Digital Marketing Attribution

How marketers can identify which channels work

- Last Interaction
- Last Non-Direct Click
- First Interaction
- Linear
- Time Decay
- Position-based

PerceptiveAnalytics
Marketing Analytics | Tableau Consulting | Excel Apps
New York • Dallas • San Francisco • Hyderabad
Perceptive-Analytics.com
(646) 583 0001 cs@perceptive-analytics.com

Overview

Consumers are exposed to ads across channels in their purchase journey. A sale is a consequence of a sequence of ads and other marketing material shown to a consumer. For a marketer, this raises the question of attribution: Which ad gets credit and how much credit does each of the ads get in the conversion path? This analysis is critical in closing the feedback loop to facilitate analysis and optimization.

Introduction

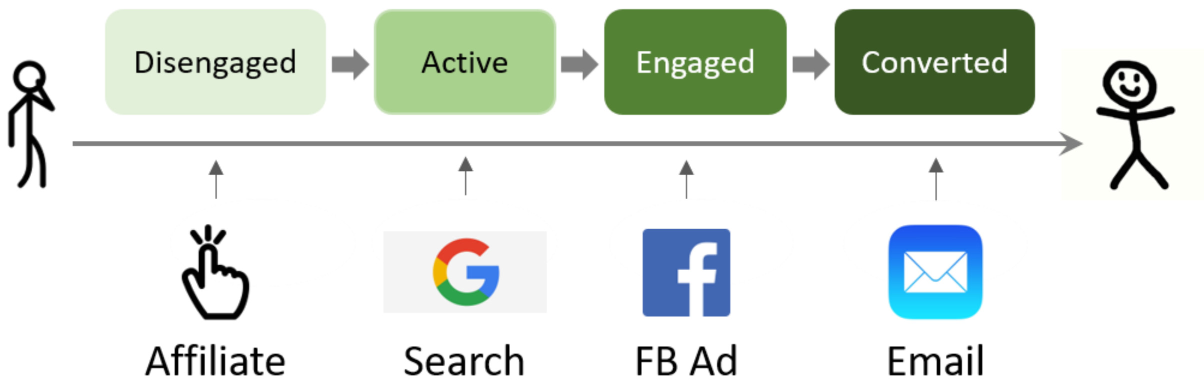
Advertisers want to know how ads (across channels) affect consumer's choice and to what level. A marketer must measure the impact of each ad channel and use this data to allocate ad budget optimally.

The challenge of assigning credit to each channel means you have to quantify the impact of each channel that led to customer acquisition. Attribution models help to identify the set of events and user actions along the customer's journey that play a role in conversion, and quantify the impact.

Attribution Models

Attribution models attempt to define how each touch point contributes to the customer conversion. A customer may get exposed to a product through an affiliate link, then may get interest and search online. You may place a pixel at this time and follow it up with a facebook followed up by an email leading to a conversion.

CUSTOMER JOURNEY



To figure out the impact, there are 2 broad categories of attribution models: 1. Heuristic based models 2. Algorithmic models. An empirical analysis is performed using clickstream data collected across multiple websites using cookies.

1. Ad attribution models using heuristic rules

A heuristic is a simplification of problem, eliminating complexity in favor of a swift analysis. In the case of multi-touch attribution modeling, this implies assigning values across channels in the sequence, regardless of actual impact the channels had on the completion of a sale. (Bill, 2012)

Last Interaction The last touchpoint receives 100% of the credit for the sale.

Last Non-Direct Click All direct traffic is ignored, and 100% of the credit for the sale goes to the last channel that the customer clicked through from before converting.

Last Google Ads Click The last Google Ad click receives 100% of the credit for the sale.

First Interaction The first touchpoint receives 100% of the credit for the sale.

Linear Each touchpoint in the conversion path shares equal credit for the sale.

Time Decay The touchpoints closest in time to the sale or conversion get most of the credit. Previous channels receive significantly less credit.

Position-based 40% credit is assigned to each the first and last interaction, and the remaining 20% credit is distributed evenly to the middle interactions.

2. Algorithmic Attribution Models

Algorithmic attribution analyzes available data to unravel the actual impact of each touch point on conversion. Instead of applying a short cut, algorithmic attribution builds custom model and assigns weights to each touch point based on user dynamics. (Yiyi Li, 2017)

2.1. **Simple Probabilistic** attribution model (Li, 2011): Computes the empirical conversion probability of each channel and calculates the probability of conversion. For a given data set, compute the empirical probability of each channel

$$P(y|x_i) = \frac{N_{positive}(x_i)}{N_{positive}(x_i) + N_{negative}(x_i)} \qquad (1)$$

Compute pair-wise conditional probabilities

$$P(y|x_i, x_j) = \frac{N_{positive}(x_i, x_j)}{N_{positive}(x_i, x_j) + N_{negative}(x_i, x_j)}, \quad (2)$$

Then compute contribution of channel at each user level conversion

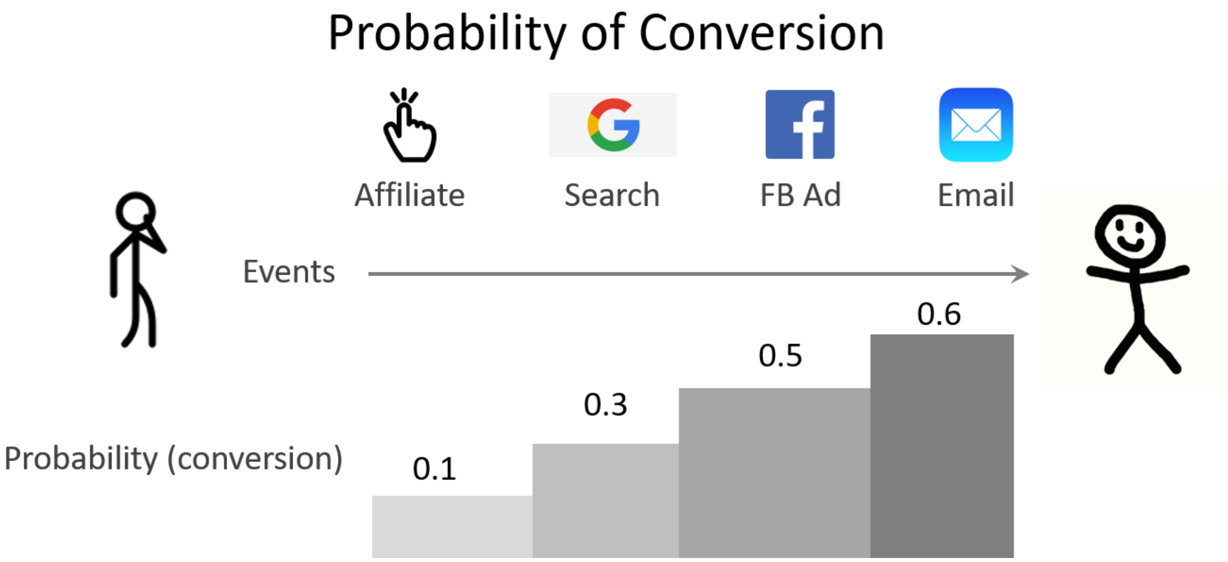
$$\begin{aligned} C(x_i) \quad &= \quad p(y|x_i) + \frac{1}{2N_{j \neq i}} \sum_{j \neq i} \Big\{ p(y|x_i, x_j) \\ &\qquad - p(y|x_i) - p(y|x_j) \Big\}, \qquad (3) \end{aligned}$$

2.2. **Bagged Logistic Regression:** (Li, 2011)

Bagging, Bootstrap Aggregation, iteratively subsamples data. Bagging generates multiple versions of a logistic classifier and uses these to get an aggregated predictor. The aggregation does a plurality vote to predict conversions. Bagging reduces variance and give significant gain in accuracy.

Step 1. For a given data set, sample a proportion [ps] of all the sample observations and a proportion [pc] of all the covariates. Fit a logistic regression model on the sampled covariates and the sampled data. Record estimated coefficients.

Step 2. Repeat Step 1 for M iterations, and the final coefficient estimate for each covariate is taken as the average of estimated coefficients in M iterations.



2.3. Higher Order Markov Chain Model (HMM): (Abhishek, 2014)

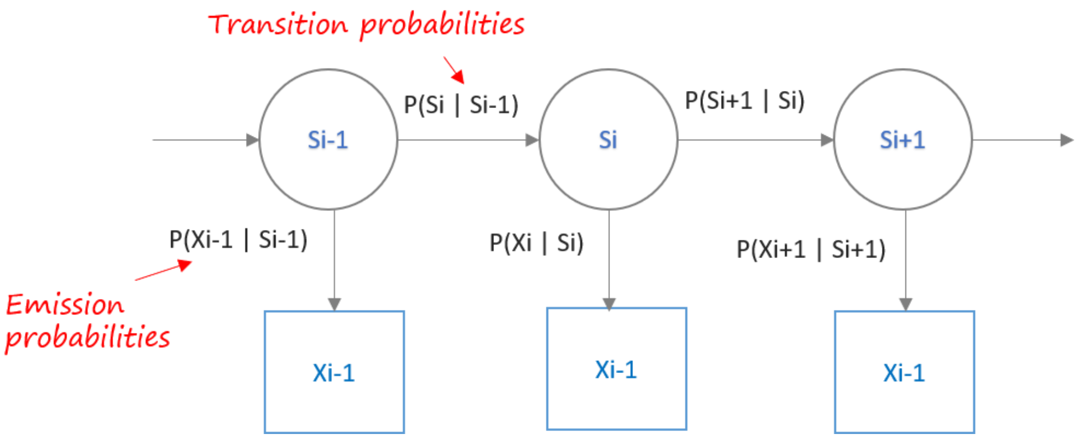
HMM captures consumers' behavior as they transition across the different states of the funnel and eventually convert. An HMM is a probabilistic sequence model - given a series of events, a probability distribution over possible sequences is computed, and the best chain of events is chosen. A higher order Markov chain model computes probability based on not just the current state but across several ordered states and assigns weighted probabilities to each combination. It provides us with the number of conversions that can be attributed to each touchpoint, as well as the value of each touchpoint. HMM is a function of (Probability Vector, Transition Matrix, Emission matrix)

Yit = f(Qit, (Mit, Nit), Pi)

Qit = transition probabilities between different states & are based on a logistic function of the number and type of ads encountered by a consumer

Eit = emission probabilities are based on a Poisson process and a Binomial model for pageviews (Nit) and conversion (Mit) that are also affected by which ads were seen.

Pi = initial state distribution (sum across all ad events)



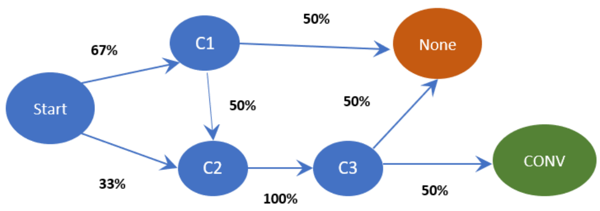
An m−th order Markov chain is defined as a stochastic process {Xt, t = 1,2, · · ·} on the state space: {1,2, · · ·, N} with the property that the next transition only depends on the last m steps. Formally,

Pr (Xt = i | Xt−1 = it−1, · · ·, X1 = i1) = Pr (Xt = i | Xt−1 = it−1, · · ·, Xt−m = it−m)

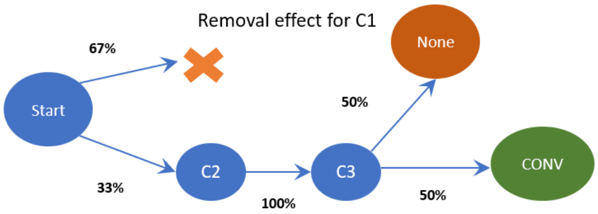
The last step is to estimate every channel/touchpoint. This can be calculated using the principle of Removal effect.

Illustration of first order Markov model

Path	Splits
(START) > C1 -> C2 -> C3 > Conv	(start) -> C1, C1 -> C2, C2 -> C3, C3 -> (conversion)
(START) > C1 -> None	(start) -> C1, C1 -> (none)
(START) > C2 > C3 > None	(start) -> C2, C2 -> C3, C3 -> (none)



From	To	Probability	Total probability
(start)	C1	2/3	67%
(start)	C1		
(start)	C2	1/3	33%
Total from (start)		3	
C1	C2	1/2	50%
C1	(none)	1/2	50%
Total from C1		2	
C2	C3	1	100%
C2	C3	1	
Total from C2		2	
C3	(conversion)	1/2	50%
C3	(none)	1/2	50%
Total from C3		2	



Absence of channel C1 in customer journey would lead to 50% loss of conversions, channels C2 & C3 to 100%.

Hence, attribution to each channel:
C1: 0.5 / (0.5 + 1 + 1) = 0.2 * 1 conversion = 0.2
C2: 1 / (0.5 + 1 + 1) = 0.4 * 1 conversion = 0.4
C3: 1 / (0.5 + 1 + 1) = 0.4 * 1 conversion = 0.4

A third-order HMM model can be used for standard attribution real time analyses (Yiyi Li, 2017).

The application of models with higher orders enables a more detailed understanding of the interplay across channels and increases the predictive performance.

2.4. **Shapley Value Model:** (K. Zhao, 2014)

The Shapley value is a cooperative game theory solution concept that assigns value among players in a cooperative game. A cooperative game is defined by a characteristic function x (q1, . . ., qM) that assigns for each coalition of players and their contribution qi the value they created.

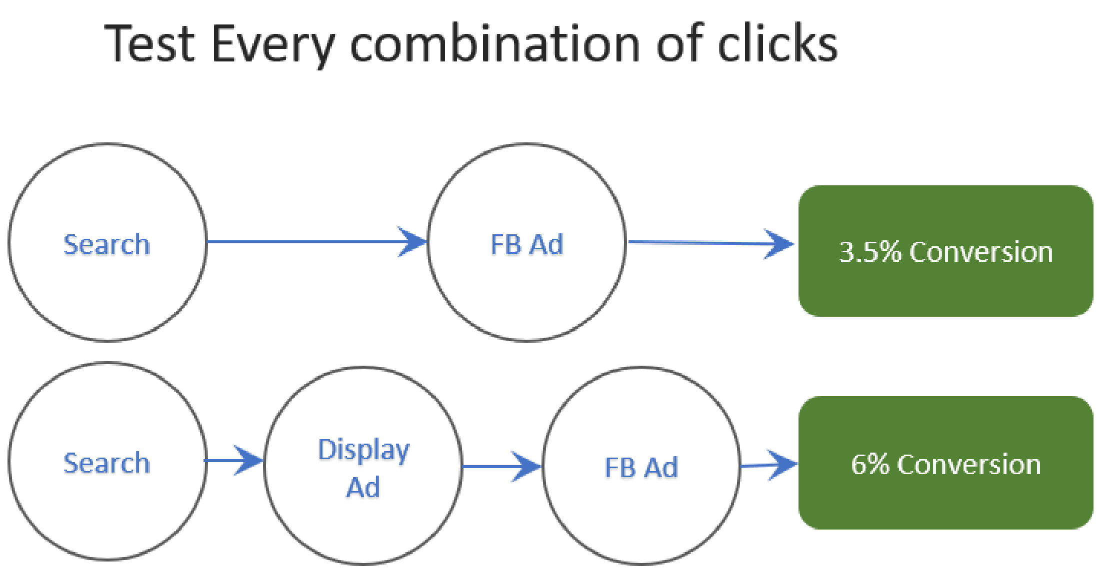
Shapley value considers an ad campaign as a cooperative game, and ad channels as players in the game. The channels work cooperatively to entice, affect and convert users. The Shapley value of each ad channel is computed based on its impact on the advertising KPI, which may comprise individual impact and interaction with the rest.

For a set of M channels, the Shapley value is defined as following:

$$\phi_i(x) = \sum_{S\subset(M\setminus i)} \frac{|S|!(|M|-|S|-1)!}{|M|!}(x_{S\cup i} - x_S)$$

where M is the set of channels and x is the set of conversion rates for different subsets of channels.

For example, for channel $x1$ in a campaign with P= {$x1$, $x2$, $x3$}, the Shapley value is, $\phi 1=R(x1)+1/R(x1,x2)+1/2R(x1,x3)+1/3R(x1,x2\ ,x3)$.



Shapley value method takes the weighted average of its marginal contribution over all possible coalitions for each channel.

Conclusion

Advertising attribution is one of the biggest challenges of the online advertising. Algorithmic attribution models have several advantages over heuristic techniques.

1. The models allow the advertisers to estimate the incremental impact of every ad that was shown to the consumer at an individual level.
2. The models allow the advertisers to discern the underlying latent state of the consumer. The advertiser can thus use this information to optimally choose the subsequent advertising activity.

A better attribution methodology allows better publishers to receive due credit, thereby increasing the efficiency of the advertising market.

References

1. Abhishek, V. F. (2014). Media Exposure through the Funnel: A Model of Multi-Stage Attribution.
2. Bill, K. (2012). Attribution Playbook Google Analytics.
3. Eva Anderl*, I. B. (2014). Mapping the customer journey: A graph-based framework for online attribution modeling.
4. K. Zhao, S. M. (2014). Shapley Value Methods for Attribution Modeling in Online Advertising.
5. Li, X. S. (2011). Data-Driven Multi-Touch Attribution Models.
6. P. Kireyev, K. P. (2013). Do Display Advertisements Influence Search? Attribution and Dynamics in Online Advertising.
7. Yiyi Li, Y. X. (2017). Modeling Multi-Channel Advertising Attribution Across Competitors.

About Perceptive Analytics

Perceptive Analytics is a Data Analytics Company recognized as a Top 10 Emerging Analytics Company. It is the winner, Fidelity Data Challenge in which 54 analytics companies participated. It also received an award at Netflix Hackathon at Tableau Conference, 2018. The clients we served include Morgan Stanley, Johnson & Johnson, Amex, Wells Fargo, and PepsiCo to name a few. We work with clients as their analytics department or alongside internal teams to deliver long-term competitive advantage.

Winner, Fidelity Investments Data Challenge

Top 10 Emerging Analytics Companies to Watch

Advanced Analytics

- Customer Analytics
- Marketing mix modeling
- Attribution modeling
- Customer segmentation
- Response modeling
- Churn modeling
- Next basket analysis

Business Intelligence & Reporting

- Visualizations & Dashboards
- ETL (Extract, Transform and Load)
- Data warehousing
- Cloud computing (AWS, Azure, Google Cloud)

Schedule Free Consulting Call

Contact

Chaitanya Sagar, CEO
cs@perceptive-analytics.com
+1 (646) 583 0001