded for each of the proposed model in the package and specifically for different Markov model order.

5. a link to the package description : https://cran.r-project.org/web/packages/ChannelAttribution/ChannelAttribution.pdf

CHAPITRE 4

Results and discussion

4.1 introduction

In this chapter we first present our two models results and then compare with findings in the studied articles and thesis. Our dataset, contains 5 unique channels and is composed of 2,311,407 users' journey (observations) which contains 6403 conversion. The positive class represents only 0,2% from the total events.

4.2 Results interpretation

Logistic Regression results

Now, because we seek to compare the performance of channels after a customer interacts once with each channel, the estimation of channel importance is as follows. The exponents of coefficients as such cannot be used to deduct directly the incremental effect of a channel since the odds ratio is not linear. However, as in [7] we can calculate the theoretical probability of the case when a user has interacted with only a particular channels using Equation 1.

The simple logistic regression model achieved 85,5% accuracy on the cross validation with a standard deviation of 10^{-3} and 85% accuracy on the test dataset. It scored a mean AUC 0.87. Here below (Figure 3.6) show the ROC graphics of the 4-fold cross-validation step.

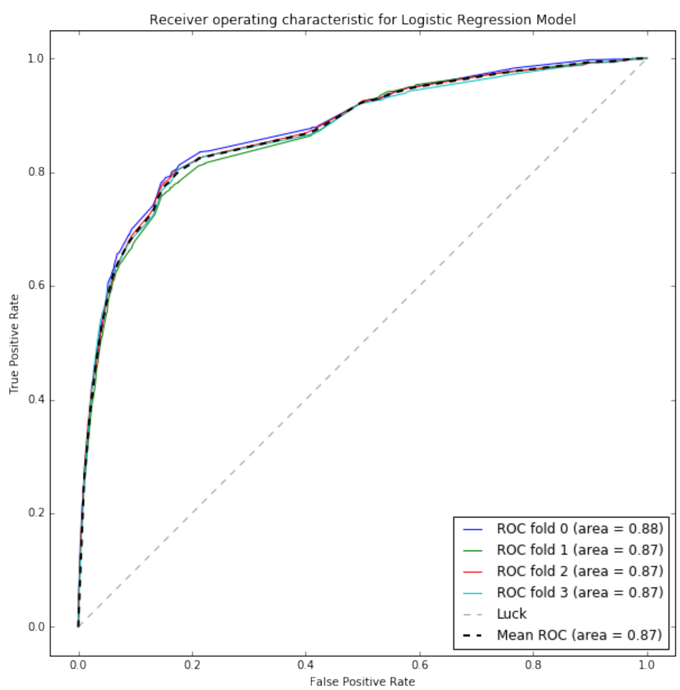

Figure 4.1 – Training ROC curve for the 4 Fold training

.

Following idea in [11], we also proceeded to a bagging of logistic regression model which yielded a better mean accuracy 86,5% so 1% more than the simple logistic regression model and a slightly better mean AUC 0.875. The author proposed to use this method in order to reduce the variance though, in our case it was not the case if not little. So we used the simple model coefficient for the rest of the estimation. This score is very high considering how skewed our data is, and how much we did take into the modeling. [7] report 0.89 for AUC for 70000 user journey dataset with a 20% conversion rate.

In Table we report the final model coefficient.

Display	SEA	SEO	Type-in	Retargeting
0.014	1.052	1.001	0.13	-0.397

Tableau 4.1 – Logistic Regression fitted predictive coefficient

As mentioned before, even if coefficient seems to be easy to understand they are hardly translated to concrete action.
In table 4.2 we report the predicted probability of converting interacting only once with each channel. By example the value in the display column is the predicted probability of converting with an input having 1 on the display column and 0 everywhere else. Based on the attribution score, we can assign conversion to which channel, what is reported in the second row of the table

	Display	SEA	SEO	Type-in	Retargeting
Attribution estimate	0.134	0.305	0.294	0.148	0.0932
Attributed conversion	858	1953	1882	948	762

Tableau 4.2 – channel and conversion Attribution

We can see that more than the half of conversions have been assigned to search channel (SEA and SEO). Results corresponds to finding in [7] where search come first followed by display. The author did not consider Type-in as a channel they deduct it from the intercept term of the logistic regression.

Markov models results

In figure 4.2 we report the transition matrix for order 1 Markov model. This is the only one we will be reporting since for higher order, we will have more columns (combination of K, where K is the order of the Markov model) and the table become unreadable. We should notive that the sum of each line is equal to 1, since it cover all the probability of what could be the next state starting from the current state (the left column of the table). The first column outline, the probability of having a conversion directly after passing by one channel. This is a rare pattern in our data though existing. We can still see small probability on the first column.

The second column reports the opposite, the probability of and end of journey without any conversion, the (null) state.

Some interesting pattern can also be extracted from this table, such that after retargeting the highest next state is type-in, which express the success of the action. The user came back to the website probably to a page he already visited before without passing by a search engine. Figure 4.3 show the difference between first order Markov model and last-touch, first-touch attribution models. The difference is huge for SEO channel which seems logic for the last touch model. Customer will end by searching his item before doing the conversion after having been checking it up few times before.

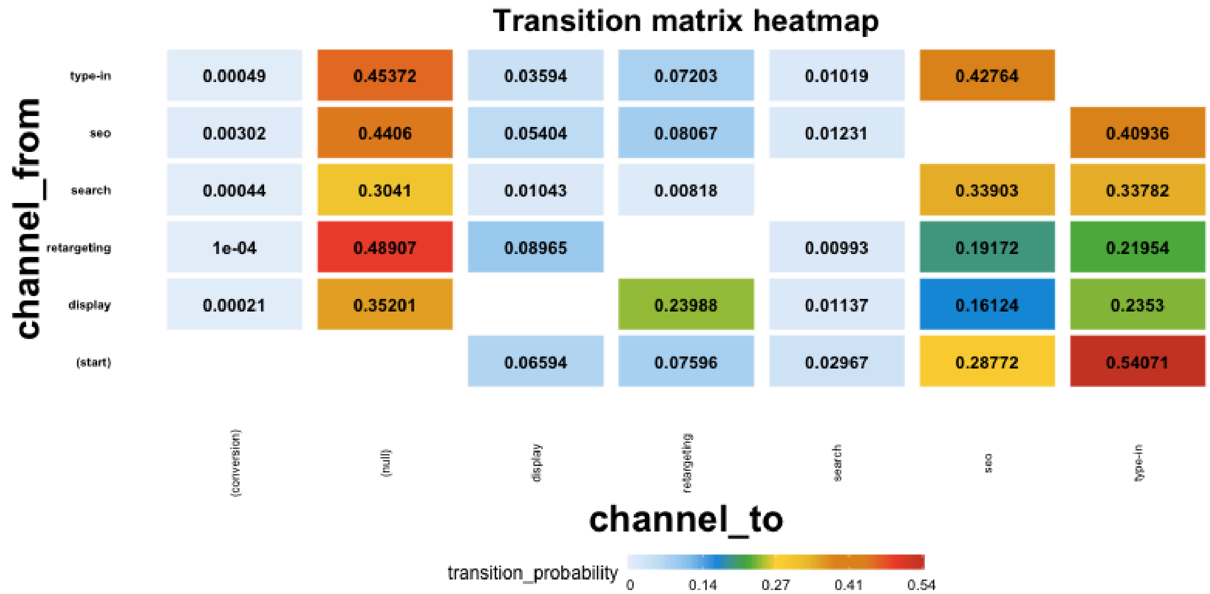

Figure 4.2 – Transition matrix estimation for order 1 markov model

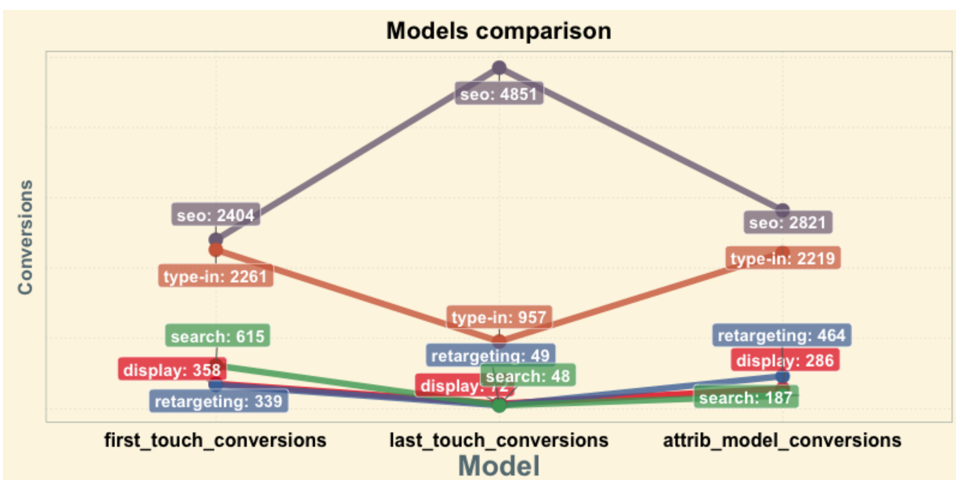

Figure 4.3 – Attribution result for order 1 markov model comparing with First touch and last touch model

What's really interesting here is the importance of retargeting when comparing with rule based models ? The dynamic attribution seems to catch more the its effect, nearly 10 times more attribution comparing to the pre-dominant technic : last touch-point model.

Surprisingly paid search do not have many conversion and most them are attributed to seo and type in. Going back to Figure 4.2 we can see that starting from search we usually ends up going to seo or type-in, which mean that we will not see sponsored link on the search engine or we will use our history to go back to the brand website. This could be explained by "exclusion" strategies adopted by the advertiser, a process permit saving money by not spending it to already interested user.

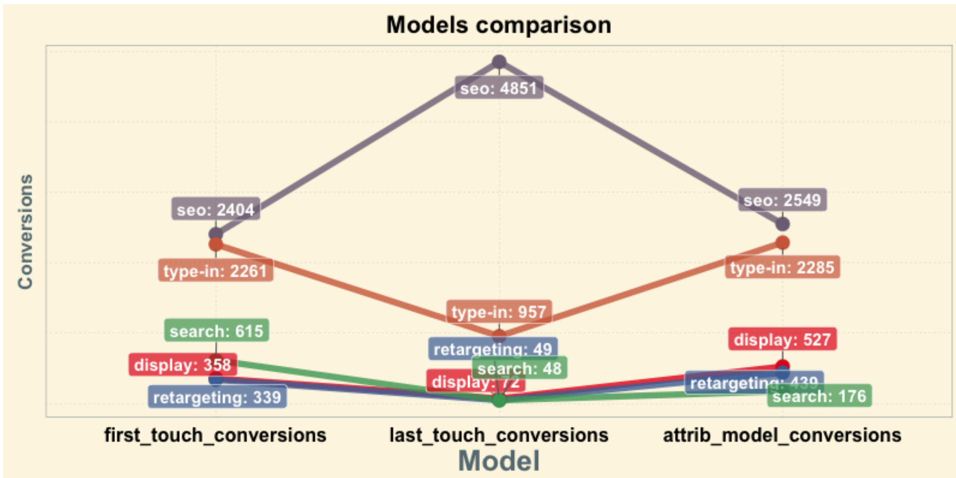

Figure 4.4 – Attribution result for order 3 markov model comparing with First touch and last touch model

Figure 4.4 shows attribution for 3 confirm the importance of retargetting and display and attribute less conversion to paid search while keeping the same level for search organic and type-in.

This leads us to think about fund relocation from search to display and retargeting, and that channel aggregation in model such logistic regression could be far from perfect.

At Artefact, we are formalizing this project finding to present it to the advertiser in-order to tune fund allocation while monitoring conversion to validate our approach. This could start with small campaign before scaling to the whole brand strategy. But still important enough to prove our point.

CHAPITRE 5

Conclusion

Throughout this report, we described my graduation internship experience at Artefact. First, we introduced the company and an overview about the project. Then we presented a review of state of the art around attribution models and their importance for marketing business units lately. After that, we detailed the data consisting of more than 2 millions user journey and the tools and algorithmic approach we have used to accomplish the attribution analysis. Before conclusion we report obtained results and discuss some of the output of the models. Attribution is at its early age is far more complex than we have seen in this first of its kind project at Artefact. As for the perspectives I suggest the following :

- ✓ Try to analyze more data going back further in historical data, and compare attribution trend over time
- ✓ Use DMP audience description to have attribution result by audience. i.e. the difference between client and non client or using available demographic data
- ✓ Analyze specific campaign, or fully digital campaign in order to be sure that no offline conversion will be happening
- ✓ Improve data acquisition to receive more channels on the DMP logs and thus have more robust attribution

This internship was a very beneficial experience to me, I'm now enlightened with new skills in the digital marketing field, I sharpened my big data processing by improving SPARK model and I also learned how to be a part of small organization and have an added value to the business.

Références

[1] Digital Ad Spending Surges to Record High as Mobile and Social Grow More Than 50%, source = http://adage.com/article/digital/iab-digital-advertising-generated-60-billion-2016/303650/, note = Accessed : 2016-08-16.

[2] Vibhanshu Abhishek, Peter Fader, and Kartik Hosanagar. Media exposure through the funnel : A model of multi-stage attribution. *Available at SSRN : http ://ssrn.com/abstract=2158421 or http ://dx.doi.org/10.2139/ssrn.2158421*, 12.

[3] Mirrokni-V. S. Muthukrishnan S Archak, N. Mining advertiser-specific user behavior using adfactors. *In M. Rappa, & P. Jones (Eds.), Proceedings of the 19th International Conference on World Wide Web (pp. 31–40). ACM.*, 10.

[4] J. Chandler-Pepelnjak. Measuring roi beyond the last ad. *Retrieved from http :* $//advertising.microsoft.com/espana/WWDocs/User/eses/Advertise/engagement_mapping.pdf$, 08.

[5] F. V. Wangenheim E. Anderl, I. Becker and J. H. Schumann. Putting attribution to work : A graph-based framework for attribution modeling in managerial practice. *23 October 2013 [Online]. Available : http* $: //papers.ssrn.com/sol3/papers.cfm?abstract_id = 2343077$, 13.

[6] F. V. Wangenheim E. Anderl, I. Becker and J. H. Schumann. Mapping the customer journey : Lessons learned from graph-based online attribution modeling. *International Journal of Research in Marketing*, 16.

[7] Rentola Olli. Analyses of online advertising performance using attribution modeling. Master's thesis, Aalto University, 14.

[8] D. L. Hoffman P. Chatterjee and T. P. Novak. Modeling the clickstream : Implications for web-based advertisin efforts. *Marketin Science vol. 22, nr. 4,pp. 520-541*, 03.

[9] K. Pauwels P. Kireyev and S. Gupta. Do display advertisements influence search ? attribution and dynamics in online advertising. *International Journal of Research in Marketing,*, 16.

[10] Khim Yong Goh Puneet Manchanda, Jean-Pierre Dubé and Pradeep K. Chintagunta. The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, 43(1) :98–108, 2006.

[11] X. Shao and L. Li. Data-driven multi-touch attribution models. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, 11.

View publication stats