Context-Aware Marketing Attribution Based on Survival Analysis - Online Appendix

In this appendix we provide details on related literature that examines marketing attribution models and outline the optimization problem as well as the corresponding analytic functions.

Related Literature

Marketing attribution is a well-known and discussed problem, such that the existing literature offers a variety of data-driven methods for this purpose (see e.g., Zaremba (2020) for an overview). For a marketing attribution model to be used in practice, it needs to fulfill the following two crucial criteria. First, the model must be able to correctly predict conversions to attest the validity of the attribution result (Lodish, 2001). Second, the model must be understandable for all stakeholders that rely on it to make strategic decisions (Little, 2004). Managers, for example, usually do not have an analytical background, but they are the ones in charge when it comes to the allocation of marketing budgets.

Logistic regression models, such as the pilot study in the marketing attribution field from Shao & Li (2011), are solid but not state of the art at predicting conversions (e.g., Zhang et al., 2014; Ji & Wang, 2017; Ren et al., 2018). Previous research indicates that logistic regression models are commonly outperformed by other models in predicting conversions. Game theoretic models based on the Shapley value (e.g., Yadagiri et al., 2015; Berman, 2018; Zhao et al., 2018) are easy to understand and offer a comprehensive strategy to derive the marketing attribution. However, these models have two major drawbacks (Romero Leguina et al., 2020). First, Shapley value attribution models cannot be used for conversion prediction, and hence, it is not possible to quantify their predictive accuracy. Second, Shapley value models do not scale well to big datasets because of their high computational complexity. In contrast to the Shapley value, models based on a neural network (e.g., Ren et al., 2018; Kumar et al., 2020; Yao et al., 2022) may be best at predicting conversions, however, they sacrifice on interpretability of their model structure and results. Hence, practitioners are unlikely to rely on them for strategic decision-making, which is the ultimate purpose of attribution models (Romero Leguina et al., 2020).

In the following, we want to discuss two other methods that achieve good predictive accuracy and at the same time have a comprehensive structure which makes them attractive choices for practitioners. Anderl et al. (2013) build a novel attribution framework based on the concepts of Markov chains. A Markov chain is a stochastic process that maps the sequence of predefined states. The entire model is defined by the transition matrix that captures all possible transitions between states and the corresponding transition probabilities. The order of a Markov chain defines the amount of "memory" allowed in the model. Markov chains of order n make transition probabilities dependent on the previous n states. In the context of marketing attribution, each customer journey represents one instantiation of a Markov chain (every channel represents one state) that ends in one of the two terminal states "conversion" and "no conversion". The attribution of one channel is then derived by calculating its removal effect, which is the change in the probability to end up in the conversion state if this channel and all dependent transactions are deleted from the model. These concepts lead to a nice and comprehensible model structure and allow the evaluation of the predictive accuracy of a Markov chain model. Anderl et al. (2013) compare different Markov chain models of first- and higher-order and find that higher-order models have better predictive accuracy than first-order models. However, the complexity of Markov chain models increases exponentially with their order which has a negative impact on interpretability (Anderl et al., 2016). Thus, the authors propose third-order Markov chain models, as they balance high predictive accuracy and the ability to interpret the model. Abhishek et al. (2012) introduce a hidden Markov chain model, which outperforms logistic regression models in terms of accurate conversion prediction, but due to its degree of complexity, loses the good interpretability of Markov chains. While Markov chain models seem promising, these models are constrained in their interpretability of the results and, therefore, the generated insights are limited. There exist correlations between marketing effects and conversions that Markov chains cannot model (Anderl et al., 2016). For example, it is plausible that the effect of marketing does change over time, however, Markov chains cannot consider a time-decaying effect.

In contrast, attribution models based on survival analysis (e.g., Zhang et al., 2014; Ji et al., 2016; Ji & Wang, 2017) do account for time-decaying effects of marketing. They aim to derive a marketing attribution strategy based on the conversion probability of a customer. Assuming an additive effect between the different marketing channels, Zhang et al. (2014) build an attribution model that accounts for an exponential time decay. However, they neglect contextual features that can be observed in customer journeys. Ji et al. (2016) do account for contextual features by modeling a time-independent conversion rate. Yet, assuming a proportional, multiplicative effect instead of an additive effect between the individual channels makes it harder to interpret the results of the model. Finally, Ji & Wang's (2017) model combines both approaches by modeling a joint distribution of a time-independent, intrinsic conversion probability of users and time-dependent conversion probability from the additive marketing channel effects. The authors compare their model with the other survival analysis models and show that it is better at predicting conversions. In its components, the model is quite comprehensive, and the time-dependent component of the joint distribution can give relevant insights into the effects of the individual marketing channels. However overall, the model is rather complex and the joint distribution that results from the time-independent conversion rate and the hazard rate is not a true distribution function because calculating the time-independent marginal distribution, i.e., eliminating the time-dependent component, does not yield the defined time-independent intrinsic conversion probability of users.

We propose an alternative survival analysis model that also accounts for contextual features and assumes an additive relationship between the effects of the marketing channels on the conversion rate. Compared to Ji & Wang's (2017) model, which can account for contextual features in form of an additional dimension of a joint probability distribution, we propose a different, simpler approach. We include contextual variables directly in the instantaneous conversion probability of users. In this way we directly control for the channel-independent but context-dependent effect of marketing, in a similar manner as control variables are included in regression models. This should allow the model to better separate context-specific from channel-specific effects and identify the "true" channel-specific effect of marketing. Consequently, beyond having a simpler structure, which facilitates model implementation and interpretation, and being mathematically correct, we further expect our model to have better predictive accuracy.

Parameter Estimation

To fit all parameters $\Theta = \{\alpha_0, \ldots, \alpha_I, \beta_1, \ldots, \beta_K, \omega_1, \ldots, \omega_K\}$ to the data, we use maximum likelihood estimation. Subject to the discussed constraints for the channel-specific parameters, the maximization problem

$$\begin{aligned} &\max_{\Theta} \log \mathcal{L}(\Theta) \\ \text{s. t.}\ \ &\beta_k \geq 0, \quad k = 1, \ldots, K \\ &\omega_k \geq 0, \quad k = 1, \ldots, K \end{aligned}$$

must be solved. If we observe that a customer journey leads to a conversion, its contribution to the likelihood is the probability to "survive" until and convert at conversion time T^u. If we observe a customer journey that does not lead to a conversion within our observation window, its contribution to the likelihood is the probability that the waiting time is greater than the last observable time T^u. Hence, using stated definitions and basic algebra, the log-likelihood function for the waiting time T is given by:

$$\begin{aligned} \log \mathcal{L}(\Theta) = &\sum_{u \in U} \left(-e^{\alpha_0 + \sum_{i=1}^{I} \alpha_i X_i^u} \left(T^u - t_1^u\right) - \sum_{t_l^u < T^u} \beta_{a_l^u} \left(1 - e^{-\omega_{a_l^u}(T^u - t_l^u)}\right) \right) \\ &+ \sum_{u \in U, Y^u = 1} \log \left(e^{\alpha_0 + \sum_{i=1}^{I} \alpha_i X_i^u} + \sum_{t_l^u < T^u} \beta_{a_l^u} \omega_{a_l^u} e^{-\omega_{a_l^u}(T^u - t_l^u)} \right) \end{aligned} \tag{1}$$

The first line in (1) represents the probability that a user does not convert before the latest timestamp of the customer journey, which is either the conversion time or the last timestamp in the observation window. This is given for both converting and non-converting customer journeys. Additionally, users that do convert, will convert just at their latest timestamp. This is represented by the term in the second line of (1).

The gradients with respect to the parameters α_i for $i \in \{0, \ldots, I\}$, β_k and ω_k for $k \in \{1, \ldots, K\}$ are given by

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \alpha_0} = &\sum_{u \in U, Y^u = 1} \frac{e^{\alpha_0 + \sum_{i=1}^{I} \alpha_i X_i^u}}{e^{\alpha_0 + \sum_{i=1}^{I} \alpha_i X_i^u} + \sum_{t_l^u < T^u} \beta_{a_l^u} \omega_{a_l^u} e^{-\omega_{a_l^u}\left(T^u - t_l^u\right)}} \\ &- \sum_{u \in U} e^{\alpha_0 + \sum_{i=1}^{I} \alpha_i X_i^u} \\ \frac{\partial \log \mathcal{L}}{\partial \alpha_j} = &\sum_{u \in U, Y^u = 1} \frac{e^{\alpha_0 + \sum_{i=1}^{I} \alpha_i X_i^u} X_j^u}{e^{\alpha_0 + \sum_{i=1}^{I} \alpha_i X_i^u} + \sum_{t_l^u < T^u} \beta_{a_l^u} \omega_{a_l^u} e^{-\omega_{a_l^u}\left(T^u - t_l^u\right)}} \\ &- \sum_{u \in U} e^{\alpha_0 + \sum_{i=1}^{I} \alpha_i X_i^u} X_j^u \end{aligned}$$

$$\frac{\partial \log \mathcal{L}}{\partial \beta_k} = \sum_{u\in U, Y^u=1} \frac{\sum_{t_l^u<T^u, a_l^u=k} \omega_k e^{-\omega_{a_l^u}(T^u-t_l^u)}}{e^{\alpha_0+\sum_{i=1}^{I}\alpha_i X_i^u} + \sum_{t_l^u<T^u} \beta_{a_l^u}\omega_{a_l^u} e^{-\omega_{a_l^u}\left(T^u-t_l^u\right)}} - \sum_{u\in U} \sum_{t_l^u<T^u, a_l^u=k} \left(1-e^{-\omega_{a_l^u}(T^u-t_l^u)}\right)$$

$$\frac{\partial \log \mathcal{L}}{\partial \omega_k} = \sum_{u\in U, Y^u=1} \frac{\sum_{t_l^u<T^u, a_l^u=k} \beta_k e^{-\omega_{a_l^u}(T^u-t_l^u)} \left(1-\omega_{a_l^u}\left(T^u-t_l^u\right)\right)}{e^{\alpha_0+\sum_{i=1}^{I}\alpha_i X_i^u} + \sum_{t_l^u<T^u} \beta_{a_l^u}\omega_{a_l^u} e^{-\omega_{a_l^u}\left(T^u-t_l^u\right)}} - \sum_{u\in U} \sum_{t_l^u<T^u, a_l^u=k} \beta_k \left(T^u-t_l^u\right) e^{-\omega_{a_l^u}(T^u-t_l^u)}$$

References

Abhishek, V., Fader, P. & Hosanagar, K. (2012), 'The long road to online conversion: A model of multi-channel attribution', *SSRN Electronic Journal* .

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.

Berman, R. (2018), 'Beyond the last touch: Attribution in online advertising', *Marketing Science* **37**(5), 771–792.

Ji, W. & Wang, X. (2017), Additional multi-touch attribution for online advertising, *in* 'Proceedings of the AAAI Conference on Artificial Intelligence', Vol. 31.

Ji, W., Wang, X. & Zhang, D. (2016), A probabilistic multi-touch attribution model for online advertising, *in* 'Proceedings of the 25th ACM International on Conference on Information and Knowledge Management', ACM, New York, NY, USA, pp. 1373–1382.

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.

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.

Ren, K., Fang, Y., Zhang, W., Liu, S., Li, J., Zhang, Y., Yu, Y. & Wang, J. (2018), Learning multi-touch conversion attribution with dual-attention mechanisms for online advertising, *in* 'Proceedings of the 27th ACM International Conference on Information and Knowledge Management', ACM, New York, NY, USA, pp. 1433–1442.

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

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 Press, New York, NY, USA, p. 258.

Yadagiri, M. M., Saini, S. K. & Sinha, R. (2015), A non-parametric approach to the multi-channel attribution problem, *in* 'Web Information Systems Engineering – WISE 2015', Vol. 9418 of *Lecture Notes in Computer Science*, Springer International Publishing, Cham, pp. 338–352.

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', 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* **1/2020**(10), 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.

Zhao, K., Mahboobi, S. H. & Bagheri, S. R. (2018), 'Shapley value methods for attribution modeling in online advertising'.