Modelling Advertising Awareness, an Interpretable and Differentiable Approach

Luz M. Blaz
Kantar
Mexico City, MEX, 03810
luzmariana.blaz@kantar.com

Moisés Arizpe
Kantar
Mexico City, MEX, 03810
moises.arizpe@kantar.com

Abstract

We present a fully interpretable learning architecture capable of describing the dynamics between media investment and advertising awareness. We propose that the effect of media investment in advertising awareness of a brand can be modeled using a system of ordinary differential equations (ODE). An observation mechanism in the form of a trainable Kalman filter is included in our model. The proposed architecture is fully differentiable and thus, it can be efficiently learnt by backpropagation and estimated at scale.

1 Introduction

Awareness measures are used extensively in market research as a gauge of brand performance and marketing effectiveness. The most commonly used are those relating to brand and advertising awareness [18]. Brand awareness refers to whether consumers can recall or recognize a brand, or simply whether consumers know about a brand [13], Advertising Awareness (AdA) can be described as whether traces of a brand's advertising reside in consumers' memories [20].

AdA is usually measured by asking brand-cued awareness measures, exemplified by the question: "Have you seen any advertising for Diners Club Card recently?"[19]. After asking this kind of question to a sample of a population, the percentage of the people in the sample that answered affirmatively, is reported as a measure of AdA. These measurements are usually tracked continuously, giving us access to time-series describing AdA levels over time.

One of the most important questions in market research concerns the relationship between an AdA and advertising investment. "Half the money I spend on advertising is wasted; the trouble is, I don't know which half," is perhaps one of the most-known marketing sayings. There are no general models for this relationship, even when common assumptions are taken. The design process of an AdA model is usually made "by eye" and based on empirical observations more than on a structured model [4]. Other approaches resort to linear regression and a set of ad-hoc data transformations to consider non-linearities [11]. These issues result in model estimates that are highly dependent on the modeler and that are difficult to interpret and automate.

To address these problems, we propose a physically-inspired interpretable model encoded as a system of Ordinary Differential Equations. Our model allows us to relate the movements seen in AdA with the spending levels used to achieve them. The model includes an observation mechanism that allows us to filter out the error introduced by our measurement instrument – a survey. The model is fully differentiable and thus, it can be efficiently learned by backpropagation and estimated at scale.

There are previous works where an ODE modelling has been implemented [1][21][16][7], but are different to ours since: They only focus on what happens after a campaign is over, and not in how that campaign affects the system [1] or they are focused on sales, rather than AdA [21][16][7] (AdA causes a predisposition to buy from the brand but there are other factors involved).

1st NeurIPS workshop on Interpretable Inductive Biases and Physically Structured Learning (2020), virtual.

2 Background and problem setup

2.1 Differentiable ODE solvers

In recent years, it has been proved that Residual Networks (ResNets) [10] can be interpreted as an Euler discretization of a continuous transformation [15][8]. And that, if sufficient layers are added, emulating shorter steps of the transformation, it is possible to parameterize the dynamics of the hidden layers of a ResNet as an ordinary differential equation (ODE) [5]

$$\frac{d\mathbf{h}(t)}{dt} = f(\mathbf{h}(t), t, \theta). \tag{1}$$

In such interpretation, the input layer corresponds to the initial condition $h(0)$, and the output layer $\mathbf{h}(T)$ is an approximation to the solution of the ODE equation with initial value $\mathbf{h}(0)$ at some time T [5].

To fit a time series using this approach, an iterative process of adjusting the f-encoded system dynamics, is performed, in such way that the Neural ODE transforms a given input $\mathbf{x} = \mathbf{h}(0)$ into an output, y, that is closer to the true values of the series at time T [6].

A PyTorch-based [17] package (`torchdiffeq`) with an implementation of an algorithm that can perform the backpropagation process through an ODE solver was developed by [5], and can be found at https://github.com/rtqichen/torchdiffeq.

2.2 Kalman filters

A Kalman filter [12] is an algorithm that uses a series of measurements observed over time, containing statistical noise and other inaccuracies, and produces estimates of unknown variables that tend to be more accurate than those based on a single measurement alone. This kind of filtering is very powerful in several aspects, one of the most relevant for our work being that it supports estimations even when the precise nature of the modeled system is unknown [22].

A classical Kalman filter assumes that for each unobserved variable, z_t , in a sequence, we have an observation x_t , and a corresponding action u_t , which is also observed. The filter models the observed sequence as follows:

$$z_{t+1} = G_n z_t + B_t u_{t-1} + w_t, \tag{2}$$
$$x_t = F z_t + v_t, \tag{3}$$

where w_t and v_t are normally distributed random variables.

AdA is not directly observable, it is measured by conducting surveys. Such surveys are subject to a variety of measurement errors [3]. We cannot a priori know the distribution of all those errors. Thus, we use a Kalman filter to model the unknown observation process. The noise and evolution parameters of the filter are learnt simultaneously to the ODE parameters via backpropagation.

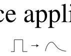
2.3 Observation model

We used a random walk one-dimensional Kalman filter to model the observed set of AdA values ($\{x_t\}$), as described by:

$$z_{t+1} = z_t + w_t,\ w_t \sim \mathcal{N}(0, q), \tag{4}$$
$$x_t = z_t + v_t,\ v_t \sim \mathcal{N}(0, r). \tag{5}$$

This Kalman filter was implemented in `Python`, and applied to all the AdA series in our studied dataset.

2.4 Dynamics

We model advertising awareness as the position of a damped harmonic oscillator and investment as an external variable force applied on it. The model seeks to create a response curve similar to the one proposed by Brown [4]  a quick bump, followed by an exponential decay.

Symbolically:

$$\frac{d}{dt}\begin{bmatrix} z \\ z' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -b^2 & -2a \end{bmatrix}\begin{bmatrix} z \\ z' \end{bmatrix} + \begin{bmatrix} 0 \\ f(t-\tau)+e+b^2r \end{bmatrix} \tag{6}$$

Where b is the stiffness of the system. a is the damping constant. r is the resting position of the spring - the percentage of the population aware of the brand when no investment is made. e is the equity force - a force that pushes people to remember the brand even if there is no investment being made. $f(t-\tau)$ is a function that represents the investment over time, it is represented as a force pushing the system closer to a desirable value of awareness. There is a delay parameter τ to account for the delay between an investment being made and its response being measured.

The forcing function, $f(t)$, was constructed by fitting a sum of an arbitrary number of Gaussian functions to the investment in advertising series:

$$f(t) = \Sigma_i^n g_{\mu_i,\sigma_i,k_i}(t), \tag{7}$$

where n is the number of campaigns, $g_{\mu_i,\sigma_i,k_i}(t)$ is a Gaussian function with magnitude k_i, mean μ_i and standard deviation σ_i. We chose a Gaussian function because it fitted the shape of campaigns -which occur in bursts- very well, and also because of its numerical benefits.

2.5 Campaign efficiency

The impact of advertisement campaigns certainly depends on how much money we invest, but also in how good - or persuasive the campaign actually is. To reflect that campaigns with similar levels of investment can generate quite different impacts, a creative efficiency factor for each campaign was added. This factor is implemented as a linear scaling factor applied to the initial campaign-specific forcing function:

$$f(t) = \Sigma_i^n \text{eff}_i g_{\mu_i,\sigma_i,k_i}(t). \tag{8}$$

At the end of the training process we have information about the internal dynamics of AdA, the delay factor between an investment and its reflection into awareness and the relative efficiency of the investment campaigns present in our dataset.

The full modeling process is schematized in the diagram presented in Figure 1.

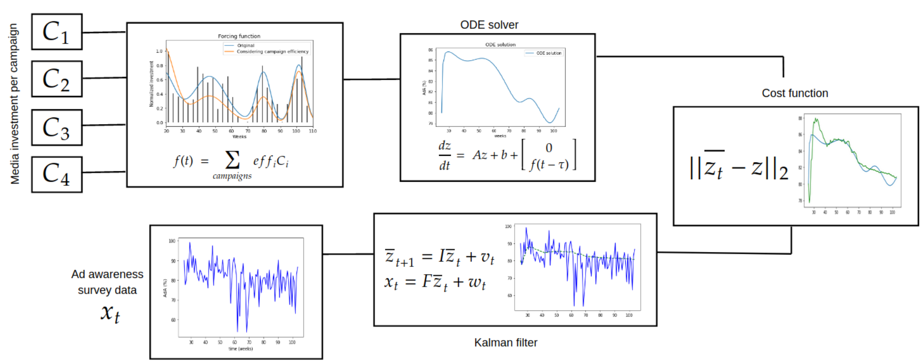


Figure 1: Schematized representation of the modeling process

3 Empirical verification

3.1 Data

We trained our model using data belonging to two big FMCG brands in France. We had access to weekly total media investment data for both brands. We normalized these time-series by dividing all quantities by the maximum amount of money spent in one week per brand.

Our measure of AdA is the proportion of people who answered affirmatively to the question "Have you seen some ad from *brand* in the last month?". This proportion was measured through a survey carried out continuously from 2016 to 2019 providing weekly estimates. The survey was conducted online with an average weekly sample size of 55 people.

3.2 Results

The training pipeline was implemented in `Pytorch`. `Torchdiffeq` was used to fit the parameters of the ODE system. The Kalman filter was implemented as a new layer using tensor algebra and Autograd.

We initialized the parameters using our a priori knowledge of the system. Namely, the order of magnitude of the observational variance and the system's initial position, the damping constant and stiffness of the system were initialized in such a way that the system represented an overdamped oscillator. We trained both systems for $6,000$ iterations.

In Table 1 we present the estimated parameters of our model. In Figure 2 we show the results of the modeling process.

Table 1: Trained parameters and final loss

Parameter	Stiffness (b)	Damping constant (a)	Equity force (e)	AdA resting position (r)	Response delay(τ)	Relative campaign efficiency (eff)
Brand A	0.91	0.14	1.09	0.13	0.40	1.51, 0.57, 0.50, 0.89
Brand B	0.56	0.16	1.01	0.04	8.04 $\times 10^{-5}$	0.49, 0.51

Parameter	State evolution variance (q)	Observational variance (r)	Final loss
Brand A	2.75 $\times 10^{-7}$	0.01	2.01
Brand B	1.05 $\times 10^{-5}$	0.01	1.81

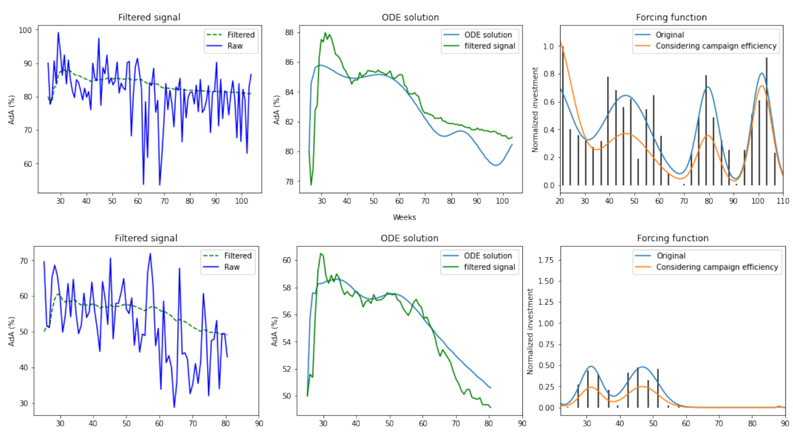


Figure 2: Trained solution for each brand. Brand A on top, brand B on bottom. From left to right: Raw signal and trained Kalman Filter, filtered signal and solution of the trained ODE system, and investment over time.

4 Conclusions and future work

We presented a fully interpretable learning architecture capable of describing the dynamics between media investment and advertising awareness. Each of the parameters of the model has a specific physical meaning that can be translated directly to a marketing context.

While conceived in the context of market research, the model is general enough to express relationships in other domains given that the response function can be approximated by a damped harmonic oscillator. The inclusion of a generic trainable observational filter allows us to model a variety of measurement mechanisms.

Neural analogs to the Kalman filter have been proposed over the last few years [14][9][23][2]. In further work we would like to include our knowledge of the system dynamics in the filtering process. We would also like to explore more general expressions of the ODE system.

Broader Impact

This research note presents a methodology for estimating advertisement campaign efficiencies at scale. These efficiencies can be used to assess the quality of those campaigns, and to develop more impactful advertisement strategies.

References

[1] Aravindakshan, A., & Naik, P. A. (2011). How does awareness evolve when advertising stops? The role of memory. *Marketing Letters*, 22(3), 315-326.

[2] Becker, P., Pandya, H., Gebhardt, G., Zhao, C., Taylor, J. & Neumann, G. (2019). Recurrent kalman networks: Factorized inference in high-dimensional deep feature spaces. arXiv preprint arXiv:1905.07357.

[3] Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (Eds.). (2011). *Measurement errors in surveys* (Vol. 173). John Wiley & Sons.

[4] Brown, G. (1986). Modelling Advertising Awareness. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 35(2), 289-299. doi:10.2307/2987534.

[5] Chen, R., Rubanova, Y., Bettencourt, J. & Duvenaud, D. (2018). Neural ordinary differential equations. *Advances in neural information processing systems* (pp. 6571-6583).

[6] Dupont, E., Doucet, A., & Teh, Y. W. (2019). Augmented neural odes. In *Advances in Neural Information Processing Systems* (pp. 3140-3150).

[7] Ezimadu, P. E., & Nwozo, C. R. (2018). Modeling dynamic cooperative advertising in a decentralized channel. *Yugoslav Journal of Operations Research*, 28(4), 539-566.

[8] Haber, E. & Ruthotto, L. (2017). Stable Architectures for Deep Neural Networks. *Inverse Problems*. 34. 10.1088/1361-6420/aa9a90.

[9] Haarnoja, T., Ajay, A., Levine, S. & Abbeel, P. (2016). Backprop KF: Learning Discriminative Deterministic State Estimators. *Advances in Neural Information Processing Systems* pp. 4376–4384.

[10] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[11] Joseph, J. V. (2006). Understanding advertising Adstock transformations. Available at SSRN 924128.

[12] Kalman, R. (1960) A New Approach to Linear Filtering and Prediction Problems. *Trans. ASME, Journal of Basic Engineering*, pp. 35-45, March 1960.

[13] Keller, K. L., Parameswaran, M. G., & Jacob, I. (2011). *Strategic brand management: Building, measuring, and managing brand equity*. Pearson Education India.

[14] Krishnan, R., Shalit, U. & Sontag, D. (2015) Deep Kalman Filters. ArXiv, abs/1511.05121.

[15] Lu, Y., Zhong, A., Li, Q., & Dong, B. (2018, July). Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. In *International Conference on Machine Learning* (pp. 3276-3285). PMLR.

[16] Ma, S., He, Y., & Gu, R. (2020). Dynamic generic and brand advertising decisions under supply disruption. *International Journal of Production Research*, 1-25.

[17] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... & Lerer, A. (2017). Automatic differentiation in pytorch.

[18] Romaniuk, J., Sharp, B., Paech, S., & Driesener, C. (2004). Brand and advertising awareness: a replication and extension of a known empirical generalisation. *Australasian Marketing Journal* (AMJ), 12(3), 70-80.

[19] Sutherland, M., & Friedman, L. (2000). Do you model ad awareness or advertising awareness?. *Journal of Advertising Research*, 40(5), 32-36.

[20] Vaughan, K., Beal, V., & Romaniuk, J. (2016). Can Brand Users Really Remember Advertising More Than Nonusers?: Testing an Empirical Generalization Across Six Advertising Awareness Measures. *Journal of Advertising Research*, 56(3), 311-320.

[21] Wang, X., Li, F., & Jia, F. (2020). Optimal Advertising Budget Allocation across Markets with Different Goals and Various Constraints. *Complexity*, 2020.

[22] Welch, G. & Bishop, G. (1995) An Introduction to the Kalman Filter *University of North Carolina, Department of Computer Science*, TR 95-041.

[23] Wilson, R., & Finkel, L. (2009). A neural implementation of the Kalman filter. In *Advances in neural information processing systems* (pp. 2062-2070).