**B2B Attribution Modeling**

**Introduction:**

Marketing attribution is a system of determining which marketing touchpoints lead to a conversion, and subsequently assigning a specific percentage of attribution to each contributing touchpoint. B2B Attribution Modeling gives the contribution of each customer touchpoint to the overall revenue. It helps cut down on unnecessary marketing channels and helps us focus on what is actually working. A typical B2B sales cycle involves a very high number of touchpoints and the entire process from prospecting to conversion spans many months.

B2B marketers usually go for a multi-touch attribution model because it involves multipletouchpoints, multiple teams, and multiple stakeholders, making it harder and complicated to attribute credit to the right touchpoint.

Involvement of both online & offline channels. So Omni channel Attribution is required i.e. integration of:

* Web analytics data (for online interactions)
* CRM data (for offline interactions such as event attendance)

**Why do we do this?**

* Leverage Attribution Modeling to optimize marketing strategy and gain competitive advantage.
* Gain insights into customer decision cycles and accordingly align your marketing strategy.
* Identify efficient conversion paths by analyzing different customer journeys and comparing key metrics such as the number of leads generated over time and the cost per lead.
* Identify and scale down on poorly performing marketing channels for bottom-line savings.
* Measure the effectiveness of different campaigns (within each channel) and suitably decipher better performing campaign strategy.

The age of data has given marketers more options than ever. Marketers find themselves overwhelmed with data, trying to figure out what content and campaigns actually work, and which marketing channels should get the credit.

**Challenges:**

* The length of the conversion funnel
* The number of interactions and
* The combination of online and offline touchpoints
* Capturing the data points accurately and extracting patterns from them is one of the biggest tasks of attribution modeling.
* The other factor is that interactions could be online as well as offline.(siloed data)

The fig-1 shows a B2B customer journey comprising of various stages in marketing and sales.

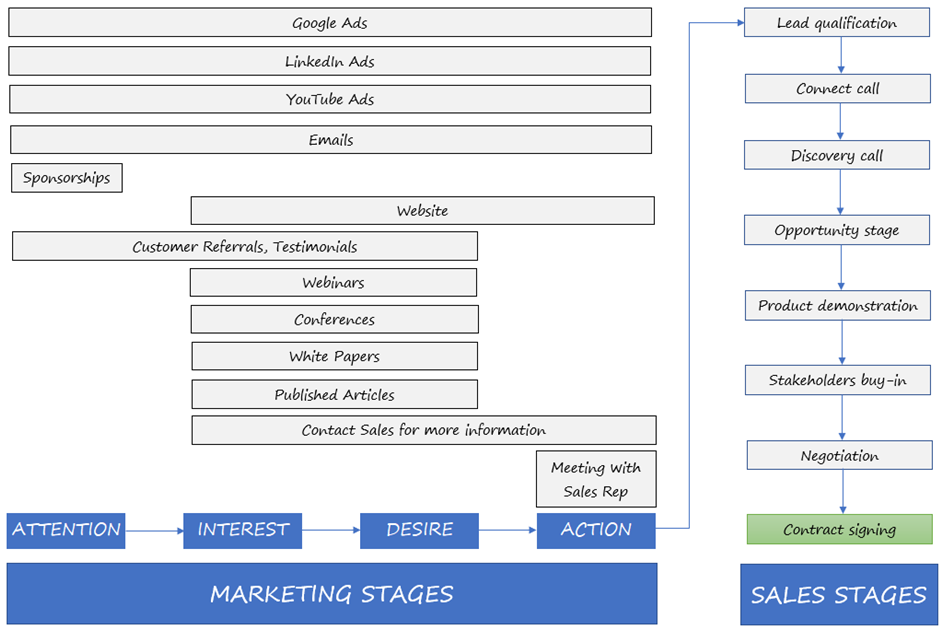


Fig. 1

The fig-2 shows a Multi-touch attribution through the lens of basketball. Scoring a point is a team effort. Similarly, each marketing touchpoint has certain amount of credit towards a customer’s (prospect’s) conversion.

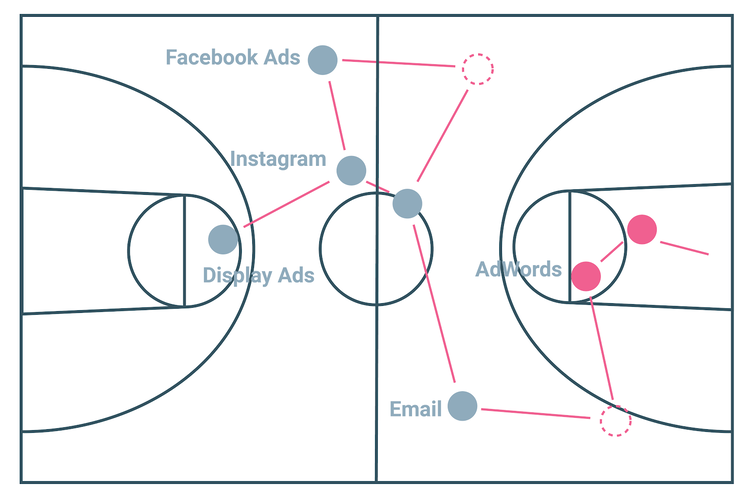


Fig.2

**Data:**

The data consists of **619,117** marketing touch-points from July (2018) till, comprising **130,008** unique customers among which 32,644 are qualified leads and generated **3297** conversions. Each row represents a prospect with the details like unique id of the prospect and the time of the interaction along with touchpoints.

The features in the data are:

* **Opportunity\_ID:** Represents the individual prospect.
* **Timestamp:** Date & Time of interaction with touchpoint.
* **Opportunity\_stage:** marketing or sales channel **(**touchpoint).

The Marketing channels in the data are:

|  |  |
| --- | --- |
| **Google ad** | Display\_ad\_campaign |
| Product |
| Webinar/Conference |
| **Email newsletter** | Marketing |
| Product |
| Support |
| **YouTube** | Announcements |
| Conference\_related |
| Product\_Support |
| **Website** | Corporate\_page |
| Product\_page |
| Support\_pages |

|  |  |
| --- | --- |
| **LinkedIn ad** | Content\_page |
| linkedin post |
| Product |
| Webinar/Conference |
| **Published article** | Financial\_times |
| Forbes\_magazine |
| Wallstreet\_journal |
| **Conference** | Booth |
| Conventions |
| **Customer referral** | |

The Sales channels in the data are: Connect call, Discovery call, Demonstration, Stakeholder’s buy in, Negotiation, Contract signed.

The top 8 unique journeys/paths followed by prospects are:

|  |
| --- |
| Email newsletter(Marketing) -> LinkedIn ad(linkedin post) -> YouTube(Product\_support) -> Website(Product\_page) |
| Google ad(Display\_ad\_campaign) -> YouTube(Product\_support) -> Website(Product\_page) |
| Google ad(Display\_ad\_campaign) -> Email newsletter(Marketing) -> Website(Product\_page) |
| Google ad(Display\_ad\_campaign) -> Email newsletter(Product) -> YouTube(Product\_support) |
| Google ad(Display\_ad\_campaign) -> Email newsletter(Product) -> LinkedIn ad(linkedin post) |
| Google ad(Display\_ad\_campaign) -> LinkedIn ad(linkedin post) -> YouTube(Product\_support) |
| Email newsletter(Product) -> Email newsletter(Product) -> YouTube(Product\_support) |
| Google ad(Display\_ad\_campaign) -> Website(Product\_page) -> LinkedIn ad(linkedin post) |
| Email newsletter(Product) -> YouTube(Product\_support) -> LinkedIn ad(linkedin post) |
| Google ad(Display\_ad\_campaign) -> YouTube(Product\_support) -> Email newsletter(Product) |

The data is skewed i.e. 80% of the data (prospects) is (are) from 8 (top) unique journeys followed by prospects.

**Modeling:**

**Markov Chain Model** (Used for attributions of each touchpoint)**:**

Markov Chains deal with a sequence of possible events in which the probability of the next event depends only on the current state of the event. Markov Processes are essentially random processes that satisfy “Markov property”. A random process (aka stochastic process) is a collection of random events whose outcomes are denoted by a set of random variables. (In B2B attribution scenario, the path taken by the buyer is considered to be a stochastic/random process.)

**How does it work?**

Markov Chains lets us model the attribution problem statistically as users making a journey from each state (channel) to finally reach a state of conversion and allows us to identify the probabilities of transitioning from one channel to another. These transition probabilities help us identify the statistical impact each channel has on overall conversions.

The components of the Markov Chain model are:

1. State space (the list of marketing channels)
2. Transition matrix  (probability of transitions between all the states of state space)
3. Initial distribution across the state space

A state-space is a collection of all the possible states the process (application) can take. The system is modelled as a sequence of interconnected states and there exists a probability of the system transitioning from one state to another. This is called the state transition probability. A square matrix that contains these transition probabilities between all possible states is called the transition matrix.

This model is one of the most-widely used ones for attribution modeling because of its scalability – it is very efficient when being applied to both large as well as small datasets. It also requires very less computing power. It has two flaws; first, they assume that the next state depends only on the current state and the second; they assume that each event in the sequence is uniformly spaced.

By integrating CRM data and click stream data (for offline and online interactions respectively), we first draw a Markov graph which contains all the different paths taken by all customers. (A sample Markov graph is shown in Fig.3.) Also, these data will give us the number of customers entering and exiting each stage of the sales process. Using these numbers, we can calculate the state transition probabilities. From the transition probabilities obtained as described above, we can calculate the removal effect of each channel i.e. the difference in the probability of conversion when that channel is removed from the Markov Chain.

The Removal Effect is the conversion contribution of a channel in a particular path. When this measure is aggregated across all paths, we get the overall conversion contribution of each channel.

The share of conversions is attributed accordingly to the corresponding channel. By calculating the number of conversions associated with each channel we can identify which channel(s) is/are responsible for conversions.

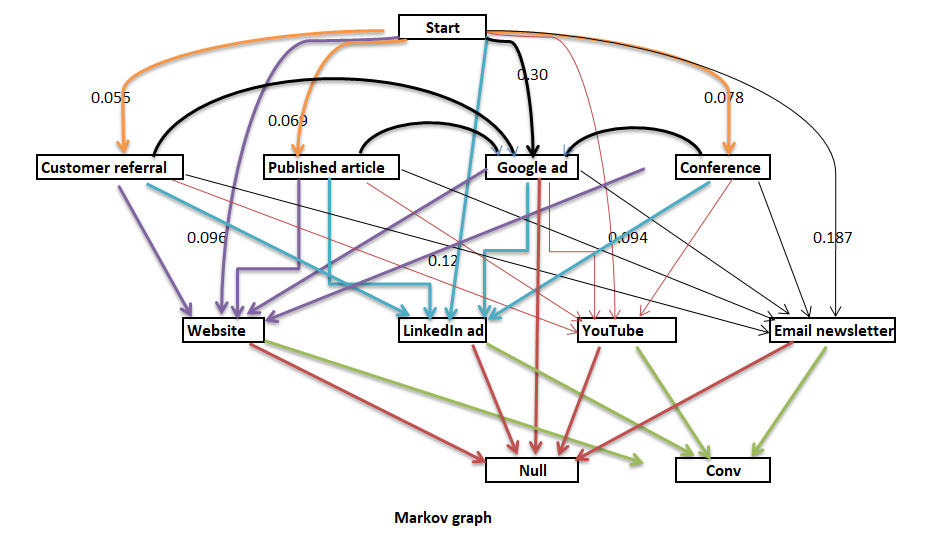


Fig.3

The biggest advantage of applying a data-driven model for B2B attribution is the quantification that it provides. We can view ad expenditure in the context of lead generation and conversions of different marketing channels. Stakeholders can analyze customer journeys, identify the relative importance of each path and build meaningful customer experiences. By gaining insight into customer behaviour and significant touchpoints, they can make better investment decisions.

Attributions obtained using Markov Chain model are:

|  |  |  |
| --- | --- | --- |
| **Channel** | **Credit** | **Percentage** |
| Website(Product\_page) | 784.06 | 24% |
| YouTube(Product\_support) | 781.89 | 24% |
| Email newsletter(Support) | 571.69 | 17% |
| LinkedIn ad(Content\_page) | 534.94 | 16% |
| Google ad(Display\_ad) | 207.91 | 6% |
| Email newsletter(Product) | 142.71 | 4% |
| LinkedIn ad(linkedin post) | 106.60 | 3% |
| LinkedIn ad(Webinar/Conference) | 44.32 | 1% |
| Email newsletter(Marketing) | 27.73 | 1% |
| LinkedIn ad(Product) | 20.06 | 1% |
| YouTube(Announcements) | 11.20 | 0% |
| YouTube(Conference\_related) | 10.15 | 0% |
| Google ad(Webinar/Conference) | 9.57 | 0% |
| Published article(Wallstreet\_journal) | 7.26 | 0% |
| Website(Corporate\_page) | 6.83 | 0% |
| Website(Support\_pages) | 6.31 | 0% |
| Conference(Conventions) | 6.09 | 0% |
| Website(Webinar/Conference\_pages) | 5.86 | 0% |
| Customer referral | 4.25 | 0% |
| Google ad(Product) | 3.83 | 0% |
| Published article(Forbes\_magazine) | 1.68 | 0% |
| Published article(Financial\_times) | 1.15 | 0% |
| Conference(Seminars) | 0.90 | 0% |

**Viterbi Algorithm:**

This algorithm is used to obtain top 10 journey sequence of events which are actually working.

The Viterbi algorithm is **a dynamic programming algorithm for obtaining the maximum a posteriori probability estimate of the most likely sequence of hidden states**—called the Viterbi path—that results in a sequence of observed events, especially in the context of Markov information sources and Hidden Markov models.

The purpose of the Viterbi algorithm is to make an inference based on a trained model and some observed data. It reduces the number of computations by storing the calculations that are repeated.

Mathematically,

A path**X = (x1, x2 …xT)** is generated which basically is a sequence of states **x ∈ S = {s1, s2 …sK}**. This generates the observation **Y = (y1, y2… yT)**with**y ∈ O = {o1, o2….oN}**. Here, N is the possible number of observations in the observation space **O**.

**Inputs:**

* Observation space **O = {o1,o2,…oN}**
* [State space](https://en.wikipedia.org/wiki/State_space) S = **{s1,s2,….,sk}**
* An array consisting of initial probabilities**π = (π1,π2….πk)**where πi stores the probability **x1 = si**
* Sequence of observations **Y = (y1, y2, …., yT)**
* [Transition matrix](https://en.wikipedia.org/wiki/Stochastic_matrix)**A**of size**K x K** where**A[i,j]** stores the [transition probability](https://en.wikipedia.org/wiki/Transition_probability) of transiting from state **Si**to**Sj**
* [Emission matrix](https://en.wikipedia.org/wiki/Hidden_Markov_model) **B**of size **K x N** where **B[i,j]**stores the probability of observing**Oj** from state **Si**

**Output:**

**The most likely hidden state sequence X=(x1, x2…xj)**

**The top 10 journey sequences of events obtained by Viterbi algorithm are:**

|  |  |
| --- | --- |
| **Sequence of events** | **Percentage/Probability** |
| Website(Product\_page) -> Email newsletter(Product) | 0.14% |
| Email newsletter(Product) -> Email newsletter(Product) | 0.04% |
| LinkedIn ad(linkedin post) -> Email newsletter(Product) | 0.03% |
| YouTube(Product\_support) -> Email newsletter(Product) | 0.02% |
| YouTube(Conference\_related) -> Email newsletter(Product) | 0.00% |
| Website(Webinar/Conference\_pages) -> LinkedIn ad(Content\_page) | 0.00% |
| YouTube(Announcements) -> Website(Corporate\_page) | 0.00% |
| Website(Support\_pages) -> YouTube(Conference\_related) | 0.00% |
| Website(Corporate\_page) -> Website(Corporate\_page) | 0.00% |
| LinkedIn ad(Content\_page) -> Website(Webinar/Conference\_pages) | 0.00% |

**Long-Short Term Memory Sequence Model:**

This model is also used to obtain the top 10 journey sequence of events with high probability. By representing each journey as a sequence of events in time, we are able to leverage sequence-specific modeling techniques. These techniques improve accuracy of future predictions by incorporating both information within and the order of a long history of prior events.

In sequence prediction challenges, Long Short Term Memory (LSTM) networks are a type of Recurrent Neural Network (RNN) that can learn order dependence. The output of the previous step is used as input in the current step in RNN. They are effective models for any type of sequential data where future events depend on past events.

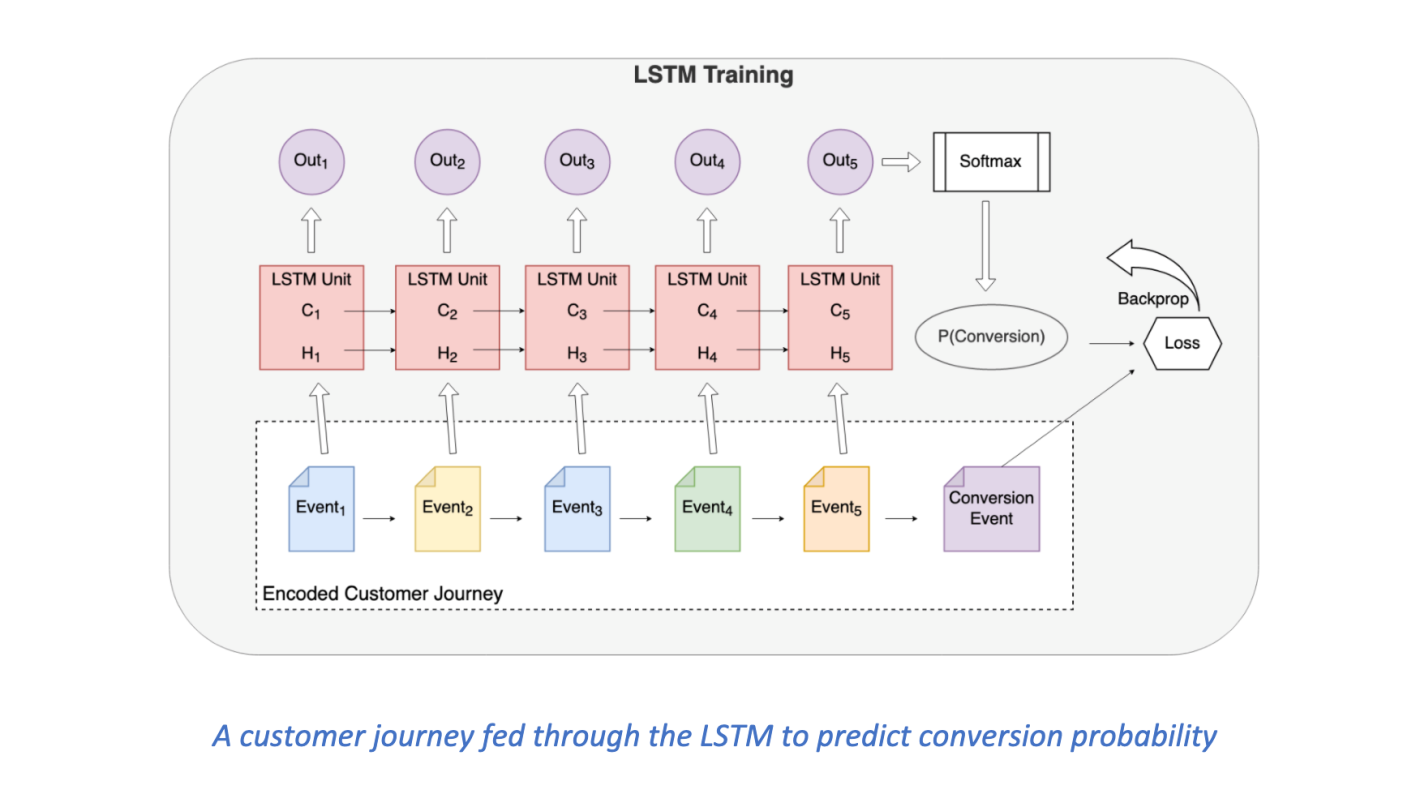


Fig.4

An improvement over Markov chains are Recurrent Neural Networks (RNNs). Each node in an RNN has an internal hidden state. The node takes a vector as input, combines it with the hidden state, and produces a prediction and a new hidden state. The prediction is a probability distribution over the states; concrete predictions are obtained by sampling from the distribution. RNNs improve over Markov chains by adding short-term memory: the hidden state depends on recently-seen inputs. The issue of RNN long-term dependency, in which the RNN is unable to predict words stored in long-term memory but can make more accurate predictions based on current data. RNN does not provide an efficient performance as the number of states rises. The LSTM may keep information for a long time by default.

Customer journeys can consist of dozens of events spread across several days, so a short memory is a serious limitation. To overcome the short memory problem, we use a variant of RNNs called Long Short-Term Memory networks (LSTMs).  LSTMs are composed of a series of nodes, each equipped with an input gate, output gate, and forget gate.  At a high level, these gates mimic the write, read, and reset operations of computer memory. The effect is dramatic: LSTMs can learn to bridge events that are dozens or even hundreds of steps apart. This makes LSTMs ideal for modeling customer journeys.

Using the data with qualified leads,

|  |  |
| --- | --- |
| **Training data** | **26115** |
| **Test data** | **6529** |
| Total | 32644 |

**The top 10 journey sequences of events obtained by LSTM model are:**

|  |  |
| --- | --- |
| **Sequence of events** | **Percentage** |
| Email newsletter(Marketing) -> LinkedIn ad(linkedin post) -> YouTube(Product\_support) -> Website(Product\_page) | 99.54% |
| Google ad(Display\_ad\_campaign) -> Email newsletter(Marketing) -> Website(Product\_page) | 99.23% |
| Email newsletter(Marketing) -> LinkedIn ad(linkedin post) -> Email newsletter(Marketing) | 97.14% |
| Email newsletter(Product) -> Email newsletter(Product) -> LinkedIn ad(linkedin post) | 80.92% |
| Email newsletter(Marketing) -> LinkedIn ad(linkedin post) -> Email newsletter(Marketing) -> Website(Support\_pages) | 79.29% |
| Google ad(Display\_ad\_campaign) -> Email newsletter(Product) -> LinkedIn ad(linkedin post) | 75.34% |
| Email newsletter(Product) -> LinkedIn ad(linkedin post) -> Website(Product\_page) | 74.15% |
| Email newsletter(Product) -> Email newsletter(Product) -> Email newsletter(Product) -> LinkedIn ad(linkedin post) | 72.12% |
| Email newsletter(Product) -> Website(Product\_page) -> LinkedIn ad(linkedin post) | 68.57% |
| Google ad(Display\_ad\_campaign) -> LinkedIn ad(linkedin post) -> Email newsletter(Product) | 68.00% |

**Results:**

|  |  |
| --- | --- |
| Training accuracy | 95.77% |
| Test accuracy | 95.40% |

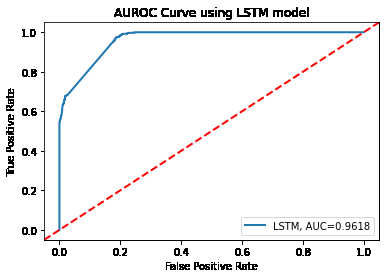


Fig.5

Fig.5 shows the AUROC curve for LSTM model, and the area under curve (AUC) is 0.9618. So this model makes better prediction whether a prospect may convert or not by analyzing the prospect’s journey.

**References:**

1. <https://www.linkedin.com/pulse/predicting-next-steps-customers-journey-using-deep-danny-portman-phd/>
2. <http://digital-thinking.de/deep-learning-clickstream-based-user-intent-prediction-with-anns/>
3. <https://www.pythonpool.com/viterbi-algorithm-python/>
4. <https://www.latentview.com/blog/markov-chains-what-are-they-and-where-do-they-matter/>
5. <https://dreamdata.io/revenue-attribution-models>
6. <https://medium.com/machine-learning-for-business-problem/how-to-implement-an-attention-rnn-into-solving-the-multi-channel-attribution-problem-6fa90d935859>