Impact of Ad Impressions on Dynamic Commercial Actions: Value Attribution in Marketing Campaigns

Joel Barajas*[1], Ram Akella[1], Marius Holtan[2], Jaimie Kwon[2], Aaron Flores[2], Victor Andrei[2]

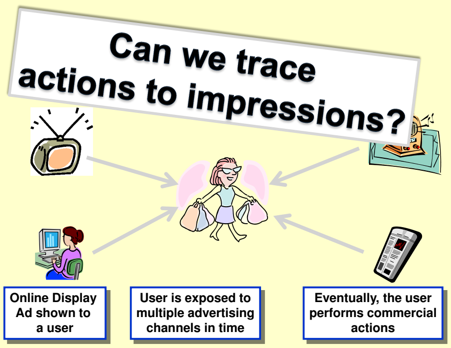
[1]UC Santa Cruz, Santa Cruz CA, USA

[2]AOL Research, Palo Alto CA, USA

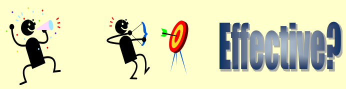
*Contact Author: jbarajas@soe.ucsc.edu

Introduction

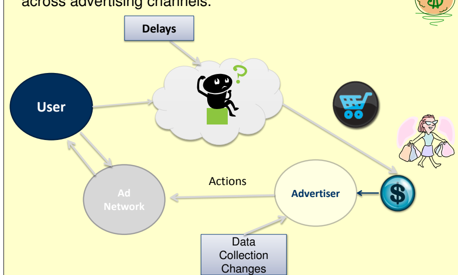
- Online display advertising is an area of rapid growth and consequently of great interest as a marketing channel.
- Eventually, many of these users perform either online conversions at the advertiser's website or offline conversions at a physical store.



- Our goal is to measure the effectiveness of display advertising when users are exposed to multiple advertising channels.



- This is particularly important for the Cost-Per-Action CPA business model when deciding how to allocate resources across advertising channels.



- A key difference in CPA is that commercial actions are collected by advertisers.
- In CPA, it could be several days before a commercial action is performed after showing an impression
  1. **Restructuring of the website**
  2. **Merging of products to a single ID**
  3. **Disaggregation of products to create a new ID**

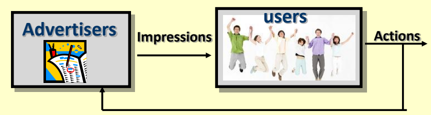
Estimates from the *advertising.com* ad networks show around 15% of users with unreliable cookies.

Related Work

- For reliable cookies, running experiments (A/B testing) have been proposed recently in [2].
  - –This is expensive as public service announcements (PSA) should be displayed to a percentage of audience.
  - –There is an advertising opportunity cost in addition to PSAs impression cost
  - –This requires significant human intervention to track and randomize properly the control group of users
- On the other hand, correcting observational data have been proposed to evaluate campaigns in [1].
  - –This method relies heavily on fully observed features
  - –With high percentage of users deleting and creating tracking cookies, user features collection becomes challenging.
- In general, these methods rely heavily on cookies, and time is not addressed properly as a confounder.
- In contrast, we model the impact of ad impressions on actions to provide daily estimates without cookies or user features.
- We account for time and seasonal confounders for attribution
- This is easily scalable to thousands of campaigns

Methodology

- We develop an interpretative method, based on Dynamic Linear Models (DLM)[3], to estimate the impact of ad impressions on actions without cookies.
- Our method is scalable for millions of users and thousands of campaigns using aggregate data.



- Time series model accounts for multi-channel effect and confounders
- Decay Rate Accounts For user persistence
- Multi-campaign model. If multiple channels are observed, each channel is represented by a campaign
- Dynamic effect of impressions on commercial actions

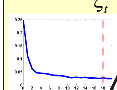
Base time series model to account for confounding effects

Combination of campaign effects

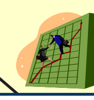
$$y_t = \theta_t^0 + \sum_{c=1}^{M} \xi_t^{(c)} + v$$

$$\theta_t^0 = \theta_{t-1}^0 + w_t^{(0)}$$

$$\xi_t^{(c)} = \lambda^{(c)} \xi_{t-1}^{(c)} + \psi_t^{(c)} x_t^{(c)} + w_t^{(\varepsilon,c)}$$



$$\psi_t^{(c)} = \psi_{t-1}^{(c)} + \varepsilon_t^{(\psi,c)}$$



Exponential decay (lead-lag) effect of impressions on actions

Dynamic Coefficient for the number of impressions (dynamic regression)

$x_t^{(c)}$ = Log of the Number of Impressions at Time t from Campaign c

y_t = Log of the Number of Actions at time t

M = Number of Campaigns

θ_t^0 = Any DLM-based Model

- **This model can be written as a DLM as follows:**

$$y_t = F'\theta_t + v_t \qquad v_t \sim N(0,V)$$

$$\theta_t = G\theta_{t-1} + w_t \qquad w_t \sim N(0,W)$$

$$F' = \left[1,(1,0)^{(1)},\ldots,(1,0)^{M}\right] \quad \theta_t' = \left[\theta_t^0,\xi_t^{(1)},\psi_t^{(1)},\ldots,\xi_t^{(M)},\psi_t^{(M)}\right]$$

$$G_t = blockdiag\left(G_t^0, \begin{bmatrix}\lambda^{(1)} & X_t^{(1)} \\ 0 & 1\end{bmatrix}, \ldots, \begin{bmatrix}\lambda^{(M)} & X_t^{(M)} \\ 0 & 1\end{bmatrix}\right)$$

Algorithm 1 Generative Model to Handle Outliers

```
Draw p|α ∼ Dirichlet(α)
for t ← 1 to T do
   Draw η_t|p ∼ Mult(1, p),   ω_t|η_t ∼ Γ(η_t/2, η_t/2),   V_t = ω_t^{-1} V
end for
```

Model Fitting

- We follow a Gibbs sampling approach detailed in Algorithm 2.

Algorithm 2 Gibbs Sampling Algorithm

Define $D_{1:T} = \left\{Y_{1:T}, X_{1:T}^{(1:N)}\right\}$, $\Omega = \{\omega_{1:T}, \eta_{1:T}, p\}$

Define $\Phi = \left\{\lambda^{(1:N)}, W_{\psi}^{(1:N)}, W_{\xi}^{(1:N)}, W^{(0)}, V\right\}$

for $s \leftarrow 1$ to $N_0 + N_s$ **do**

Draw $\theta_{1:T}^s \sim p\left(\theta_{1:T} | \Phi^{s-1}, \Omega^{s-1}, D_{1:T}\right)$

Draw $\Phi^s \sim p\left(\Phi | \theta_{1:T}^s, \Omega^{s-1}, D_{1:T}\right)$

Draw $\Omega^s \sim p\left(\Omega | \theta_{1:T}^s, \Phi^s D_{1:T}\right)$

end for

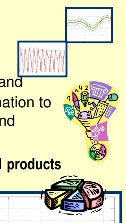
$\boxed{\theta_{1:T} | \Phi, \Omega, D_{1:T}}$ – We sample the latent state based on Forward Filtering Backward Sampling FFBS [3].

$\boxed{\Phi | \theta_{1:T}, \Omega, D_{1:T}}$ – To sample the variances, we assume standard Inverse Gamma conjugate priors, and truncated Normals for the decay rate λ

$\boxed{\Omega | \theta_{1:T}, \Phi, D_{1:T}}$ – We sample the outlier processing variables based on Algorithm 1.

Results

- We test two base models:
  1. Random Walk (simplest dynamic model)
  2. Weekly Seasonal model [4]



- We test the model with the number of actions and impressions, (Mω) and using the log-transformation to relax the linear relationship between actions and impressions (Mωlog).
- **We analyze 2,885 campaigns associated with 1,251 products during 6 months.**

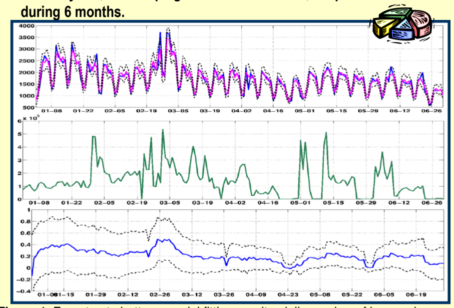


Figure 1: From top to bottom: model fitting results, daily number of impressions, proportion of actions attributed to impressions. *X*-axis is time in dates.

Model	Fitted	Forecast	Fitted $\omega_t = 1$	Forecast $\omega_t = 1$
Random Walk Base model, MRSE				
Mω	7.91 ± 1.85	61.77 ± 7.13	14.87 ± 2.40	72.13 ± 7.58
Mω log	1.33 ± 0.32	**13.25±2.21**	5.49 ± 1.02	20.00 ± 2.57
Weekly Seasonal Base model, MRSE				
Mω	8.26 ± 2.13	61.13 ± 7.34	12.65 ± 2.11	70.75 ± 7.67
Mω log	**0.72±0.14**	15.51 ± 2.81	**3.92±0.92**	21.20 ± 3.12

Table 1: Model evaluation results, scaled by 0.01, averaged over products. 95% confidence intervals are shown. Mean relative squared error, MRSE $e=(Y_t\text{-}\hat{Y}_t)/Y_t$

Model	% attributed	Campaign Significance (+)	Campaign Significance (-)	Campaign Significance (±)
Random Walk Base model				
Mω	14.07 ± 1.36	23.13	0.71	76.09
Mω log	21.31 ± 1.63	18.65	0.58	**80.71**
Weekly Seasonal Base model				
Mω	10.39 ± 1.23	19.66	1.31	78.96
Mω log	19.84 ± 1.64	14.83	0.60	**83.98**

Table 2: Averaged campaign evaluation results. Distribution of campaign effect significance (%).

- Higher performance is reported when the log transformation is included suggesting a non-linear relationship between actions and impressions.
- This model reports the highest campaign percentage with non-significant effect

Comparison with A/B Testing

Method	Campaign 1 Low	Campaign 1 Med	Campaign 1 High	Campaign 2 Low	Campaign 2 Med	Campaign 2 High
A/B	0.009	0.199	0.458	-0.034	0.115	0.312
Mωlog	0.013	0.051	0.107	-0.049	0.347	0.809

Table 3: A/B testing results compared to the attribution given by Mωlog for the Random Walk base model.

- Confidence intervals for A/B testing are not tight due to the sparsity of actions. Both methods report one positive significant campaign at 90% confidence level and one leaning towards positive effect.

Comparison with Ground Truth

- We test our method with the PROMO dataset to compare with a ground truth [4].
- We use products with less than 6 relevant campaigns for 365 days.
- We detect **84.6%** of effective campaigns correctly, and **73.2%** of the days a campaign is effective per product.

Discussion and Current Work

- We observed several campaigns with non-significant average effect on actions, which is consistent with A/B testing results.
- When experimentation is possible and we can track users, can we combine A/B testing with time series?
- A/B testing as described in [2] is expensive. Can this experimentation be less costly?
- Achieving statistical significance for campaign lifts is non-trivial for commercial actions. Can we aggregate prior information to improve these estimations

Acknowledgements

- This work is partially funded by CONACYT UC-MEXUS grant 194880, CITRIS and AOL Faculty Award.

References

[1] D. Chan, R. Ge, O. Gershony, T. Hesterberg, and D. Lambert. Evaluating online ad campaigns in a pipeline: causal models at scale. In *Proceedings of SIGKDD*, pages 7–16. ACM, 2010.

[2] R. A. Lewis, J. M. Rao, and D. H. Reiley. Here, there, and everywhere: correlated online behaviors can lead to overestimates of the effects of advertising. In *Proceedings of WWW2011*, pages 157–166. ACM, 2011.

[3] G. Petris, S. Petrone, and P. Campagnoli. *Dynamic Linear Models with R*. use R! Springer-Verlag, 2009.

[4] J. Barajas, R. Akella, M. Holtan, J. Kwon, and B. Null. Measuring the effectiveness of display advertising: a time series approach. In *WWW (Companion Volume)*, pages 7–8, 2011.

[5] Causality workbench team. PROMO: Simple causal effects in time series, 2008.