Learning Multi-touch Conversion Attribution with Dual-attention Mechanisms for Online Advertising

Kan Ren, Yuchen Fang, Weinan Zhang,
Shuhao Liu, Jiajun Li, Ya Zhang, Yong Yu, Jun Wang*

Apex Data & Knowledge Management Lab
Shanghai Jiao Tong University
*University College London

CIKM, 2018

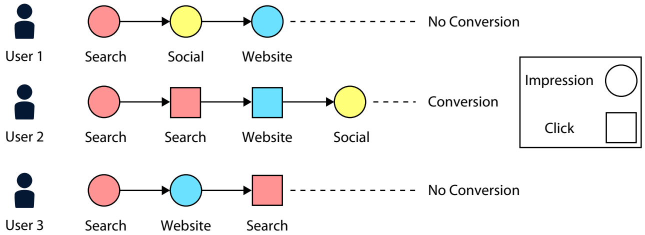
Outline

Problem Background

Figure: John Wanamaker

- John Wanamaker: "Half the money I spend on advertising is wasted; the trouble is I don't know which half."

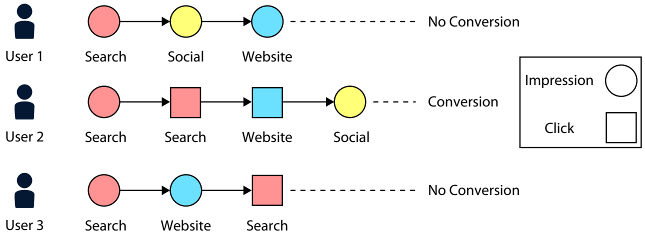
Problem Background



Two views of the problem

Sequence View: Touch point attributes positively/negatively to the conversion.
Channel View: Which channel appeals the user the most?

Problem Background



Two views of the problem

Sequence View: Touch point attributes positively/negatively to the conversion.
Channel View: Which channel appeals the user the most?

Problem

To analyze the *effects* of the **touch points** from different **channels** to the final user conversion.

Related Works

Rule-based Methods

	Model	Description	Growth strategy
	Last-click (DEFAULT)	Gives all credit for the conversion to the last-clicked keyword	Most conservative
	First click	Gives all credit for the conversion to the first-clicked keyword	Most growth-oriented
	Linear	Distributes the credit for the conversion equally across all clicks on the path	Moderate
	Time decay	Gives more credit to clicks that happened closer in time to the conversion	Conservative
	Position-based	Gives 40% of credit to both the first- and last-clicked keyword, with the remaining 20% spread out across the other clicks on the path	Growth-oriented
	Data-driven	Gives credit to clicked keywords based on how imperative they were in the conversion process	Based on account's performance

Too simple and heuristic, cannot help subsequent advertising strategy.

Google Ads: https://support.google.com/google-ads/answer/7002714?hl=en

Data Insights

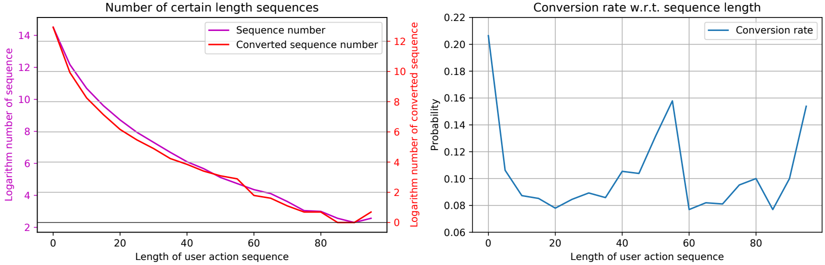


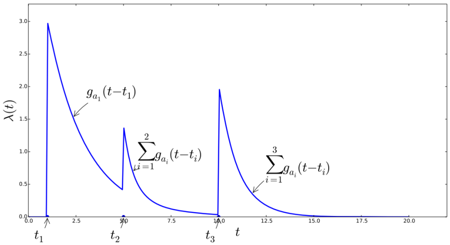
Figure: Left: Sequence length distribution; Right: CVR distribution against the sequence length.

- Longer behavior sequence $\nRightarrow$ higher conversion rate.
- Not all the ad touch points have additive positive influence, some may have counteractive effects.

Related Works

Data-driven Methods

- Logistic regression with learned coefficients for the attribution. [Shao et al. In KDD'11.]
- Additive point process to model the conversion rate over time and derive the attribution for each point. [Zhang et al. ICDM'16. Ji et al. CIKM'16, AAAI'17.]



Problem Challenge: Multi-touch Conversion Attribution

Cons of the traditional methods

- Rule-based methods are
  - heuristical
  - wrong to subsequence usage of attributed results
- Simple probability methods
  - predict upon single point
  - ignore sequential influence
- Consider only one type of user behaviors.

- Our Solution

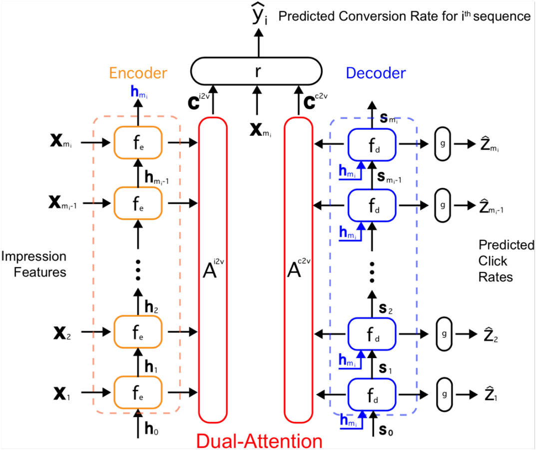
Our Solution

Attention-based Conversion Prediction

- Use **recurrent neural network** to model the sequential user activities.
- Learn to assign "**attention**" to the touch points to model the conversion *attributions*.
- Simultaneously model **impression-level** and **click-level patterns** for conversion estimation.

Dual-attention Mechanism for Conversion Attribution

Attention as Attribution Credits (cont.)



Conversion Rate Estimation

$$\hat{y}_i = p(y = 1 \mid \mathbf{x}, \mathbf{z}) = r(\boldsymbol{x}_{m_i}, \boldsymbol{c}^{i2v}, \boldsymbol{c}^{c2v})$$

Balance of Impression and Click

$$r(\boldsymbol{x}_{m_i}, \boldsymbol{c}^{i2v}, \boldsymbol{c}^{c2v}) = r_{\text{conv}}\left((1-\lambda)\cdot \boldsymbol{c}^{i2v} + \lambda \cdot \boldsymbol{c}^{c2v}\right)$$

$$\lambda = \frac{\exp\left[f_\lambda(\boldsymbol{x}_{m_i}, \boldsymbol{c}^{c2v})\right]}{\exp\left[f_\lambda(\boldsymbol{x}_{m_i}, \boldsymbol{c}^{i2v})\right] + \exp\left[f_\lambda(\boldsymbol{x}_{m_i}, \boldsymbol{c}^{c2v})\right]}$$

Attentional Combination

$$\boldsymbol{c}^{i2v} = \sum_{j=1}^{m_i} a_j^{i2v} \boldsymbol{h}_j$$

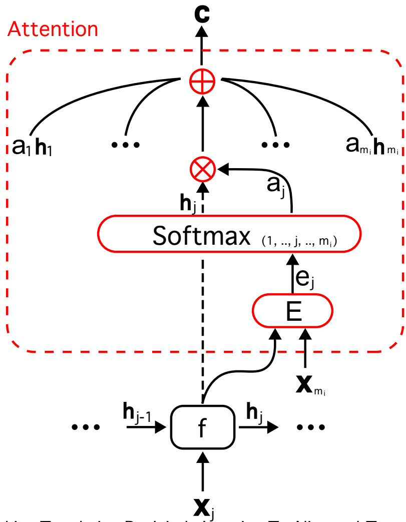
$$\boldsymbol{c}^{c2v} = \sum_{j=1}^{m_i} a_j^{c2v} \boldsymbol{s}_j$$

Attribution Calculation

$$\text{Attr}_j = (1-\lambda)\cdot a_j^{i2v} + \lambda \cdot a_j^{c2v}$$

Attention Implementation

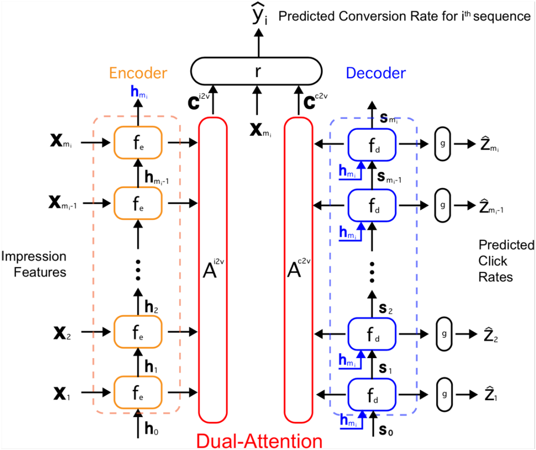
{ (Query, Key, Value) }



Dzmitry Bahdana et al. Neural Machine Translation By Jointly Learning To Align and Translate. ICLR 2015.

Dual-attention Mechanism for Conversion Attribution

Attention as Attribution Credits (cont.)



Conversion Rate Estimation

$$\hat{y}_i = p(y=1 \mid \mathbf{x}, \mathbf{z}) = r(\boldsymbol{x}_{m_i}, \boldsymbol{c}^{i2v}, \boldsymbol{c}^{c2v})$$

Balance of Impression and Click

$$r(\boldsymbol{x}_{m_i}, \boldsymbol{c}^{i2v}, \boldsymbol{c}^{c2v}) = r_{\text{conv}}\left((1-\lambda)\cdot \boldsymbol{c}^{i2v} + \lambda \cdot \boldsymbol{c}^{c2v}\right)$$

$$\lambda = \frac{\exp\left[f_\lambda(\boldsymbol{x}_{m_i}, \boldsymbol{c}^{c2v})\right]}{\exp\left[f_\lambda(\boldsymbol{x}_{m_i}, \boldsymbol{c}^{i2v})\right] + \exp\left[f_\lambda(\boldsymbol{x}_{m_i}, \boldsymbol{c}^{c2v})\right]}$$

Attentional Combination

$$\boldsymbol{c}^{i2v} = \sum_{j=1}^{m_i} a_j^{i2v} \boldsymbol{h}_j$$

$$\boldsymbol{c}^{c2v} = \sum_{j=1}^{m_i} a_j^{c2v} \boldsymbol{s}_j$$

Attribution Calculation

$$\text{Attr}_j = (1-\lambda)\cdot a_j^{i2v} + \lambda \cdot a_j^{c2v}$$

The Usage of Attribution

Attribution of the j-th Touch Point

$$\mathrm{Attr}_j = (1-\lambda) \cdot a_j^{i2v} + \lambda \cdot a_j^{c2v} \ . \tag{1}$$

- Now that we have obtained the attributed credits, what else can we do with it?
  - None of the related works consider the subsequent usage of the obtained attribution values.

The Usage of Attribution

Attribution of the j-th Touch Point

$$\mathrm{Attr}_j = (1-\lambda)\cdot a_j^{i2v} + \lambda \cdot a_j^{c2v} \ . \tag{1}$$

- Now that we have obtained the attributed credits, what else can we do with it?
  - None of the related works consider the subsequent usage of the obtained attribution values.

Example

To guide the subsequent **budget allocation** over the channels for the advertiser.

Back Evaluation for Attibution Guided Budget Allocation

Attribution Calculation for the k-th Channel (y_i: converted)

$$\mathrm{Attr}(c_k|y_i) = \sum_{j=1}^{m_i} \mathrm{Attr}_j \cdot 1(c_j = c_k) \tag{2}$$

Back Evaluation for Attibution Guided Budget Allocation

Attribution Calculation for the k-th Channel (y_i: converted)

$$\mathrm{Attr}(c_k|y_i) = \sum_{j=1}^{m_i} \mathrm{Attr}_j \cdot 1(c_j = c_k) \tag{2}$$

Inferred ROI of Channel (Sahin Cem Geyik et al. ADKDD'14)

$$\mathrm{ROI}_{c_k} = \frac{\sum_{\forall y_i=1} \mathrm{Attr}(c_k|y_i) \cdot V \cdot 1(y_i = 1)}{\text{Money spent on channel } c_k}, \tag{3}$$

Back Evaluation for Attibution Guided Budget Allocation

Attribution Calculation for the k-th Channel (y_i: converted)

$$\mathrm{Attr}(c_k|y_i) = \sum_{j=1}^{m_i} \mathrm{Attr}_j \cdot 1(c_j = c_k) \quad (2)$$

Inferred ROI of Channel (Sahin Cem Geyik et al. ADKDD'14)

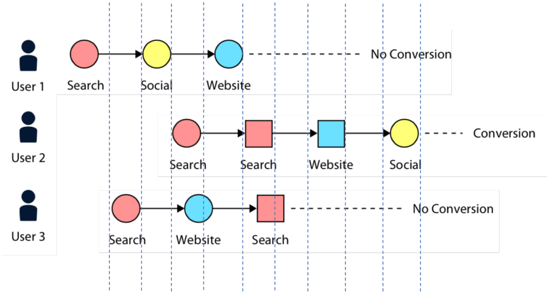
$$\mathrm{ROI}_{c_k} = \frac{\sum_{\forall y_i = 1} \mathrm{Attr}(c_k|y_i) \cdot V \cdot 1(y_i = 1)}{\text{Money spent on channel } c_k} , \quad (3)$$

Attribution guided Budget Allocation

For channel c_k, the allocated budget

$$b_k = \frac{\mathrm{ROI}_{c_k}}{\sum_{v=1}^{K} ROI_{c_v}} \times B . \quad (4)$$

Back Evaluation Flow



- Allocate the budget for each channel w.r.t. the calculated attributions of the model.
- Replay the history user behavior sequences according to the timestamp of each touch point, and judge
  - If the left budget of the channel has run out, then those sequence would be "*blocked*" and put into the **blacklist**;
  - If the replay has reached the tail of sequence, the result of (non-)conversion would be added for the model performance.

Experiments

Experimental Setup

Conversion Estimation

Given the user behavior sequence, compare the models on the measurement (AUC and Log-loss) for conversion rate estimation.

Attribution Guided Budget Allocation

After the back evaluation, compare different models over

- Conversion Number (CN)
- Conversion Rate (CVR)
- Profit $= V_0 \cdot \text{CN} - \sum \text{cost}$
- Cost-per-action (CPA) w.r.t. the models.

Datasets

- Miaozhen Dataset
  - Zhang et al. ICDM'16; Ji et al. CIKM'16, AAAI'17.
- Criteo
  - http://ailab.criteo.com/criteo-attribution-modeling-bidding-dataset/

Compared Settings

- **LR** is the Logistic Regression model [24].
- **SP** is a Simple Probabilistic model [7].
- **AH** (AdditiveHazard) model [37] using additive point process.
- **AMTA** is the Additional Multi-touch Attribution model [12] which was state-of-the-art.
- **ARNN** is the normal Recurrent Neural Network (i.e., only with encoder part) method.
- **DARNN** is our proposed model with dual-attention mechanism.

Conversion Estimation

Table: Conversion estimation results on two datasets. AUC: the higher the better; Log-loss: the lower the better.

Models	Miaozhen AUC	Miaozhen Log-loss	Criteo AUC	Criteo Log-loss
LR	0.8418	0.3496	0.9286	0.3981
SP	0.7739	0.5617	0.6718	0.5535
AH	0.8693	0.6791	0.6791	0.5067
AMTA	0.8357	0.1636	0.8465	0.3897
ARNN	0.8914	0.1610	0.9793	0.1850
DARNN	**0.9123**	**0.1095**	**0.9799**	**0.1591**

Back Evaluation of Budget Allocation

Models	CPA $\frac{1}{2}$	CPA $\frac{1}{4}$	CPA $\frac{1}{8}$	CPA $\frac{1}{16}$	CPA $\frac{1}{32}$	Profit $\frac{1}{2}$	Profit $\frac{1}{4}$	Profit $\frac{1}{8}$	Profit $\frac{1}{16}$	Profit $\frac{1}{32}$
LR	31.79	29.47	29.77	27.83	27.46	8.022	6.938	4.386	3.238	1.954
SP	24.84	22.98	21.29	21.39	20.93	13.07	10.28	7.694	4.648	2.776
AH	24.69	21.84	20.37	18.89	19.32	27.03	22.08	15.38	10.32	5.491
AMTA	24.71	21.91	20.43	18.89	19.41	27.01	21.96	15.29	10.33	5.446
ARNN	26.66	23.98	22.61	19.86	18.96	29.10	**23.32**	15.81	11.68	**7.010**
DARNN	**23.47**	**21.24**	**18.50**	**16.85**	**17.63**	**29.25**	22.56	**17.58**	**12.09**	6.26

Models	Conversion Num. $\frac{1}{2}$	Conversion Num. $\frac{1}{4}$	Conversion Num. $\frac{1}{8}$	Conversion Num. $\frac{1}{16}$	Conversion Num. $\frac{1}{32}$	CVR $\frac{1}{2}$	CVR $\frac{1}{4}$	CVR $\frac{1}{8}$	CVR $\frac{1}{16}$	CVR $\frac{1}{32}$
LR	576	427	275	181	107	0.0928	0.0910	0.0873	0.0827	0.0748
SP	452	315	191	112,	62	0.1205	0.1251	0.1223	0.1122	0.1028
AH	1286	925	607	385	208	0.1120	0.1194	0.1197	0.1183	0.1079
AMTA	1285	922	605	385	207	0.1118	0.1192	0.1195	0.1183	0.1073
ARNN	**1527**	**1073**	**684**	**452**	**262**	0.1073	0.1137	0.1119	0.1206	0.1174
DARNN	1315	922	646	419	223	**0.1226**	**0.1274**	**0.1339**	**0.1321**	**0.1206**

Figure: CPA: the lower, the better.

Budget Settings

We set budget constraint as $1/n$ of the total costs in the training dataset.

Visualization & Insights

Visualization of the Attribution

Sequence Level

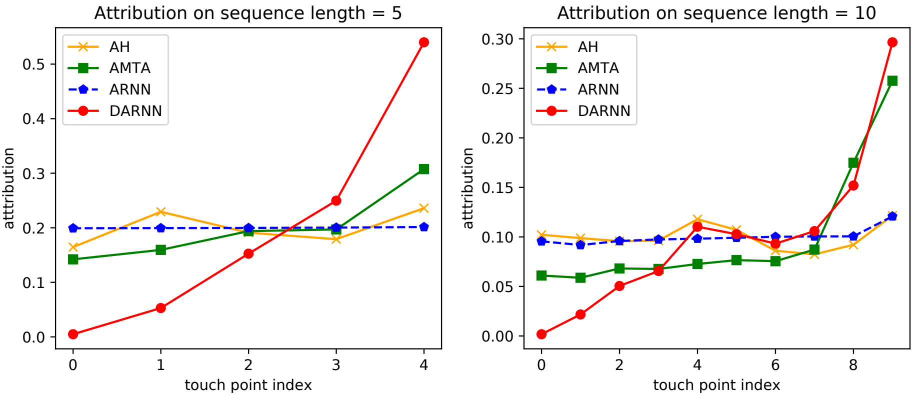


Figure: Touch point level attribution statistics (Miaozhen).

Visualization of the Attribution

Channel Level

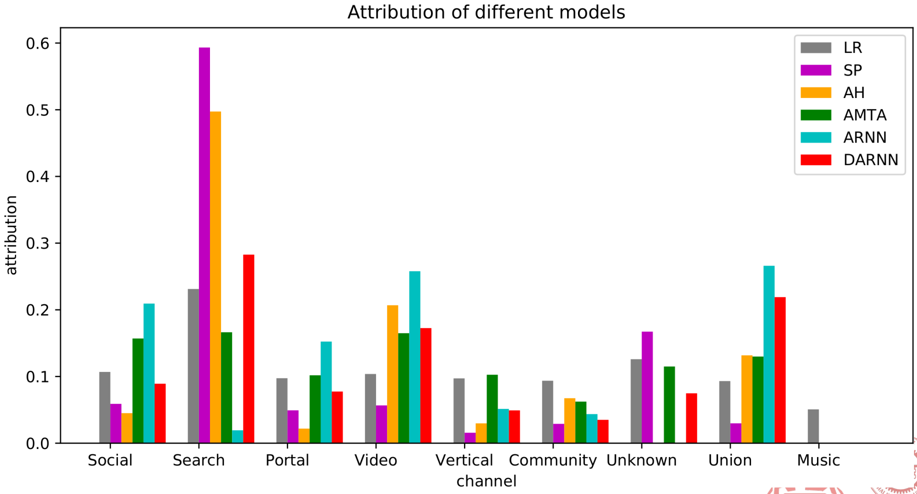


Figure: Attribution of different channels on Miaozhen.

Visualization of the Attribution Preferences

Click-level v.s. Impression-level, $\text{Attr}_j = (1 - \lambda) \cdot a_j^{i2v} + \lambda \cdot a_j^{c2v}$

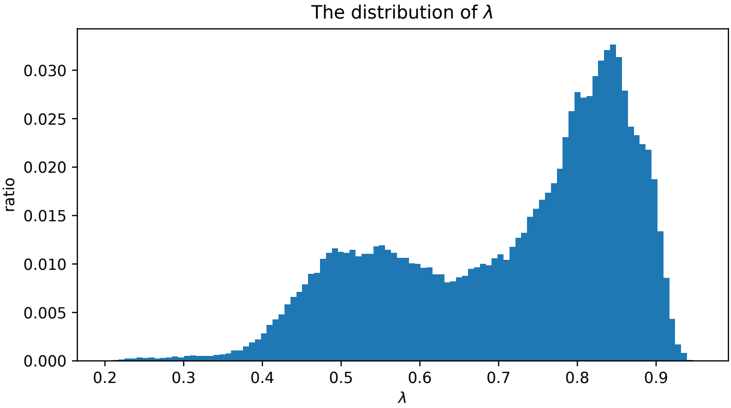


Figure: The distribution of λ over Criteo dataset.

Summary

- First work using *attentional* recurrent model for conversion attribution.

Summary

- First work using *attentional* recurrent model for conversion attribution.
- First work proposing a replay protocol for offline evaluation over the obtained attribution.

Summary

- First work using *attentional* recurrent model for conversion attribution.
- First work proposing a replay protocol for offline evaluation over the obtained attribution.
- Perhaps the subsequent budget allocation should be guided by the data-driven attributions.

Summary

- First work using *attentional* recurrent model for conversion attribution.
- First work proposing a replay protocol for offline evaluation over the obtained attribution.
- Perhaps the subsequent budget allocation should be guided by the data-driven attributions.
- Reproductive code: https://github.com/rk2900/deep-conv-attr.

Summary

- First work using *attentional* recurrent model for conversion attribution.
- First work proposing a replay protocol for offline evaluation over the obtained attribution.
- Perhaps the subsequent budget allocation should be guided by the data-driven attributions.
- Reproductive code: https://github.com/rk2900/deep-conv-attr.
- Outlook
    - To attribute with the consideration of *cost*.