

CLV - Attribution modeling

Mehdi Farhangian

12/21/2019

Attribution modelling is a must-have technique for marketers, as it enables them to determine the contribution of their marketing channels in their sales. In the conventional attribution model, all customers' paths are treated equally. However, as we explained here, they do not help customers develop smart marketing strategies. In reality, we have different segments of customers and a marketing channel might have a different meaning for customers, with different histories of purchases.

In order to demonstrate how attribution modelling can be developed and how it can be customised by considering the lifetime value of customers, we used a practice-oriented attribution framework based on something known as Markovian graph-based data mining techniques. This is an approach that enables us to measure the removal effect of all the touchpoints in the customers' journey - in order to determine the contribution of these touchpoints compared to the revenue generated. These touch points can be pages, events, campaigns, channels and so on. In this article, we focus only on one of these touchpoints, the Marketing channel.

Features involved in Customer Lifetime Value (CLV) or Lifetime Value (LTV) such as frequency of purchase, recency and monetary value can play a crucial role in the way that we determine the importance of a channel. In this article, we only focus on the frequency of purchase and how it can be taken into account in attribution modelling. All the customers' paths are not equal and we believe that the first purchase of a customer should be analysed differently from their next purchases. A customer in different stages of their life cycle have different perceptions of a website and the influence of channels are not equal in their different journeys.

In this article, we argue that in order to implement an effective attribution model that leads to a smart and data-driven marketing strategy, the life cycle of customers should be taken into account. If we can determine the effect of each marketing channel on the customers in different stages of their life cycle, we can develop smart marketing strategies. These strategies can have a huge effect on the customers' retention. According to statistics, a 2 percent increase in customer retention has the same effect on profits as cutting costs by 10 percent.

##Markov Chain and Attribution Modelling

Markov calculates the transition probabilities between channels and the importance of a channel can be defined as the change in conversion rate if the channel is dropped from the graph. The details of this approach are explained in this article.

We selected Markov approach for a variety of reasons. Compared to similar models, Markov provides better understanding and provides transferable results into managerial decisions. Unlike Google's Data-Driven Attribution modelling, the Markov approach is not a black box and the results can be repeated. Unlike Heuristic approaches (such as Last Click, First Click, Linear Model) that analyse one path at a time, the Markov model analyses relationships between different paths to understanding the role of channels in conversion.

In this article we explain the usage of Markov graph in attribution modelling, and how this approach can be customised by incorporating all the customers' life cycles.

As an example, let's say that we have four unique Marketing Channels which are M1, M2, M3 and M4. In our model, each customer journey represents a chain in Markov graph. So, we need to define states for each graph. In this case, we define three states:

- Start : Represents the starting point of the customers' journeys
- Conversion : Represents successful transactions
- Null : Represents unsuccessful transactions

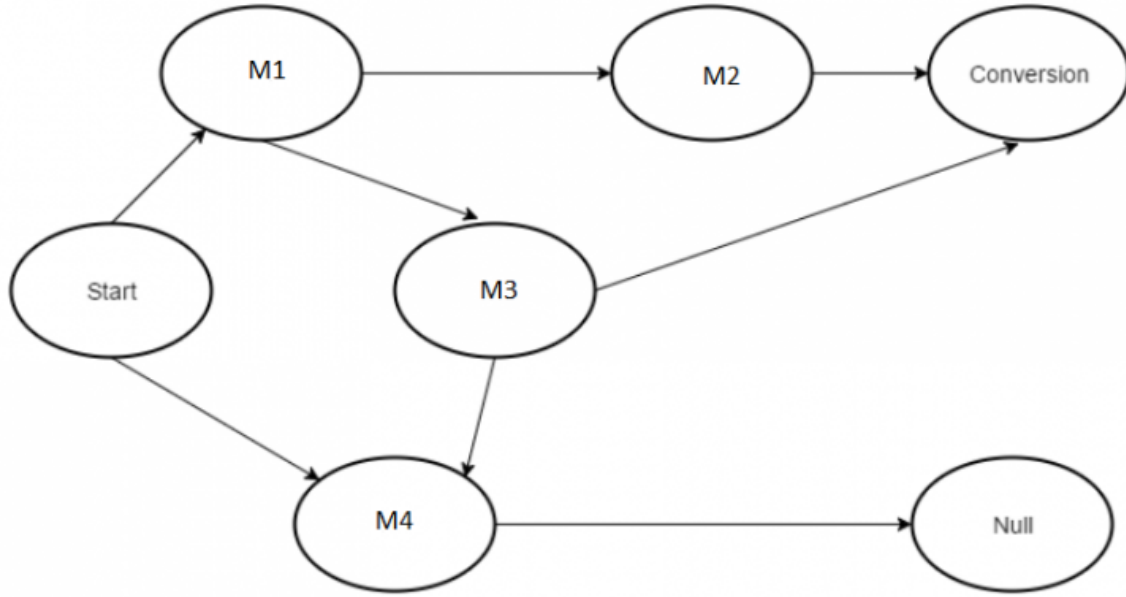


Figure 1: A caption

Let's say we have the following sequences of channels or customers' journeys:

How do I take the life time value of customers into account, in attribution modelling? image_1

So, in total, we have four customer journeys:

M1 > M2 > transaction. M1 > M3 > transaction. M1 > M3 > M4 > unsuccessful transaction. M4 > unsuccessful transaction. The following table shows these sequence with extra states that are split into pairs:

Based on the graph and this information, we can calculate the transition probabilities between states as follows:

After calculating the transition probabilities the graph should show the following:

In this approach in order to know the importance of each channel, we measure how many conversion could have happened, if we remove one of the channels. We called this the removal effect.

Let's continue this example by removing M3.

If we remove M3 we have to omit some of the pairs, that contained this node. As a result we have:

Probability of conversion with M3 is $0.5 * 0.5 = 0.25$ Probability of conversion without M3 is $0.5 * 0.5 = 0.25$ It means the removal effect of M3 is $0.25 / 0.5 = 0.5$ Mathematically speaking, we can define the transition probability as follows:

Simply put this means the transaction probability which is W_{ij} , is equal to the probability of the previous state (given the current state). With one condition: The transition probability is between 0 and 1, and the sum of transition probabilities equals 1. As a result if we didn't have M3 in the customers' journey, we have lost 50% of the transactions. In order to calculate the total number of conversion attributed to a channel, we can use the following formula:

Number of Conversions Attributed to a Channel = Total Number of Conversions * Removal Effect

Customer Journey	Extra States
M1 > M2	Start >M1 > M2 > conversion
M1> M3	Start >M1> M3 > conversion
M1> M3>M4	Start >M1> M3>M4 > Null
M4	Start> M4 >Null

Figure 2: A caption

From	To	Probability	Percentage
start	M1	1/2	50%
start	M4	1/2	50%
M1	M2	1/2	50%
M1	M3	1/2	50%
M2	Conversion	1/1	100%
M3	Conversion	1/2	50%
M3	M4	1/2	50%
M4	Null	1/1	100%

Figure 3: A caption

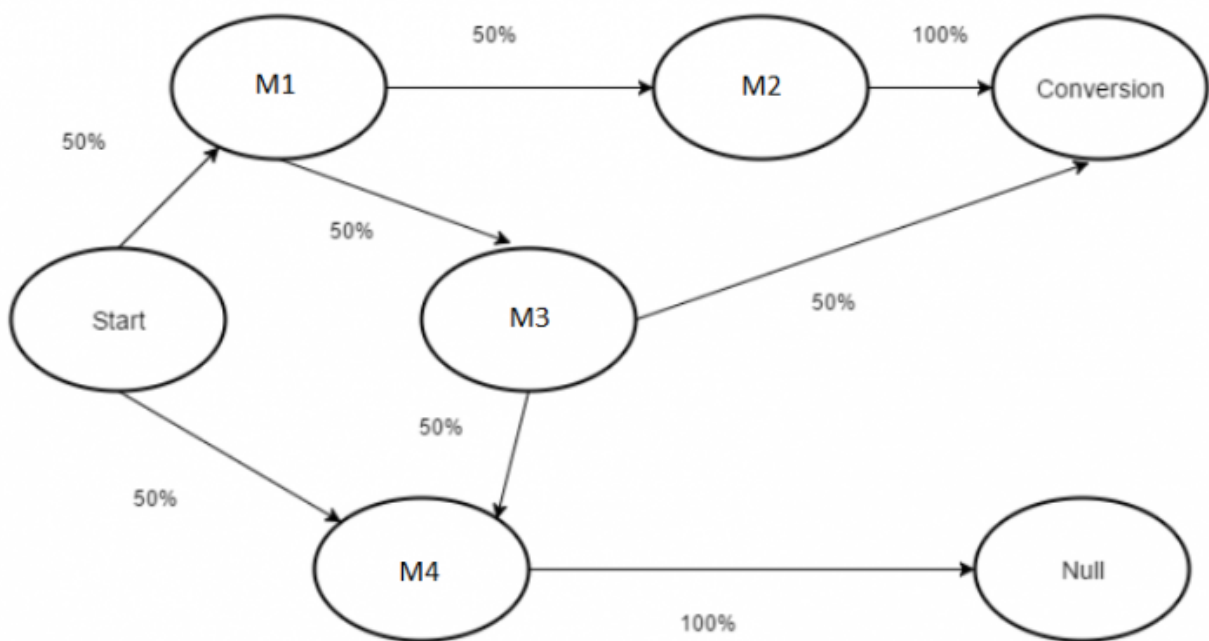


Figure 4: A caption

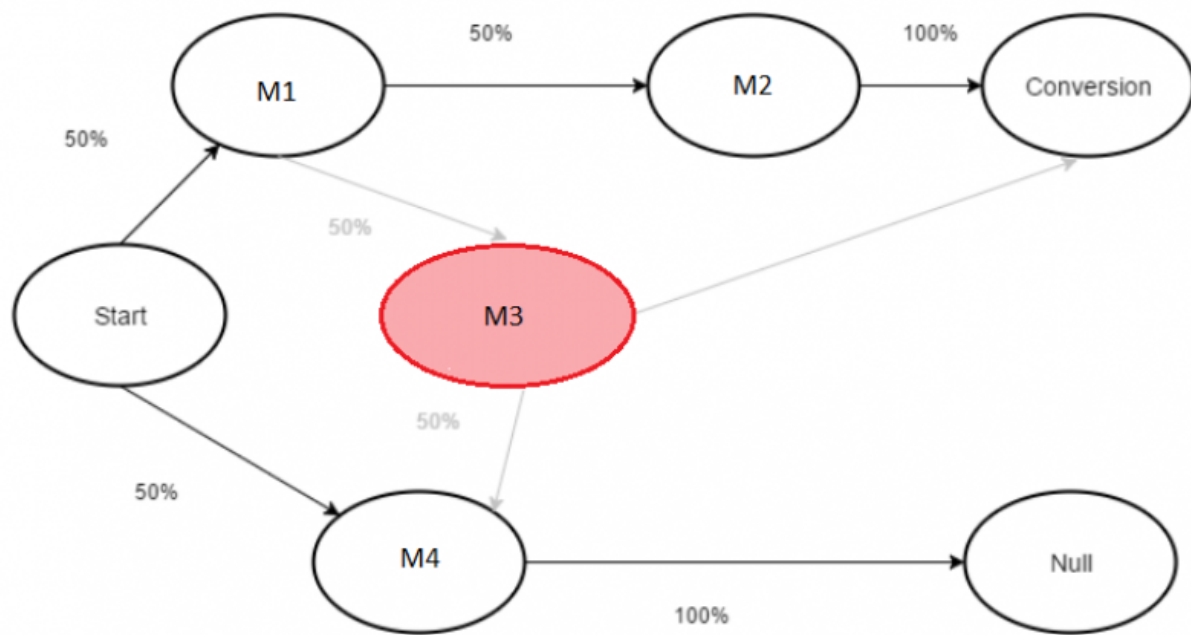


Figure 5: A caption

Customer	Journey
C1	M1 -> M2 -> Purchase, M3 -> M1-> M4 -> Purchase, M2 -> M1 -> no purchase
C2	M2 -> M4 -> M4 -> No Purchase
C3	M3 -> M1 -> Purchase

Figure 6: A caption

Customer	Journey
C1	M1 -> M2 -> Purchase
C1	M3 -> M1-> M4 -> Purchase
C1	M2 -> M1 -> no purchase
C2	M2 -> M4 -> M4 -> No Purchase
C3	M3 -> M1 -> Purchase

Figure 7: A caption

Considering the Frequency of Purchase in Attribution Modeling The above example shows how attribution modelling can be calculated using a Markov approach. However in the real world, not all customers are one-time buyers. Both data-driven and heuristic attribution models ignore this fact, and measure the credit of marketing channels regardless the frequency of purchases. Lets consider a more realistic example, we have four marketing channels (M1, M2, M3 and M4), and three customers which are (C1, C2, C3), and the customers' journeys are listed as follows:

The first customers had three journeys which are M1 -> M2 -> Purchase, M3 -> M1-> M4 -> Purchase and M2 -> M1 -> No Purchase. The common approach in attribution modelling is separating all the journeys as follows:

How_do_I_take_the_life_time_value_of_customers_into_account,_in_attribution_modelling?_image_8

The problem with this approach is that the lifecycle of customers is ignored. For example, the first two rows in the above dataset should not have the same meaning since the customers who already have a purchase experience do not respond the same as new customers to marketing channels.

In this article we introduce a framework based on a Markov-chain concept, that can be used for attribution modelling and consider the role of customers lifecycle. It should be noted that the mathematical formulas and codes used are out of the scope of this article. Instead, we will introduce a comprehensive framework that helps you to solve your attribution modelling problem. It's good to note that the implementation of

this framework is not complicated for someone with a basic understanding of R programming language and the Markov chain. In the next few articles, we will provide more details about the implementation of this framework.

Compute the customer journey duration based on how many purchases are done by customers previously In order to calculate the customers' journeys duration, we can look at the distribution of durations, from their first interaction with the website to their purchase, and select 95% of occurrences. We should calculate this duration for each group of customers based on their frequency of purchase. For example we might find out that 95% of purchases occurred after 50 days of the first interaction, and then we can select only 50 days and remove the previous days. For the second-time buyers we select all the customers who had more than one purchases, and select 95% of occurrences of the second purchase. For example we might find out that 95% of customers, had a second purchase after 60 days from their first interaction. Continuing this process we are then able to find out the customer journey duration for then-time buyers.

Splitting paths based on how many purchases are done by customers previously In this instace we split our dataset into different categories. The first dataset contains the journey of customers with one purchase and the nth dataset contains the journey of customers that happened after their n-1th purchases. In the above example, M1 -> M2 -> Purchase, goes to the first dataset that includes all the paths in the first lifecycle of customers, and M3 -> M1-> M4 -> Purchase, is selected for the second dataset and so on.

Replacing some channels One important step that is often neglected is cleaning data. Often we are not interested in giving credits to some channels such as direct channels or channels with no value. In these cases, we need to remove these channels or replace them with the previous channel.

Higher order of Markov Chains A simple way of attribution modelling based on Markov Chain concept, is the first-order or "memory-free" Markov graph. This is where the probability of reaching one state, depends on only the previous state. We can improve the accuracy of our attribution model by using a higher order of Markov chain. Once this is done we can then compute the transition probabilities, based on the previous two, three or more channels. To change the order of the Marcov chains, we can use ChannelAttribution Package.

After splitting our datasets to the n dataset (based on the number of purchases that are done previously), we can conduct attribution modelling on each dataset. The importance of a channel might be different for the one-time customers, as compared to second-time customers and n-time custom. Different strategies are required based on the frequency of purchase for each customer.

So, what can we say about attribution modelling and the lifetime value of our customers? In this article, we discussed that the lifetime value of a customer, should influence the way that we measure the attribution of a channel. We also discussed that attribution modelling should be customised for each segment of customers. When it came to different features involved in the lifetime value of customers, we focused on the frequency of purchase. We also demonstrated how a conventional Markov-based approach, measures the attribution of marketing channel. As well as this we introduced a framework and guideline to incorporate the frequency of purchases in attribution modelling. In the next articles, we will cover more features involved in the customers lifetime value and go into the detail of implementing of these customised attribution modelling.