International Journal of Research in Marketing 33 (2016) 457–474

Contents lists available at ScienceDirect

IJRM

International Journal of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

Mapping the customer journey: Lessons learned from graph-based online attribution modeling

Eva Anderl [a,*], Ingo Becker [b], Florian von Wangenheim [b], Jan Hendrik Schumann [a]

[a] Universität Passau, Innstr. 27, 94032 Passau, Germany
[b] ETH Zurich, Weinbergstr. 56/58, 8092 Zurich, Switzerland

ARTICLE INFO

Article history:
First received on April 24, 2014 and was under review for 12½ months
Available online 19 March 2016

Keywords:
Online advertising
Attribution
Marketing models
Markov models
Multichannel

ABSTRACT

Advertisers employ various channels to reach customers over the Internet, who often get in touch with multiple channels along their "customer journey." However, evaluating the degree to which each channel contributes to marketing success and the ways in which channels influence one another remains challenging. Although advanced attribution models have been introduced in academia and practice alike, generalizable insights on channel effectiveness in multichannel settings, and on the interplay of channels, are still lacking. In response, the authors introduce a novel attribution framework reflecting the sequential nature of customer paths as first- and higher-order Markov walks. Applying this framework to four large customer-level data sets from various industries, each entailing at least seven distinct online channels, allows for deriving empirical generalizations and industry-related insights. The results show substantial differences from currently applied heuristics such as last click attribution, confirming and refining previous research on singular data sets. Moreover, the authors identify idiosyncratic channel preferences (carryover) and interaction effects both within and across channel categories (spillover). In this way, the study can help advertisers develop integrated online marketing strategies.

1. Introduction

Online advertising is essential to the promotional mix of many industries (Raman, Mantrala, Sridhar, & Tang, 2012). Today, advertisers employ a variety of online marketing channels[1] to reach potential customers, including paid search and display marketing, as well as e-mail, retargeted displays, affiliates, price comparison, and social media advertising. At the same time, customers visit the advertisers' websites on their own initiative—for instance, by directly typing in the related web address. Using various channels, many customers visit company websites multiple times before concluding a purchase transaction (Li & Kannan, 2014). 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). Given the proliferation of online channels and the complexity of customer journeys,[2] measuring the degree to which each channel actually contributes to a company's success is demanding.

Despite the widespread and ongoing practice of many advertisers to apply comparatively simple heuristics (e.g., last click attribution), such that the value is attributed solely to the marketing channel directly preceding the conversion (The CMO Club, & Visual IQ,

* Corresponding author at: Universität Passau, Innstr. 27, 94032 Passau, Germany. Tel.: +49 851 509 2421; fax: +49 851 509 2422.
E-mail addresses: eva.anderl@uni-passau.de (E. Anderl), ibecker@student.ethz.ch (I. Becker), fwangenheim@ethz.ch (F. von Wangenheim), jan.schumann@uni-passau.de (J.H. Schumann).

[1] In this study, we use the term "online marketing channels" as an umbrella phrase referring to various online marketing instruments, including search engine advertising, display, or social media advertising.

[2] We define an online customer journey of an individual customer as including all touch points over all online marketing channels preceding a potential purchase decision that lead to a visit of an advertiser's website.

http://dx.doi.org/10.1016/j.ijresmar.2016.03.001
0167-8116/

Inc, 2014), this challenge of attributing credit to different channels (Neslin & Shankar, 2009) has recently begun to receive increased attention in academia and practice alike (Berman, 2015). Academics have proposed a variety of substantiated analytical attribution frameworks, including logistic regression models (Shao & Li, 2011), game theory-based approaches (Berman, 2015; Dalessandro, Perlich, Stitelman, & Provost, 2012), Bayesian models (Li & Kannan, 2014), mutually exciting point process models (Xu, Duan, & Whinston, 2014), VAR models (Kireyev, Pauwels, & Gupta, 2016), and hidden Markov models (Abhishek, Fader, & Hosanagar, 2015). Furthermore, several industry players such as Adometry (Google), Convertro (AOL), or VisualIQ have introduced a range of attribution methodologies (Moffett, 2014). Even though sophisticated attribution methods become accessible to a broader audience, and marketing executives call for such performance measures (Econsultancy, 2012), in practice the full browsing history of a user is rarely taken into account when calculating channel effectiveness in multichannel settings (Li & Kannan, 2014).

Based on a survey among marketers using advanced interactive attribution offerings, Osur (2012) reports that the two most widely applied objectives of attribution software are to "measure the value and performance of digital channels" and to "measure how one digital channel affects the performance of another [channel]" (p. 4). In this research, we address both topics, but we extend the issue further by trying to identify generalizable answers to these challenges. For advertisers, it is valuable to know what insights from attribution apply in a company-specific context and what insights may be generalized across companies (industries). Such insights can help to better explain factual channel effectiveness and can also shed light on the interplay of channels in multi-touch environments. Marketing scholars specifically call for further research using customer-level path data across several firms and industries to determine spillover effects in a more generalizable way (Li & Kannan, 2014). Empirical generalizations are important for both theory generation and evaluation and can provide valuable guidance to managers (Kamakura, Kopalle, & Lehmann, 2014). For instance, if advertisers anticipate idiosyncratic channel preferences among some users on their path to purchase, the advertisers could transfer this knowledge into a more adequate channel selection.

To achieve our research objectives, we needed to develop and apply an attribution mechanism with the capability of determining the effectiveness of individual marketing channels and deriving insights on the interplay of channels in a multichannel environment across very different data sets and conditions. In particular, we suggest an attribution framework based on Markovian graph-based data mining techniques, extending an approach originally developed in the context of search engine marketing (Archak, Mirrokni, & Muthukrishnan, 2010). We model individual-level multichannel customer journeys as first- and higher-order Markov graphs, using a property that we call removal effect to determine the contribution of online channels and channel sequences. The graph-based structure of our model reflects the sequential nature of customer journeys, enabling insights into the interplay of channels. Applying this framework to four large, real-world customer-level data sets from three different industries enables us to derive both cross-industry generalizations and industry-specific findings. In doing so, we make the following contributions:

First, we contribute novel insights into online marketing effectiveness of single channels within a multichannel setting. We estimate our graph-based framework on four data sets from different industries and compare the results against two well-known heuristic attribution techniques, namely, first and last click attribution, as well as two logit models. Prior research indicates that heuristic approaches of attributing conversion to the very last (or first) click can produce incorrect conclusions (Abhishek et al., 2015; Li & Kannan, 2014; Xu et al., 2014). A comparison of our results across four data sets enables us to confirm and refine these results and move toward empirical generalizations. We find that firm-initiated channels, where the advertiser initiates the marketing communication (Bowman & Narayandas, 2001; Wiesel, Pauwels, & Arts, 2011), are consistently undervalued by the heuristic attribution approaches. For customer-initiated channels, which are triggered by potential customers, on their own initiative (Wiesel et al., 2011), the contribution of paid search and direct type-ins is consistently overestimated by the last click approach. For other customer-initiated channels, additional factors such as industry and brand characteristics seem to play a role.

Second, higher-order models enable us to shed light on the interplay of channels in a multichannel setting. By comparing the results of our attribution framework across data sets, we generalize findings from prior literature indicating that a majority of channels exhibits idiosyncratic channel carryover (Li & Kannan, 2014). Furthermore, we observe spillover effects both within and between channel categories. Customer-initiated channels show substantial removal effects if they are followed by other customer-initiated channels, whereas spillover effects between firm-initiated channels are, by and large, negligible.

Third, we propose a novel variant that adds to existing advanced attribution modeling techniques (Abhishek et al., 2015; Berman, 2015; De Haan, Wiesel, & Pauwels, 2016; Kireyev et al., 2016; Li & Kannan, 2014; Xu et al., 2014) by representing customer path data as first- and higher-order Markov walks. This graph-based approach, adapted from research on paid search (Archak et al., 2010), represents a useful addition to the emerging attribution literature. Whereas first- and higher-order models offer support in measuring channel contribution in a multichannel setting, higher-order models, in particular, allow investigation of channel sequences and spillovers between channels.

Finally, our framework is beneficial in the approach to several explicit problems that online marketers confront. For instance, the framework may help to calibrate online channel budgets and move toward an improved budget allocation, even though endogeneity issues and different channel elasticities can make budget allocation difficult. Furthermore, using our framework, advertisers can more accurately calculate the conversion probability of a customer, given his or her previous customer journey. This information can be used to support state-of-the-art applications such as real-time bidding decisions in advertising exchanges.

2. Research background

Academic research on attribution and the interplay of online channels has only recently gained momentum. Jordan, Mahdian, Vassilvitskii, and Vee (2011) examine allocation decisions for publishers, using multiple attribution approaches, and derive optimal allocation and pricing rules for publishers who are selling advertising slots. In a study of the economic welfare consequences of the