YAHOO! RESEARCH

Time-Aware Prospective Modeling of Users for Online Display Advertising

Djordje Gligorijevic, Jelena Gligorijevic and Aaron Flores

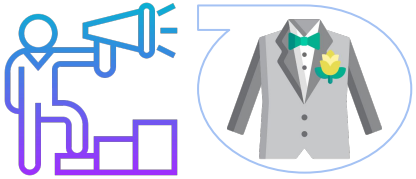
Presented by: Djordje Gligorijevic

verizon media

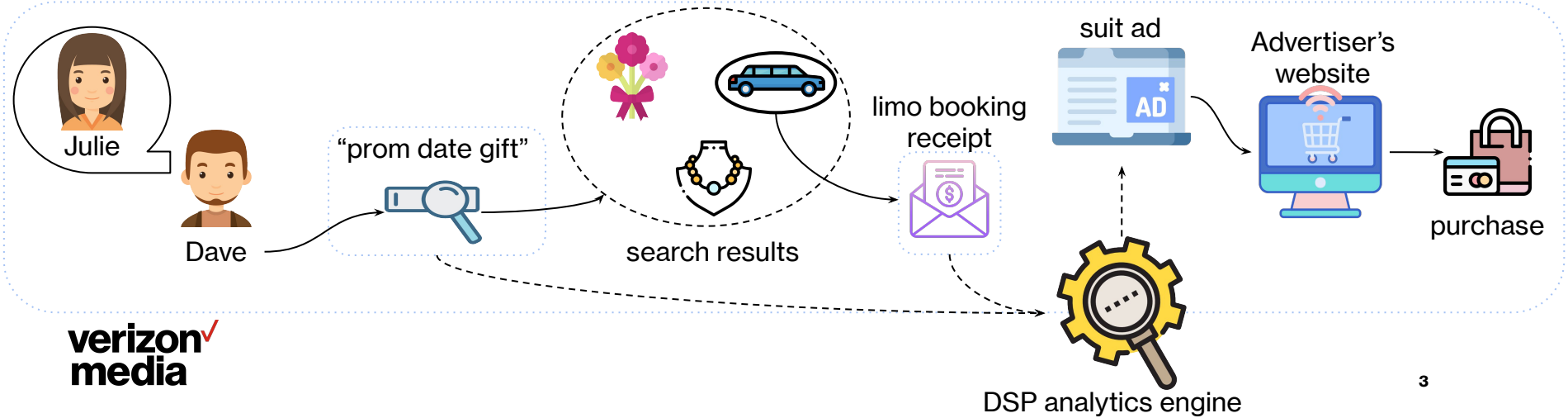
Prospective Display Advertising Introduction

verizon media

Prospective display advertising

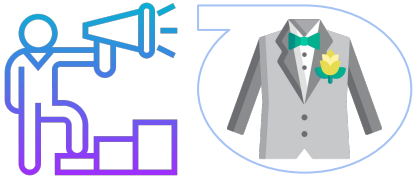


Retail adv. running a **prospecting** man suits sale campaign

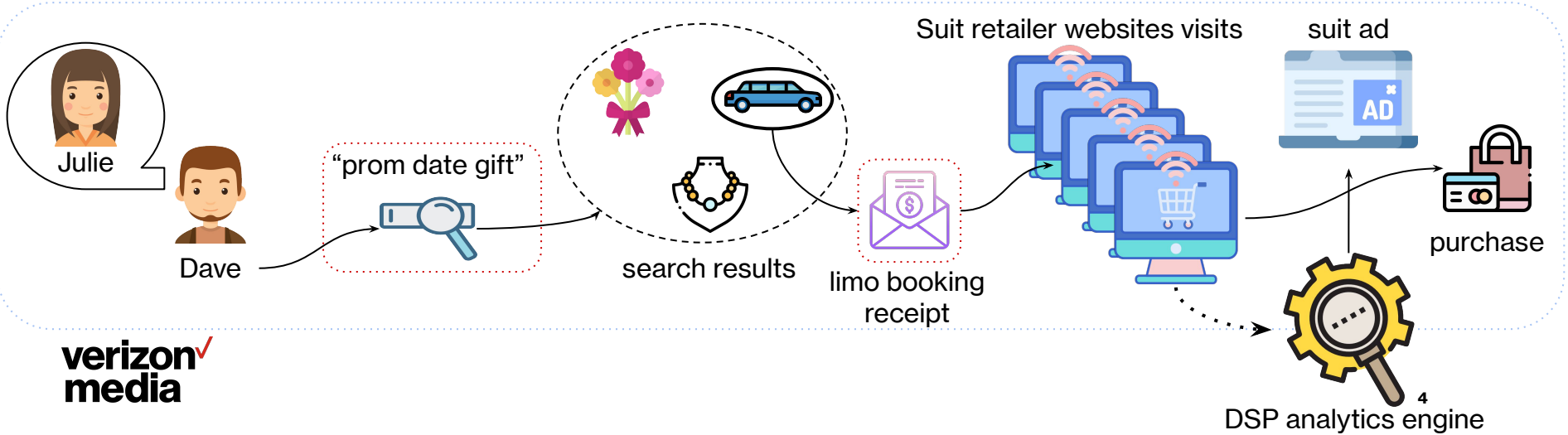


verizon media

Prospective display advertising - Reality



Retail adv. running a **prospecting** man suits sale campaign



verizon media

Problem statement

verizon media

Problem definition

Challenge: More and more advertisers are *interested* in prospective advertising while current systems tend to *underperform* there.

Problem: Powerful signals often referred as *retargeting events* overwhelm predictive systems

Conversion	Adv. site visit	Site visit prior to conv.	Percentage
TRUE	FALSE	FALSE	0.01%
TRUE	TRUE	FALSE	0.02%
TRUE	TRUE	TRUE	99.97%

Table 1: Percentages of conversions with respect whether advertisers' site visit (a retargeting) event occured, and if it occurred before the conversion or not.

- A simple rule based system can achieve Recall of 99.97% (on this retail advertiser example)
- Thus a few retargeting events can dominate over many other useful events
- Particularly noticeable for retail advertisers audiences

Proposed solution

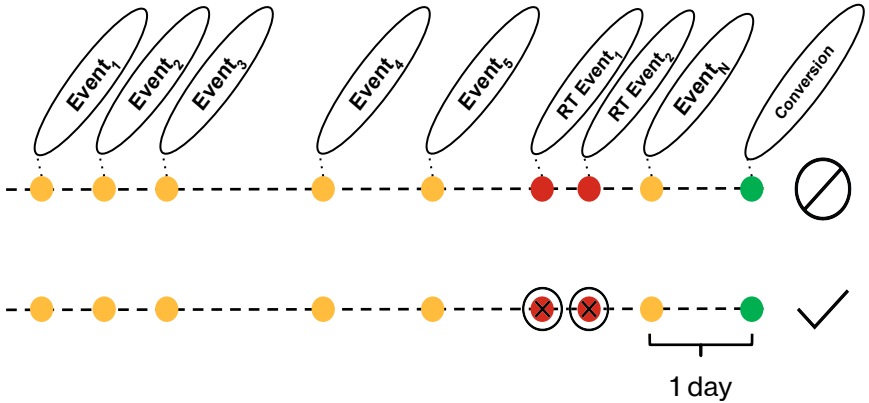
The idea:
99.97% of all conversions are coming from retargeting users - observed data should be altered

Dataset generation:

For each user, generate ***events sequence*** and remove all ***known*** retargeting events up to each conversion

Modeling goals:

to design more powerful models that can capture ***early usefull signals*** becomes a neccessity



verizon media

Data

verizon media

Dataset illustrated

Dataset: User activities collected in a chronological order

Canonicalized and normalized activities are derived from heterogeneous sources:

- Yahoo Search,
- Yahoo and AOL Mail receipts,
- Content reads on publisher's webpages such are Yahoo and AOL news, HuffPost, TechCrunch, Tumblr, etc.,
- Advertising data from Yahoo Gemini and Verizon Media DSP,
- Flurry mobile analytics,
- Conditional data from all advertisers (e.g., ad impressions, conversions, and advertiser site visits).

Final data product is a sequence of activities with a timestamp

verizon media

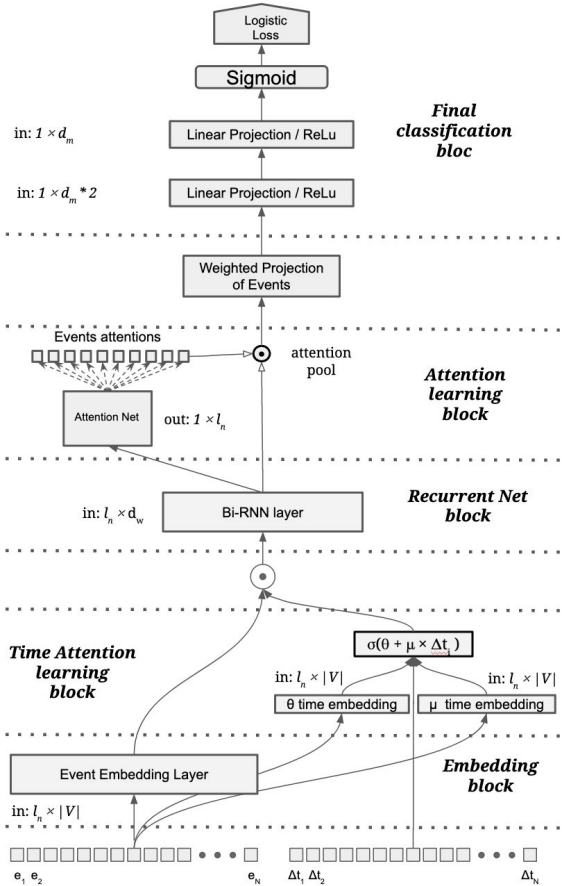
Proposed Approach

verizon media

Proposed approach: Deep Time-Aware conversIoN model DTAIN

Architecture

- DTAIN takes 2 sets of inputs: events and timesteps
- Consists of **5 blocks**: embedding, recurrent, two attention and a classification block
- **Temporal Attention** captures differences between event occurrence and inference timestamp through *mu* and *theta* parameters

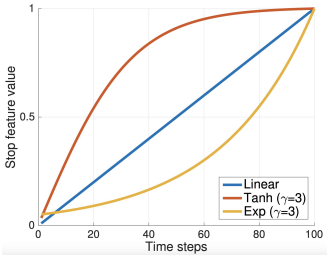


verizon media

Dj. Gligorijevic, J. Gligorijevic and A. Flores "Time-Aware Prospective Modeling of Users for Online Display Advertising", AdKDD 2019

Temporal Modeling in Deep Learning

- Temporal information is most frequently modeled as a **decay function**, though:
  - **Stop features** [1]
    - **Linear** $v_t = \frac{t}{T}$
    - **Tanh** $v_t = tanh\left(\gamma * \frac{t}{T}\right) + 1 - tanh(\gamma)$
    - **Exp** $v_t = exp\left(\gamma * \frac{t-T}{T}\right)$



  - **Attention regularization** [2] (where Δ_t is time gap between event and prediction time:

$$a_i = \frac{exp(v_i^T u - \lambda \Delta_t)}{\sum_t exp(v_t^T u - \lambda \Delta_t)}$$

  - **Attention modeling** using the temporal signal [3, 4] by handcrafting **time features** $A_j(\Delta t_i)$

$$a_i = Softmax\left(\sum_{j=1}^{k} p_{j,i} A_j(\Delta t_i)\right)$$

verizon media

Dj. Gligorijevic, J. Gligorijevic and A. Flores "Time-Aware Prospective Modeling of Users for Online Display Advertising", AdKDD 2019

Temporal Modeling in Deep Learning

- **Proposed approach** is motivated by *Euler's forward method* of solving *linear dynamic systems* [5]

$$\Delta_t = \tau_{e_j} - \tau_{e_i}$$
$$\delta(e_i, \Delta_t) = S(\theta_{e_i} - \mu_{e_i}\Delta_t)$$
$$S(x) = \frac{1}{1 + e^{-x}}$$

- Learns **event-specific impact** onto prediction [4]
- Single dimensional *learnable parameters:*
  - theta is the ***initial impact*** of the event
  - mu is ***temporal change*** of the event
  - Final impact of the event is *scaled* to 0-1 scale using Sigmoid function

- The larger theta and the smaller mu -> the greater impact does the event have onto prediction
- The closer to 0 they are -> the smaller initial and/or temporal impact the event has

Experimental Evaluation

verizon media

Experimental setup

The proposed DTAIN model was evaluated on two datasets and against 4 competitive baselines

Datasets:

1) Proprietary Verizon Media dataset of a single retail advertiser
   - 788,551 users in train and 196,830 in test set, downsampled to obtain ~7.5% positives
2) Public youchoose.com dataset from RecSys 2015 challenge
   - 1,965,359 sessions in train and 279,999 in test set, downsampled to obtain ~11% positives

Baselines:

1) CNN
2) GRU
3) GRU + Attention layer
4) GRU + Self Attention layer

Experimental results: Proprietary VerizonMedia dataset

- **Verizon Media dataset:**
  - 985,381 user sessions, 74,407 conversions
  - **long-time** sequences of activities
  - **prediction task**: to predict if a user is going to convert for the given advertiser (binary classification task)

	ROC AUC	Accuracy	Precision	Recall	Bias
CNN	0.8806	0.8110	0.2457	0.7871	1.0161
GRU	0.9018	0.8520	0.3004	0.7972	1.1983
GRU+Attn	0.8968	0.8438	0.2882	0.7982	0.8047
GRU+SelfAttn	0.8804	0.8364	0.2743	0.7756	0.9273
DTAIN	**0.9263**	**0.8602**	**0.3219**	**0.8537**	**0.9871**

Table 3: Performance metrics on the proprietary user trails dataset for all algorithms.

- **The proposed DTAIN model outperforms other baselines** on the conversion prediction task w.r.t. ROC AUC, Accuracy, Precision, Recall and Bias
- **Improvements** over all baselines are prominent thanks to the long-time sessions (>100 days)

Experimental results: Proprietary VerizonMedia dataset, contd.

- **Verizon Media dataset:**
  - 985,381 user sessions, 74,407 conversions
  - **long-time** sequences of activities
  - **prediction task**: to predict if a user is going to convert for the different conversion rules given by the advertiser (multi-class classification task)

- Due to class disbalance that occurs when splitting the binary into multi-classification task we report PRC-AUC

- **The proposed DTAIN model outperforms other baselines** on the majority of metrics

	PRC AUC	Accuracy	Precision	Recall	Bias
Task 0					
CNN	0.9880	0.8139	0.9810	0.8153	1.0069
GRU	0.9896	0.8544	0.9821	0.8588	1.0030
GRU+Attn	0.9907	0.8511	0.9837	0.8537	0.9933
GRU+SelfAttn	0.9877	0.8456	0.9795	0.8515	0.9941
DTAIN	**0.9926**	**0.8613**	**0.9876**	**0.8614**	**0.9982**
Task 1					
CNN	0.2523	0.9602	0.3161	0.2506	0.8836
GRU	0.2711	0.9629	0.3635	0.2720	**0.9715**
GRU+Attn	**0.3013**	0.9630	0.3788	**0.3139**	1.1163
GRU+SelfAttn	0.2452	0.9606	0.3277	0.2648	1.0645
DTAIN	*0.2880*	**0.9652**	**0.4000**	0.2539	1.0680
Task 2					
CNN	0.2495	0.9584	0.3287	0.2419	0.7588
GRU	0.2567	0.9597	0.3485	0.2464	0.8849
GRU+Attn	0.2374	0.9584	0.3355	**0.2582**	1.0453
GRU+SelfAttn	0.2081	0.9587	0.3081	0.1951	**0.9887**
DTAIN	**0.2776**	**0.9633**	**0.4083**	0.2348	0.9460

Table 4: Performance metrics on the proprietary user trails dataset for different tasks.

verizon media

Interpretability analysis of the DTAIN model

On a dataset with 500 conversions and 500 last events in each trail we analyze attentions:

Figures (a) and (b) display attentions of GRU+Attn and DTAIN model:

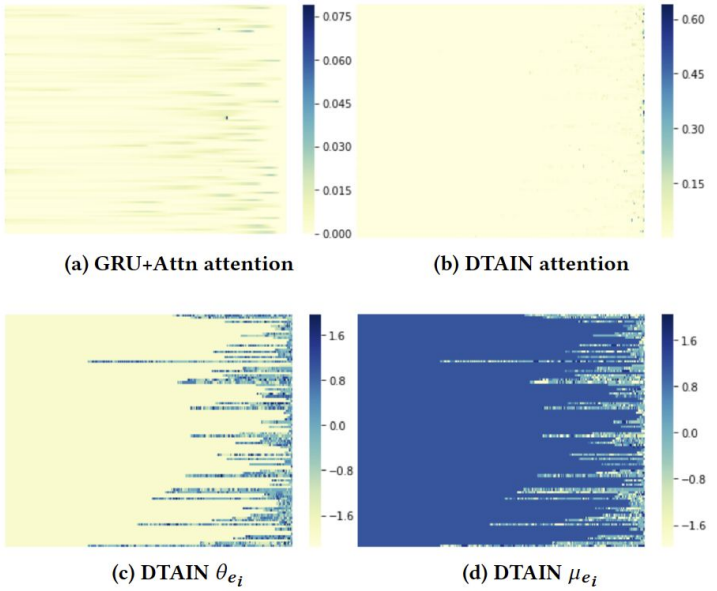
- GRU+Attn looks on events mostly in the latter half
- DTAIN shows interesting pattern where it only focuses to last few events.

Analyzing temporal attention signals for theta (c) and mu (d):

- events both near and far from conversion are exploited

We suspect that the temporal-attention has captured the impacts of each event thus by biRNN modeling the information was compressed in last few event positions.

Figure 4: Heat maps of events attentions scores for 100 randomly sampled converters



(a) GRU+Attn attention (b) DTAIN attention

(c) DTAIN θ_{e_i} (d) DTAIN μ_{e_i}

verizon media

Experimental results: Public RecSys 2015 challenge dataset

- **Youchoose.com dataset:**
  - 2,245,358 sessions, 241,887 buys
  - **short-time** sequences of activities
  - **prediction task**: to predict if a session is going to end in purchase (binary classification task)

	ROC AUC	PRC AUC	Accuracy	Precision	Recall
CNN	0.7534	0.2870	0.6779	0.2087	0.7041
GRU	0.7504	0.2725	0.6958	0.2142	0.6746
GRU+SelfAttn	0.7029	0.2391	0.6734	0.1907	0.6184
GRU+Attn	0.7639	0.2973	**0.6997**	**0.2195**	0.6904
DTAIN	**0.7666**	**0.3019**	0.6943	0.2186	**0.7047**

- **The proposed DTAIN model outperforms other baselines** on the purchase prediction task w.r.t. ROC AUC, PRC AUC and Recall and is comparable to the second best baseline w.r.t. Accuracy and Precision.
- **Improvements** over GRU + Attention model are expectedly **smaller** (short sessions)
- However, **adding temporal information helps**, as it aslo models **initial impact** of the events to the conversion, thus providing additional information to the classifier.

Next steps

1. Analyze different dataset generation strategies
2. Predict first occurence of retargeting events
3. Design regularization techniques that act on events highly associated with the target
4. Extend model optimization through labeling such events as adversarial ones

Dj. Gligorijevic, J. Gligorijevic and A. Flores “Time-Aware Prospective Modeling of Users for Online Display Advertising”, AdKDD 2019

References

[1] Pei W, Tax DM. Unsupervised Learning of Sequence Representations by Autoencoders. arXiv preprint arXiv:1804.00946. 2018 Apr 3.

[2] S. K. Arava, C. Dong, Z. Yan, A. Pani, et al. Deep neural net with attention for multi-channel multi-touch attribution. arXiv preprint arXiv:1809.02230, 2018

[3] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M Dai, Nissan Hajaj, Peter J Liu,Xiaobing Liu, Mimi Sun, Patrik Sundberg, Hector Yee, et al.2018. Scalable and accurate deep learning for electronic health records.arXiv preprint arXiv:1801.07860 (2018).

[4] T. Bai, S. Zhang, B. L. Egleston, and S. Vucetic. Interpretable representation learning for healthcare via capturing disease progression through time. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 43–51. ACM, 2018

[5] X. H. Cao, C. Han, and Z. Obradovic. Learning a dynamic-based representation for multivariate biomarker time series classifications. In 2018 IEEE International Conference on Healthcare Informatics (ICHI), pages 163–173. IEEE, 2018

Q&A

verizon media