2D-ATT: Causal Inference for Mobile Game Organic Installs with 2-Dimensional Attentional Neural Network

IEEE BIG DATA 2020

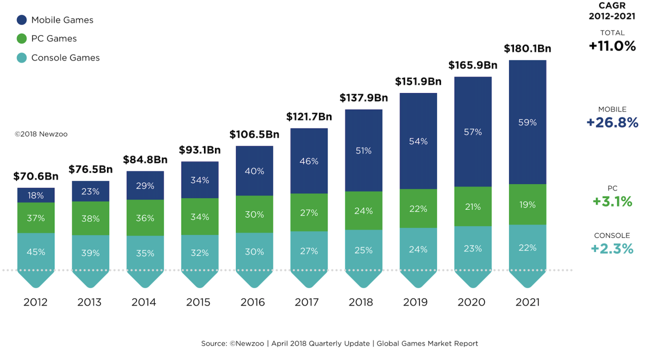
Boxiang Dong[1] Hui (Bill) Li[2] Yang (Ryan) Wang[2] Rami Safadi[2]

[1]Montclair State University
Montclair, NJ

[2]Jam City
Los Angeles, CA

December 11, 2020

Mobile Game Industry



- By 2021, the gaming market is projected to reach $180 billion.
- Mobile games take a large fraction.

User Acquisition

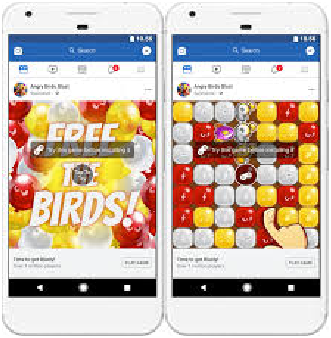
The intense competition in the industry exhorts the importance of user acquisition (UA) in mobile gaming operations.

User Acquisition

A new app download must fall into one of the two categories:

Paid Install obtained by advertising

Organic Install cannot be attributed to any advertisement source.



Objective

Organic installs are of crucial value for a mobile game's ecosystem.

- no upfront UA cost
- more loyal and active

Objective Understand the driving forces of organic installs

Application Intelligent UA budget allocation

Challenges

Challenge I many potential causal factors

- game quality
- app visibility
- in-game social referrals

Challenge II temporal lags

- delay between game improvement and the growth of organic installs

Limitations of Existing Work

Statistic Inference [XHD+20, CF18]

Definition 1 (Granger causality[Granger, 1969]) *Suppose we have a stationary sequence of random variables $\{(X_t, Y_t)\}$ $(t \in \mathbb{N})$, where X_t and Y_t are on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let S_X and S_Y be observations of $\{X_1, \cdots, X_t\}$ and $\{Y_1, \cdots, Y_t\}$, respectively.*

Granger causality defines $\{X_t\}$ as the cause of $\{Y_t\}$ if

$$P(Y_{t+1}|S_X, S_Y) \neq P(Y_{t+1}|S_Y)$$

and states that $\{X_t\}$ is not the cause of $\{Y_t\}$ if

$$P(Y_{t+1}|S_X, S_Y) = P(Y_{t+1}|S_Y) \tag{1}$$

- rigid assumption on data (e.g., no confounder, additive causal impacts)
- no temporal lag

Model Explanation [PJS13, RSG16, LL17, SGK17]

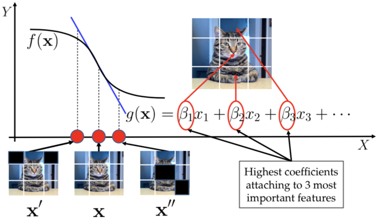
Attention Mechanism [SMK19, GLAF19]

Limitations of Existing Work

Statistic Inference [XHD+20, CF18]

- rigid assumption on data (e.g., no confounder, additive causal impacts)
- no temporal lag

Model Explanation [PJS13, RSG16, LL17, SGK17]



- no temporal pattern
- low fidelity

Attention Mechanism [SMK19, GLAF19]

Limitations of Existing Work

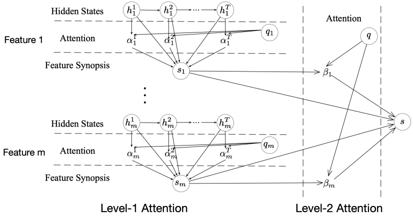
Statistic Inference [XHD+20, CF18]

- rigid assumption on data (e.g., no confounder, additive causal impacts)
- no temporal lag

Model Explanation [PJS13, RSG16, LL17, SGK17]

- no temporal pattern
- low fidelity

Attention Mechanism [SMK19, GLAF19]



- The importance of different features at different temporal lags are not comparable.

Contributions

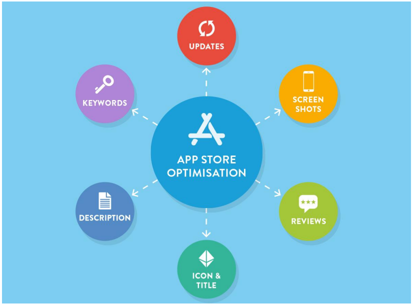
We are the first to quantitatively evaluate the causal impact of different factors with temporal lags on organic installs in the mobile game industry.

- a deep recurrent neural network to learn the dynamics of each feature;
- an innovative attention mechanism to learn the importance of each feature at various temporal lags; and
- experimental validation of the effectiveness.

Outline

Preliminaries: Organic Installs

- Organic installs refer to the scenario where users install the app without directly responding to any advertising campaign.
- Three widely recognized driving factors:
  - App store optimization.



  - Encouraging social exposure.
  - Game quality improvement.

Preliminaries: Organic Installs

- Organic installs refer to the scenario where users install the app without directly responding to any advertising campaign.
- Three widely recognized driving factors:
  - App store optimization.
  - Encouraging social exposure.

  - Game quality improvement.

Preliminaries: Causal Inference

Counterfactual Inference

- Compare outcomes in two almost-identifical worlds.
- Controlled experiments are expensive, unethical, and even impossible.

Observational Inference

- Infer the causal structure of the data generation system from observational data.

Problem Formulation

Input

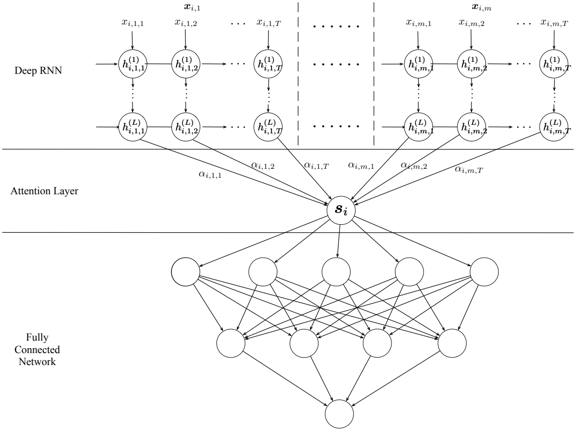
$$
\text{Features}\left\{
\begin{array}{cccccc}
x_{1,1} & x_{1,2} & \cdots & x_{1,t} & \cdots & x_{1,T} \\
x_{2,1} & x_{2,2} & \cdots & x_{2,t} & \cdots & x_{2,T} \\
\vdots & \vdots & & \vdots & & \vdots \\
x_{j,1} & x_{j,2} & \cdots & x_{j,t} & \cdots & x_{j,T} \\
\vdots & \vdots & & \vdots & & \vdots \\
x_{m,1} & x_{m,2} & \cdots & x_{m,t} & \cdots & x_{m,T}
\end{array}
\right.
$$

$$\xrightarrow{\hspace{10em}} \text{Time}$$

Target y

Output The causal impact of $\{x_{j,t}\}$ on y

Solution in a Nutshell

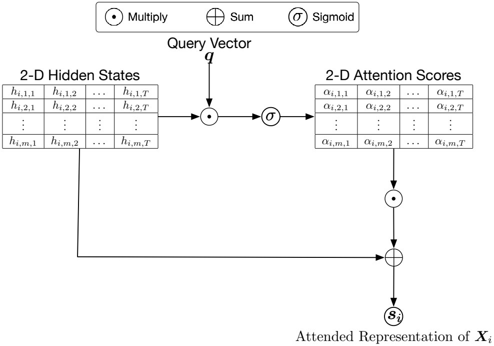


Deep RNN Capture temporal patterns

2D-ATT Reveal information flow in the generative process

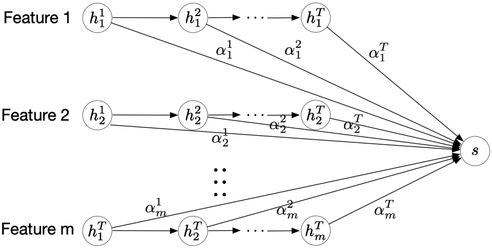
FCN Predict target value

2D-ATT



- Attention score $\alpha_{i,j,t} = \frac{\boldsymbol{u}_{i,j,t}\boldsymbol{q}^\intercal}{\sum_{1\leq j\leq m, 1\leq t\leq T} exp(\boldsymbol{u}_{i,j,t}\boldsymbol{q}^\intercal)}$
- Query vector q represents the fixed question "*what are the important hidden states with regard to the prediction of the target variable*" and is to be optimized
- Attended representation $\boldsymbol{s}_i = \sum_{1\leq j\leq m, 1\leq t\leq T} \alpha_{i,j,t} h_{i,j,t}$

2D-ATT



- The attention mechanism resembles the information flow in the generative process of the target variable.
- The attention scores represent the contribution of each feature at every time step to the target.
- By steering the prediction model to the most important input fragments, it can improve the prediction accuracy in return.

Experiments: Synthetic Datasets

Synthetic 1 Linear, Uniform with instantaneous effects

Synthetic 2 Linear, Gaussian with instantaneous effects

Synthetic 3 Nonlinear, non-Gaussian without instantaneous effects

$$
\begin{aligned}
X_t &= 0.8 * X_{t-1} + 0.3 * N_{X,t} \\
Y_t &= 0.4 * Y_{t-1} + (X_{t-1} - 1)^2 + 0.3 * N_{Y,t} \\
Z_t &= 0.4 * Z_{t-1} + 0.5 * cos(Y_{t-1}) + sin(Y_{t-1}) + 0.3 * N_{Z,t},
\end{aligned}
$$

where $N_{\cdot,t} \sim \mathcal{U}([-0.5, 0.5])$. Z is regarded as the target variable.

Therefore, $Y_{t-1} \rightarrow Z_t$ is the ground truth causal effect.

Synthetic 4 Non-additive interaction

Experiments: Synthetic Datasets

Approach	Synthetic 1	Synthetic 2	Synthetic 3	Synthetic 4
T-Causality [30]	N/A	N/A	N/A	N/A
G-Causality [20]	N/A	N/A	N/A	N/A
SHAP [15]	$X_{t-3}, X_t, X_{t-5}, X_{t-2} \to Y_t$ ✓	$Y_{t-1} \to Z_t$ ✓ $Y_t, W_{t-2}, W_{t-1}, Y_{t-2} \to Z_t$ ✗	$Y_{t-1} \to Z_t$ ✓ $Y_t \to Z_t$ ✗	$X_{t-1}, X_{t-2} \to Y_t$ ✓
AME [23]	N/A	$W, Y \to Z$ ✓	$X \to Z$ ✓ $Y \to Z$ ✗	N/A
2D-ATT	$X_t, X_{t-2}, X_{t-5} \to Y_t$ ✓	$W_t \to Z_t$ ✓ $W_{t-1}, W_{t-2} \to Z_t$ ✗	$Y_{t-1} \to Z_t$ ✓	$X_{t-1}, X_{t-2} \to Y_t$ ✓

TABLE I

CAUSAL EFFECTS DISCOVERED BY ALL APPROACHES ON THE SYNTHETIC DATASETS

(✓denotes a true-positive causal effect, ✗denotes a false-positive causal effect)

Observations:

- 2D-ATT is capable of identifying complex causal effects at various levels of temporal delays.
- SHAP heavily depends on the accuracy of the prediction model.
- AME cannot deal with temporal lags.
- T- and G-Causality are vulnerable to observational noise.

Experiments: Real-world Dataset

- Provided by Jam City
- 9 mobile games across two platforms (iOS and Android)
- From December 2013 to June 2020
- Numerical features

  paid_install the number of paid users obtained;
  ua_cost the user acquisition cost;
  DAU daily active users;
  DAP daily active payers, i.e., users who make within-app purchases;

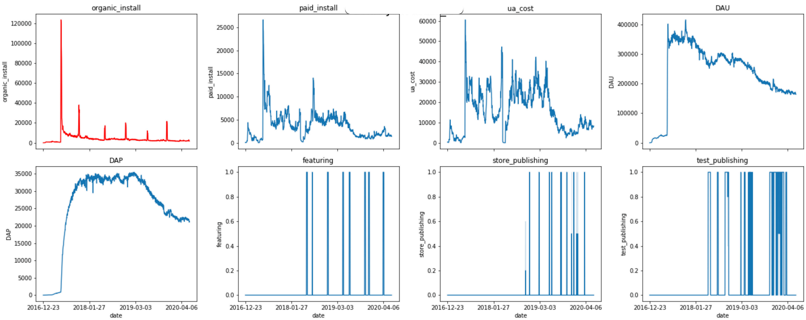
- Binary features

  featuring indicates whether there is an app store featuring event, i.e., if the app store displays the game in the main page;
  store_publishing indicates whether the game's information is updated in the app store;
  test_publishing indicates whether there is any testing for new game content in the app store;

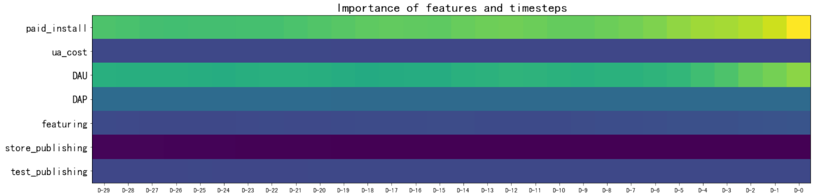
- Target variable

  organic_install the number of organic installs.

Experiments: Real-world Dataset



Experiments: Real-world Dataset



Observations:

- Paid installs and DAU are the most important causal factors, with closer statistics playing a more important role.
- We are the first to discover that paid installs are the most significant causal factor of organic installs.
- This demonstrates that the transformation from quantitative change (i.e., more paid users) to qualitative change (i.e., more organic users) also holds in the mobile game industry.

Conclusion

We present a novel attention mechanism to discover causal relationships with temporal lags from multivariate time series data.

In the future, we plan to

- design an intelligent UA budget allocation algorithm based on the discovered causal effects; and
- design a user attribution model for all the advertising channels.

References I

[BDF91] Peter J Brockwell, Richard A Davis, and Stephen E Fienberg.
Time series: theory and methods: theory and methods.
Springer Science & Business Media, 1991.

[CF18] Yoichi Chikahara and Akinori Fujino.
Causal inference in time series via supervised learning.
In *IJCAI*, pages 2042–2048, 2018.

[FM82] Jean-Pierre Florens and Michel Mouchart.
A note on noncausality.
Econometrica: Journal of the Econometric Society, pages 583–591, 1982.

[GFT+08] Arthur Gretton, Kenji Fukumizu, Choon H Teo, Le Song, Bernhard Schölkopf, and Alex J Smola.
A kernel statistical test of independence.
In *Advances in neural information processing systems*, pages 585–592, 2008.

[GLAF19] Tian Guo, Tao Lin, and Nino Antulov-Fantulin.
Exploring interpretable lstm neural networks over multi-variable data.
arXiv preprint arXiv:1905.12034, 2019.

[HT90] Trevor J Hastie and Robert J Tibshirani.
Generalized additive models, volume 43.
CRC press, 1990.

[LL17] Scott M Lundberg and Su-In Lee.
A unified approach to interpreting model predictions.
In *Advances in neural information processing systems*, pages 4765–4774, 2017.

References II

[PJS13] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf.
Causal inference on time series using restricted structural equation models.
In *Advances in Neural Information Processing Systems*, pages 154–162, 2013.

[RSG16] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin.
" why should i trust you?" explaining the predictions of any classifier.
In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[SGK17] Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje.
Learning important features through propagating activation differences.
arXiv preprint arXiv:1704.02685, 2017.

[SGSK16] Avanti Shrikumar, Peyton Greenside, Anna Shcherbina, and Anshul Kundaje.
Not just a black box: Learning important features through propagating activation differences.
arXiv preprint arXiv:1605.01713, 2016.

[ŠK14] Erik Štrumbelj and Igor Kononenko.
Explaining prediction models and individual predictions with feature contributions.
Knowledge and information systems, 41(3):647–665, 2014.

[SMK19] Patrick Schwab, Djordje Miladinovic, and Walter Karlen.
Granger-causal attentive mixtures of experts: Learning important features with neural networks.
In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4846–4853, 2019.

References III

[XHD+20] Haoyan Xu, Yida Huang, Ziheng Duan, Jie Feng, and Pengyu Song.
Multivariate time series forecasting based on causal inference with transfer entropy and graph neural network.
arXiv preprint arXiv:2005.01185, 2020.

Q & A

Thank you!

Questions?