10-9-2023

Context-Aware Marketing Attribution Based on Survival Analysis

Kilian Züllig
Ulm University, Germany, kilian.zuellig@uni-ulm.de

Stefan Napirata
Ulm University, Germany, stefan.napirata@uni-ulm.de

Steffen Zimmermann
Ulm University, Germany, steffen.zimmermann@uni-ulm.de

Context-Aware Marketing Attribution Based on Survival Analysis

Research Paper

Kilian Züllig, Stefan Napirata, and Steffen Zimmermann

Ulm University, Institute of Business Analytics, Ulm, Germany
{kilian.züllig,stefan.napirata,steffen.zimmermann}@uni-ulm.de

Abstract. Companies increasingly invest in digital marketing channels to promote their products and services. While the expenditures for each marketing channel are known, the contribution of marketing channels to a successful conversion, and therefore the value they generate, is unknown, but highly relevant for strategic decision-making. In this paper, we develop a novel, context-aware additive hazard marketing attribution (CAHMA) model based on survival analysis to address this problem. In addition to channel-specific, time-decaying effects of marketing on the users' conversion rate, we control for the effects of contextual features, such as the device or country from which users interact with marketing channels. Based on a prototypical implementation, we demonstrate the model's applicability and evaluate it on real-world data from the industry. We find that CAHMA outperforms other models in terms of accuracy while offering unique interpretability of the results and hence, providing deep insights for practitioners into the effects of marketing.
Keywords: Marketing Attribution, Online Advertising, Survival Analysis, Conversion Prediction

1 Introduction

Over the past decade, digital marketing has become the most important way for companies to reach potential new and existing customers to promote their products or services. Spending on digital marketing has more than doubled in the last 5 years and is expected to keep increasing in future (Statista, 2022*b*). Thereby, Big Data analytics enables businesses to provide customers with more targeted and personalized advertisements to motivate them to move along the purchase funnel toward the action desired by the advertising company, i.e., conversion (Kannan et al., 2016). Along their customer journey towards conversion, users typically have multiple interactions, i.e., touchpoints, with different marketing channels such as search engine advertising (SEA), display or social media advertising. While companies can easily quantify the exact amount spent for each marketing channel, their contribution to a successful conversion, and therefore the value generated from specific channels, is unknown. Marketing attribution is a strategy for determining the individual contribution of marketing channels to conversions and allocating the generated value accordingly. In order to measure and manage the efficiency of marketing and to improve its return on investment, marketing attribution has been

18th International Conference on Wirtschaftsinformatik
September 2023, Paderborn, Germany

recognized as a central task in digital marketing (Kannan et al., 2016), underlining the importance of accurate marketing attribution models. Yet, determining the accuracy of marketing attribution is not trivial, because there is no ground truth for the value generated by channels. Most likely, not even customers are aware of the exact influence that different marketing channels had on their own purchase decisions. In this vein, marketing attribution is a two-folded problem. First, an attribution strategy is necessary to allocate the value of conversions to marketing channels and second, a measurement of how well this allocation reflects reality is required. For the latter, it has become standard to assume that good marketing attribution models should also be good at predicting whether a customer journey leads to a conversion or not (Shao & Li, 2011). Aiming for the highest predictive accuracy, data-driven marketing attribution models tend to increase in complexity (e.g., Kumar et al., 2020). While predictive accuracy is indeed an important criterion for convincing managers of a model's credibility (Lodish, 2001), managers must understand how the results are generated in order to rely on them for strategic decisions (Little, 2004).

Existing literature proposes a variety of data-driven approaches (see e.g., Zaremba (2020) for an overview), with a recent trend towards neural-network-based models (e.g., Yang et al., 2020; Kumar et al., 2020; Yao et al., 2022). While these models may be best at predicting conversions, they strongly sacrifice on interpretability, especially regarding their black-box model structure. Consequently, practitioners are unlikely to rely on them for strategic decision-making, which is the ultimate purpose of attribution models (Romero Leguina et al., 2020).

On the other hand, there are advanced, data-driven models that also find application in the industry. Among these, Markov chains provide a comprehensible model structure and solid predictive accuracy (Anderl et al., 2013). While this seems promising, there are relevant aspects that cannot be modeled with Markov chains (Anderl et al., 2016). For example, the effect of a user's touchpoint with a marketing channel on the conversion probability obviously decreases over time, which cannot be accounted for by Markov chains. Models based on additive survival analysis (e.g., Zhang et al., 2014; Ji & Wang, 2017) also have a very intuitive and comprehensible model structure and, in addition, allow to incorporate such time-decaying effects. Even more, beyond time-dependencies, the environment and circumstances under which users interact with marketing channels (e.g., season or country) or, if available, user characteristics (e.g., users' demographics or preferences), might also have an impact on their conversion probabilities. This additional information can also be included in additive survival analysis models, using contextual features. Zhang et al. (2014), however, do not account for such contextual features. Ji & Wang (2017) significantly increase the complexity of their model when including contextual features, such that it loses interpretability, besides having some unresolved mathematical issues.[1]

In this paper, we propose a novel context-aware additive hazard marketing attribution model (CAHMA) based on survival analysis. The model assumes an intuitive, additive relationship between the effects of marketing channels and accounts for the influence of contextual features. Unlike Ji & Wang's (2017) model, we directly control for channel-

[1] For a more detailed discussion of the related literature see: `https://github.com/CAHMAchameleon/CAHMA`.

independent but context-dependent effects, similar to how control variables are used in regression models. This allows to better separate context-specific effects and identify the "true" channel-specific effect of marketing. Based on a prototypical implementation, we demonstrate the applicability of CAHMA and evaluate it on a real-world dataset from GetYourGuide, a leading international booking platform for travel products. The data was provided by the platform in search of a new marketing attribution model with high predictive accuracy under the condition of strong interpretability of both the model structure and results. We find that CAHMA outperforms other interpretable data-driven marketing attribution models. Considering highly skewed data for marketing attribution, CAHMA is particularly accurate in detecting the relevant minority class, i.e., converting customer journeys. At the same time, CAHMA provides unique insights into the channel-independent effects of contextual features in addition to the channel-specific influence strength and time-decaying effect of marketing.

2 Context-Aware Additive Hazard Marketing Attribution Model

Before we present our model and its evaluation, we introduce the concepts of survival analysis which is commonly used to build models for censored time-to-event data (Lawless, 2011). Let T be a non-negative random variable describing the waiting time until the occurrence of an event (here, conversion of a user). The main goal of our model is to estimate the probability distribution of T, which is the basis for our marketing attribution and conversion prediction. Using standard notations from probability theory, $F(t)$ denotes the cumulative distribution function and $f(t)$ the probability density function of T. The survival function $S(t)$ is the counter probability that the conversion does not occur until time t. The hazard rate $\lambda(t)$ is defined as the probability of conversion within the next, infinitesimally small moment after time t, given that the user has not converted until t. Applying basic calculus (see e.g., Lawless, 2011), one can derive the following two identities

$$\text{(i)}\ \ S(t) = e^{-\int_0^t \lambda(s)\, ds}, \qquad \text{(ii)}\ \ f(t) = S(t)\lambda(t).$$

As $S(t)$ and consequently $f(t)$ depend only on the hazard rate, the random variable T, and therefore the entire model, is completely defined by the hazard rate $\lambda(t)$. We detail later, how modeling the hazard rate is sufficient to determine the marketing attribution and conversion prediction.

2.1 CAHMA Model

Based on survival analysis, we introduce our novel CAHMA model for context-aware marketing attribution. We denote users (i.e., potential customers) by $u \in \{1, \ldots, U\}$ and marketing channels by $k \in \{1, \ldots, K\}$. A touchpoint corresponds to an interaction of a user u with a marketing channel k and is denoted by a tuple (a_l^u, t_l^u), where a_l^u denotes the marketing channel of the l-th interaction of user u and t_l^u denotes the timestamp of this interaction. The customer journey of a user u, consisting of l_u touchpoints, is defined as $J^u = \{(a_l^u, t_l^u)_{i=1}^{l_u}, Y^u, T^u, X^u\}$. At the end of a customer journey either a conversion or no conversion is observed, which is denoted by $Y^u \in \{0, 1\}$. Furthermore, every customer journey has a latest timestamp T^u, marking the end of the journey. If the customer journey leads to a conversion, i.e., $Y^u = 1$, then T^u is the time of conversion. If the customer journey does not lead to a conversion, i.e., $Y^u = 0$, then T^u corresponds to the end of the observation window. Additionally, we include available contextual features, e.g., the season or country in which the touchpoint was recorded. The vector X^u contains the contextual features observed during the customer journey of user u.

Similar to Zhang et al. (2014), we assume an additive effect between the channel-specific and time-dependent effect $\kappa_{a_l^u}(t)$ of each touchpoint. Extending the model of Zhang et al. (2014), we further account for the influence of contextual features in a time-independent baseline hazard $b(X)$. Therefore, the hazard rate for customer journey J^u at time t is given by

$$\lambda_u\left(t|J^u\right) = \begin{cases} b(X^u) + \sum_{t_l^u < t} \kappa_{a_l^u}(t), & \text{for } t \geq t_1^u \\ 0, & \text{for } t < t_1^u. \end{cases}$$

In line with Zhang et al. (2014) and Ji & Wang (2017), we assume a positive influence of each touchpoint on the hazard rate, which fades over time. Hence, we model the channel-specific effect of each touchpoint on the instantaneous conversion probability of user u at time t by the kernel function $\kappa_{a_l^u}(t) = \beta_{a_l^u}\omega_{a_l^u}e^{-\omega_{a_l^u}(t-t_l^u)}$. For each interaction with channel k, i.e., if $a_l^u = k$, β_k represents the channel-specific influence strength of the touchpoint on the hazard rate and ω_k is the corresponding speed of decay over time. $(t - t_l^u)$ denotes the elapsed time since the l-th interaction. All β_k's must be non-negative, as otherwise, the hazard rate could be negative, which is impossible by the definition of a probability and the assumptions of survival analysis. We further constrain all ω_k's to be non-negative as well, to maintain the assumption of time-decaying effects of marketing.

The time-independent baseline hazard, introduced in the proposed method, allows controlling for channel-independent effects on the conversion probability that are instead driven by contextual features of user u. The baseline hazard is given by $b(X^u) = e^{\alpha_0 + \sum_{i=1}^{I} \alpha_i X_i^u}$, where each X_i^u represents one out of I different contextual feature from the feature vector X^u. The parameters $\alpha_i \in \{1, \ldots, I\}$ estimate the effects of contextual features and are not restricted in their domain. The coefficient α_0 is a constant. While different functional forms of the baseline are possible, we use an exponential function, ensuring that the baseline is always greater than zero. Moreover, the differences in the effects of two features can be easily analyzed and interpreted. By modeling a positive, time-independent baseline hazard we establish a mathematically correct survival analysis model and estimate a true probability distribution of the time to conversion.

Figure 1 illustrates an exemplary hazard rate for one customer journey over time, as assumed in our model. The horizontal line represents the context-dependent baseline hazard which is constant for one customer journey. Each kernel function $\kappa_k(t)$ corresponds to one peak that fades over time. When the user interacts with one of the marketing channels, the hazard rate increases (c.f. vertical increase). The influence strength β_k of a channel corresponds to the area between the curve and the dotted line below and the time-decaying effect ω_k defines the slope of the curve. The concept of the baseline, the influence strength, and the time-decaying effect allow for a very descriptive and understandable model structure with interpretable model parameters and results. For the sake of interpretability, we calculate the half-life of a channel, which is the time until

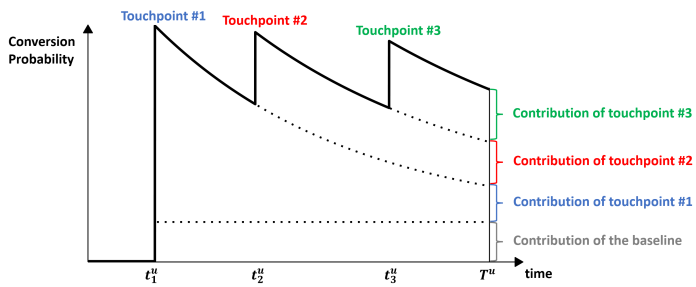

Figure 1. Hazard Rate.

the effect of a channel on the conversion rate is half its quantity. Using an exponential function, the half-life is constant over time, which facilitates interpretation.

The central objective of marketing attribution is to take a retrospective view on customer journeys to assign value to individual touchpoints or channels. However, there is no ground truth for the assigned values making it impossible to directly evaluate the quality or accuracy of the attribution. Thus, an alternative measure for the quality of marketing attribution models is required. It is widely assumed that a marketing attribution model, that accurately captures the effects of marketing, should also perform well in the task of conversion prediction. Consequently, predictive accuracy is commonly used to evaluate the performance of the models instead of directly evaluating the marketing attribution. In the following, we first derive the marketing attribution strategy based on CAHMA before explaining how the method can predict conversions.

Marketing Attribution Strategy. Through its additive structure, the proposed model allows calculating the contribution of each touchpoint (a_l^u, t_l^u) to the instantaneous conversion probability, i.e., the hazard rate, at any given time t from the value of the channel-specific kernel function $\kappa_k(t)$. In the case of a conversion, the contribution of a marketing channel is given by the relative contribution at conversion time T^u of all touchpoints in the customer journey that belong to that channel. Similar to previous models based on additive survival analysis (Zhang et al., 2014; Ji & Wang, 2017), that is

$$attr_k^u = \frac{\sum_{t_l^u < T^u, a_l^u = k} \kappa_k(T^u)}{\sum_{t_l^u < T^u} \kappa_k(T^u)}.$$

The baseline hazard $b(X^u)$ models the context-dependent but channel-independent effect of marketing, which is collectively triggered by all touchpoints in the customer journey. Thus, we ignore the contribution of the baseline in the attribution. This is equivalent to distributing the effect of the baseline over all channels in the customer journey according to their relative contribution to the channel-specific effects.

Conversion Prediction. The proposed model allows to directly derive the conversion probability of users from the hazard rate by integrating over the time after their last interaction $t_{l_u}^u$. Hence, the conversion probability within a time window $\tilde{t}$ is given by

$$P(T \leq t_{l_u}^u + \tilde{t}|J^u) = 1 - e^{-\tilde{t}b(X^u|J^u) - \sum_{l=1}^{l_u} \beta_{a_l^u}(e^{\omega_{a_l^u}\tilde{t}} - 1)e^{-\omega_{a_l^u}\left(t_{l_u}^u + \tilde{t} - t_{l_u}^u\right)}} .$$

It follows from this equation that for an infinitely large time window, i.e., $\tilde{t} \to \infty$, the baseline hazard will drive the survival function to zero and the probability of eventual conversion to 1. This property of survival analysis models ensures that the time to conversion T is a true random variable. In our practical context of marketing attribution, it can be argued that not every user will convert. However, our model only assumes that every user would theoretically convert within an infinite time window. The predicted conversion time may be too large to be observed if it exceeds the time during which a customer journey is trackable, or even the remaining lifetime of the user.

Parameter Estimation. We use maximum likelihood estimation to fit all parameters $\Theta = \{\alpha_0, \ldots, \alpha_I, \beta_1, \ldots, \beta_K, \omega_1, \ldots, \omega_K\}$ to the data. The objective log-likelihood function is given by

$$\log \mathcal{L}(\Theta) = \sum_{u \in U, Y^u = 0} \log\left(S(T^u|J^u)\right) + \sum_{u \in U, Y^u = 1} \log\left(S(T^u|J^u)\lambda(T^u|J^u)\right).$$

As there is an analytic solution for the gradients of our log-likelihood function, the optimization problem can be solved using any gradient optimization algorithm that allows for box constraints.[2]

2.2 Demonstration and Evaluation

Prototypical Implementation The prototype of CAHMA is implemented in the programming language R. Calculating the log-likelihood function and its gradients requires operations on a very large number of grouped subsets of the data (i.e., single customer journeys). Using the open source package "data.table" from R, our prototype can handle substantial marketing attribution datasets even on a local computer. However, for a productive analytics environment, nowadays many companies rely on modern cloud computing platforms (Statista, 2022*a*) such as Microsoft Azure or Amazon AWS that allow for efficient distributed computing over large clusters with analytics engines like Apache Spark or Hadoop. To be able to leverage the advantages of such platforms, we apply the MapReduce programming model, which is the foundation for Spark or Hadoop. MapReduce allows distributing big datasets over large clusters to perform computations efficiently in parallel (Dean & Ghemawat, 2008). Due to the additive structure of the log-likelihood function and its gradients, their calculation can be decomposed into many smaller subproblems. Following MapReduce, these subproblems are then distributed over a cluster of worker processes for parallel computation (map) before aggregating the results in one global solution (reduce). While MapReduce together with data.table already enables efficient computations in R, the developed prototype can be transferred

[2] For more details on the log-likelihood, its gradients and the optimization problem see `https://github.com/CAHMAchameleon/CAHMA`.

to Apache Spark to run even more efficiently on modern cloud computing platforms, further improving the practical applicability of CAHMA.

Dataset. To evaluate the prototype of CAHMA, we were provided with a real-world dataset by GetYourGuide. The dataset was provided in anonymized form. GetYourGuide invests heavily in its digital marketing campaigns to generate new and reattract previous customers. The dataset contains information on the interactions of potential customers with various marketing channels used by GetYourGuide over a nine-month time window from June through February.[3] In total, the dataset contains almost 33 Mio. touchpoints of 14.8 Mio. unique users[4] and 11 distinct channels. Each record of the dataset represents one touchpoint. Additionally, the dataset contains the exact transaction time and contextual features such as the device which was used for an interaction or the country in which it was recorded.

Assuming that reconversions of users can be influenced by their marketing touchpoints, which occurred before their previous transactions, we do not restart customer journeys after conversions. Instead, all previous touchpoints of users are added to their subsequent customer journeys which extend beyond the users' previous transactions. By this approach, we aim to avoid underestimating the contribution of early touchpoints for repeated conversions. Overall, we obtain 14,950,959 distinct customer journeys, of which significantly less than 5% ended in conversions. This highlights the strong imbalance of the data. We randomly split the dataset into disjoint training and evaluation sets, each containing 50% of the customer journeys. This allows us to assess out-of-sample performance on unseen data, which is unbiased by overfitting. To enrich customer journeys with contextual features, we use dummy-coded variables of the devices on which, the countries in which, and the seasons during which the customer journeys were tracked. The dataset contains more than 200 different countries, with some of them having even less than 10 occurrences. Hence, we code the ten most relevant countries in terms of their frequency and conversion rate individually and segment the remaining countries into nine geographic regions to limit the number of required dummy variables. This coding is solely based on information from the training data to avoid any data leakage to the evaluation dataset. The seasons are defined as the three-month time windows of summer (June - August), fall (September - November), and winter (December - February). Spring is omitted because our data contains no records for this time frame. For the devices, we distinguish between Android, iOS, other mobile devices, and desktop computers. Beyond the described data preparation steps and basic data cleaning, including the removal of duplicates and incomplete records, we do not further manipulate the dataset, e.g., to reduce the imbalance. This way we ensure to evaluate CAHMA on a realistic dataset available in practice.

Competing Models. Because of their practical relevance, we compare our proposed model to existing additive survival analysis models, which are based on the same

[3] For confidentiality reasons, we cannot provide details on commercially highly sensitive information such as the exact conversion rate or channel specifics.

[4] The dataset is limited to information associated with website visitors that have consented to the specific data collection according to data protection legislation.

assumptions and the standard Markov chain model by Anderl et al. (2013). All models are trained and evaluated on the same datasets. The competing models are

– **AdditiveHazard** (Zhang et al., 2014): This model based on survival analysis provides the basis for our proposed model, using an additive hazard rate but without considering contextual features.
– **AMTA** (Ji & Wang, 2017): This model is also based on survival analysis and provides an alternative approach that allows the inclusion of contextual features. The model considers contextual features by the users' intrinsic, time-independent conversion rate as an additional influence on conversion. The authors attempt to model the overall conversion probability as a joint distribution of the time-independent conversion probability and the marketing-induced conversion probability.
– **MarkovChain** (Anderl et al., 2013): This model is the current standard at GetYourGuide and, as already discussed, represents a popular choice for data-driven marketing attribution models among practitioners. As suggested by Anderl et al. (2013), we apply third-order Markov chains as it optimizes the general trade-off between predictive accuracy and model complexity which hinders interpretability.

Marketing Attribution. Figure 2 shows the contribution of the different channels based on the attribution generated by the four models. For comparison, the figure further includes the relative frequency with which the channels occur in the dataset. All models provide similar attributions, which appear to be correlated with the frequency of the channels. This supports the general plausibility of the results. The most remarkable differences between the overall frequency of channels and the attributed values can be observed for Channel 5, Channel 7, and Channel 9. All four models estimate Channel 5 to be significantly more relevant than suggested by its frequency. In all survival analysis models, the attribution of Channel 9 strongly deviates from the one in MarkovChain and the frequency of the channel, which suggests that Channel 9 is estimated to be much less influential. While the other survival analysis models assign significantly more value to Channel 7, CAHMA attributes only slightly more value to it than MarkovChain.

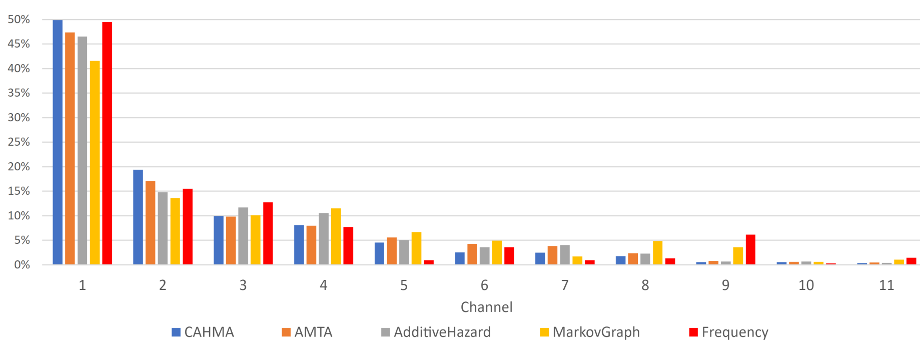

Figure 2. Results of the attribution for the different models.

Predictive Accuracy. To access predictive accuracy, the Receiver-Operator-Characteristic (ROC) curve and the area under the ROC curve (AUC) is a standard metric for the performance of probabilistic classifiers. However, the dataset at hand is highly skewed with a conversion rate of significantly less than 5%. For such highly skewed data, AUC tends to be overly optimistic in the measurement of an algorithm's performance (Branco et al., 2016). In these cases, Precision-Recall (PR) curves are more informative (Davis & Goadrich, 2006; Saito & Rehmsmeier, 2015). Furthermore, marketing attribution models can only attribute value for customer journeys which actually lead to a conversion. Consequently, it is primarily important for these models to accurately detect and map converting customer journeys. Moreover, marketing data is right censored as it is unknown whether non-converting customer journeys terminated because the user actually did not convert or simply for technical reasons, such as the deletion of the tracked cookie. This induces noise in the negative majority class. As PR curves are more sensitive to improvements in the positive minority class (Davis & Goadrich, 2006; Loezer et al., 2020; Saito & Rehmsmeier, 2015), they are particularly suitable to evaluate marketing attribution models. Thus, we include the area under the PR curve (PR-AUC) as the primary, central metric to measure the predictive accuracy. Additionally, we report the ROC curve and AUC as a supplementary metric. Panel A in Figure 3 depicts the

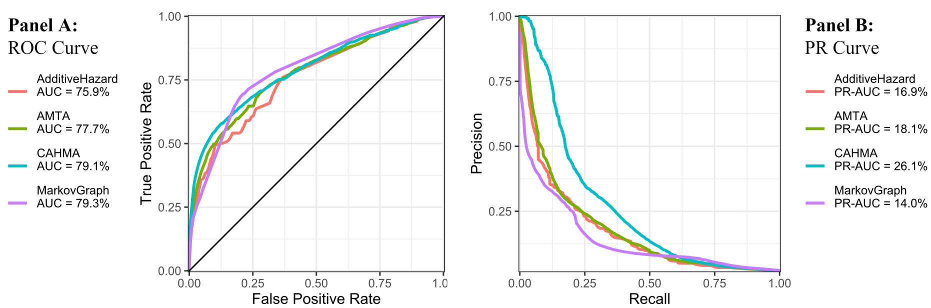

Figure 3. Results of the conversion prediction using AUC and PR-AUC.

achieved ROC curves and corresponding AUC values for CAHMA and the competing models based on the out-of-sample performance on the evaluation set. CAHMA achieves a higher AUC value than AdditiveHazard and AMTA, as related models based on survival analysis. MarkovChain achieves a slightly higher AUC than CAHMA, however, the ROC curve of MarkovChain does not strictly dominate the curve of CAHMA. For low false positive rates, CAHMA's ROC curve increases much steeper and thus lies above MarkovChain's ROC curve in this region, indicating a faster initial increase in the true positive rate at a lower cost of false positives. Thus, CAHMA outperforms MarkovChain in predicting the important minority class. This outperformance becomes more evident when comparing the more informative PR curves of all models (c.f. Panel B in Figure 3). Considering the PR-AUC, CAHMA significantly outperforms all competing models. Although significantly worse than CAHMA, both alternative survival analysis models also achieve a higher PR-AUC than MarkovChain. Overall, the results indicate that CAHMA is considerably better than the competing models at detecting converting customer journeys. The low PR-AUC of MarkovChain is driven by the sharp decline in its PR curve for low recall values. Hence, MarkovChain is more likely to miss converting customers, which is undesirable considering the highly skewed, right censored data and the objective to generate insights into why customers ultimately convert.

Interpretability. The general structure of MarkovChain is easy to understand and provides a transparent strategy for deriving the attribution. However, the ability to interpret the results is limited as MarkovChain cannot account for contextual features.

In contrast, additive survival analysis models in general have a significantly stronger interpretability of the results due to their model parameters. The model parameters of CAHMA give additional explanations for the observed differences between the frequency and the attribution of the channels and provide deeper insights into how the attribution was derived. Beyond the previous work by Zhang et al. (2014) and Ji & Wang (2017), we normalize the influence strength β_k to the interval from 0% to 100% and calculate the half-life of a channel's effect from its time-decaying property ω_k. This allows us to easily compare the parameters and analyze how different channels exert their influence. Figure 4 shows the transformed values of the parameters β_k and ω_k. Channel 1 has a relatively small influence strength and a very low half-life, indicating that each time a user interacts with this channel, it only exerts a small influence that decays very quickly with time. However, due to its high frequency (c.f. Figure 2) in the dataset, Channel 1 adds a small effect very often and is thus still attributed the highest value. On the other hand, touchpoints with, e.g., Channel 5 or Channel 7 are very scarce but exert a much higher and/or longer-lasting influence and are thus still attributed a considerable value above their relative frequency. In general, differences among the influence strengths and

Panel A: Relative Influence Strength

Panel B: Half-life in Days

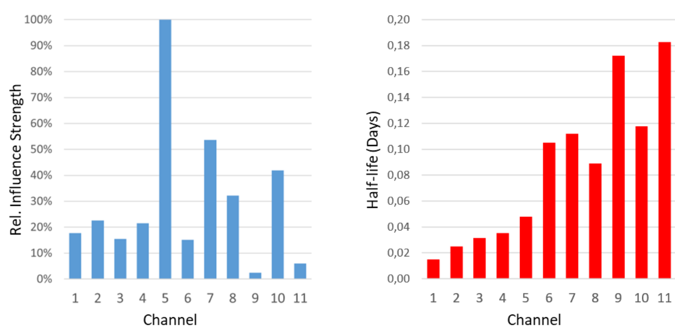

Figure 4. Overview of model parameters

half-lives of marketing channels are very plausible. For example, each individual click on a search result when looking for tickets to attractions is likely to have less impact and to be forgotten more quickly than when a user is individually targeted by a personalized email or receives a link shared by a friend or idol. This interpretability of influence and persistency of channels over time is a key strength of additive survival analysis models.

In addition to the channel-specific influence strength of marketing and its half-life, CAHMA allows for determining the channel-independent effect of marketing. The average contribution of the baseline hazard at conversion time is 4.72%. As elaborated in Section 2.1, we distribute this contribution over all channels in the customer journeys to derive the marketing attribution. However, controlling for the influence of contextual features provides additional insights. Contextual features that are correlated with a high conversion probability will increase the baseline hazard rate and vice versa. As the baseline hazard is modeled by an exponential form, we can directly compare the influence on the baseline hazard of two contextual features with parameters α_1 and α_2. For example, the estimated parameters for the devices iOS and Android are 3.10 and 2.43, respectively. The resulting ratio $e^{3.10-2.43} = 1.95$ suggests, that, ceteris paribus, the channel-independent baseline hazard for iOS users is about 95% higher than for Android users. Consequently, with an identical customer journey, iOS users would be more likely to convert. Overall, our model suggests that iOS users have the highest and users of other mobile devices have the lowest baseline hazard. Fall and winter have a substantially lower impact on a user's hazard rate compared to summer. In general, interactions from European countries have greater influence compared to other regions.

3 Discussion and Conclusion

Overall, as presented in Table 1, our proposed model meets the evaluation criteria very well. CAHMA showed high predictive accuracy compared to the competing models. This supports the validity of CAHMA's attribution. The high predictive accuracy of CAHMA does not come at the expense of interpretability. The structure of the proposed model is very comprehensible and inherits the strong interpretability of the results of additive survival analysis models. In addition, the baseline can be used to examine the extent to which different contextual features influence the user's probability to convert. This allows to directly compare these influences as well as determining the value generated by channel-independent marketing effects. The latter is not included in Ji & Wang's (2017) model. Therefore, our model enables a unique interpretation of the contextual features.

Table 1. Summary of the evaluation results.

Model	Predictive Accuracy	Interpretability	
	PR-AUC	of the model	of the results
MarkovChain	14.0%	+	
AdditiveHazard	16.9%	+	+
AMTA	18.1%	+	+
CAHMA	**26.1**%	**+**	**++**

We contribute to the literature on data-driven marketing attribution models in general and to models based on survival analysis in particular. Adding a context-dependent baseline hazard allows to control for channel-independent effects of marketing and, thus, to better capture the "true" effect of marketing channels. Also, the baseline resolves the mathematical issue of previous additive survival models, as it aligns CAHMA with the assumptions of survival analysis and properties of probability theory.

With interpretability as an important criterion for marketing attribution models, our research has additional practical contributions. Estimating an influence strength and a half-life of the effect of each marketing channel generates further insights into "how" marketing channels exert their influence and affect users. This interpretation has not been the focus of researchers and is impossible to achieve for Markov chain models, which are commonly used in the industry. Integrating contextual features in our model in form of a baseline hazard enables flexible extension and adaption of the model to specific use cases and provides further interesting insights, as the effects of different contextual features on the user's conversion probability can be compared (c.f. Section 2.2). While we model contextual features such as the season or country in which the user interacted with the marketing channels, our model is not restricted to these kinds of features. For example, campaign-level data could be used in the baseline hazard as contextual features to further model the influence of campaign characteristics over various channels on the conversion probabilities of users. This would then allow for comparing the general, channel-independent effect of marketing campaigns in addition to channel-level marketing attribution. Altogether, by implementing a prototype of CAHMA as part of an applied research project with GetYourGuide, we demonstrate the practical value of the model. Considering its interpretability and flexible extensibility, we outline the valuable support that CAHMA can provide for strategic marketing decisions.

Our work has limitations that can serve as starting points for future research. First, we evaluate our model on one dataset which only covers 9 months of data missing the spring season. As spring is likely a busy season for the travel products, it would be interesting to analyze additional data to estimate the impact of the spring season. In general, it would be interesting to see, whether our results hold for other datasets from various industries. Furthermore, our dataset only contains online channels such as SEA, display or social media advertising and is solely click-based. However, impressions, i.e., touchpoints where a user sees an advertisement without clicking on it, and exposure to offline marketing can have a relevant influence on users' purchase decisions. Although we expect the proposed model to successfully process impressions and offline channels as well, this could be verified in future work. Third, the model could be evaluated using additional criteria such as robustness and algorithmic efficiency (Anderl et al., 2013). While our in- and out-of-sample predictive accuracy are stable and, as discussed, applying MapReduce promises good scalability, a more formal assessment of these criteria could be performed in future. Finally, future research could further analyze how CAHMA performs compared to neural networks despite their lack of interpretability. This would allow to quantify the trade-off between predictive accuracy and interpretability. Although interpretability is crucial, some practitioners may consider black-box models if the gain in predictive accuracy is sufficiently large.

Based on the proposed model, further extensions are possible and indicate promising research opportunities. While we model an exponential form for the baseline hazard to guarantee mathematical validity, other functional forms are possible. In the same vein, the approach of additive survival analysis is not limited to the implemented exponential kernel function for the channel-specific marketing effects. Alternative kernel functions to allow for different time-dependent effects could be investigated in future work. Finally, in line with the assumptions of survival models, our model assumes that every user converts within an infinite timeframe. To account for the fact that not every user does indeed convert, the proposed model could be extended by additionally considering the probability that the waiting time until conversion exceeds the time a user can be observed due to technical and lifetime restrictions. However, it must be ensured that the additional complexity does not impair the interpretability and thus practical relevance of the model.

References

Anderl, E., Becker, I., von Wangenheim, F. & Schumann, J. H. (2013), 'Putting attribution to work: A graph-based framework for attribution modeling in managerial practice', *SSRN Electronic Journal* .

Anderl, E., Becker, I., von Wangenheim, F. & Schumann, J. H. (2016), 'Mapping the customer journey: Lessons learned from graph-based online attribution modeling', *International Journal of Research in Marketing* **33**(3), 457–474.

Branco, P., Torgo, L. & Ribeiro, R. P. (2016), 'A survey of predictive modeling on imbalanced domains', *ACM Computing Surveys* **49**(2), 1–50.

Davis, J. & Goadrich, M. (2006), The relationship between precision-recall and roc curves, *in* 'Proceedings of the 23rd international conference on Machine learning - ICML '06', ACM, pp. 233–240.

Dean, J. & Ghemawat, S. (2008), 'Mapreduce', *Communications of the ACM* **51**(1), 107–113.

Ji, W. & Wang, X. (2017), Additional multi-touch attribution for online advertising, *in* 'Proceedings of the AAAI Conference on Artificial Intelligence', AAAI, pp. 1360–1366.

Kannan, P. K., Reinartz, W. & Verhoef, P. C. (2016), 'The path to purchase and attribution modeling: Introduction to special section', *International Journal of Research in Marketing* **33**(3), 449–456.

Kumar, S., Gupta, G., Prasad, R., Chatterjee, A., Vig, L. & Shroff, G. (2020), Camta: Causal attention model for multi-touch attribution, *in* '2020 International Conference on Data Mining Workshops (ICDMW)', IEEE, pp. 79–86.

Lawless, J. (2011), *Statistical Models and Methods for Lifetime Data*, 2nd edn, Wiley.

Little, J. D. C. (2004), 'Models and managers: The concept of a decision calculus', *Management Science* **50**(12_supplement), 1841–1853.

Lodish, L. M. (2001), 'Building marketing models that make money', *Interfaces* **31**(3_supplement), 45–55.

Loezer, L., Enembreck, F., Barddal, J. P. & de Souza Britto, A. (2020), Cost-sensitive learning for imbalanced data streams, *in* 'Proceedings of the 35th Annual ACM Symposium on Applied Computing', ACM, pp. 498–504.

Romero Leguina, J., Cuevas Rumín, Á. & Cuevas Rumín, R. (2020), 'Digital marketing attribution: Understanding the user path', *Electronics* **9**(11), 1822.

Saito, T. & Rehmsmeier, M. (2015), 'The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets', *PloS one* **10**(3), e0118432.

Shao, X. & Li, L. (2011), Data-driven multi-touch attribution models, *in* 'Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11', ACM, pp. 258–264.

Statista (2022*a*), 'Cloud computing: usage in companies in germany 2022', `https://www.statista.com/statistics/459998/usage-of-cloud-computing-by-companies-in-germany/?locale=en`. Accessed: 09.03.2023.

Statista (2022*b*), 'Digital advertising - worldwide', `https://www.statista.com/outlook/dmo/digital-advertising/worldwide`. Accessed: 09.03.2023.

Yang, D., Dyer, K. & Wang, S. (2020), 'Interpretable deep learning model for online multi-touch attribution', *arXiv preprint arXiv:2004.00384* .

Yao, D., Gong, C., Zhang, L., Chen, S. & Bi, J. (2022), Causalmta: Eliminating the user confounding bias for causal multi-touch attribution, *in* 'Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining', ACM, pp. 4342–4352.

Zaremba, A. (2020), 'Conversion attribution:what is missed by the advertising industry? the opec model and its consequencesfor media mix modeling', *Journal of Marketing and Consumer Behaviour in Emerging Markets* **10**(1), 4–23.

Zhang, Y., Wei, Y. & Ren, J. (2014), Multi-touch attribution in online advertising with survival theory, *in* '2014 IEEE International Conference on Data Mining', IEEE, pp. 687–696.