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

1. Problem Background
2. Our Solution
3. Experiments
4. Visualization & Insights
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

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>Growth strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last-click (DEFAULT)</td>
<td>Gives all credit for the conversion to the last-clicked keyword</td>
<td>Most conservative</td>
</tr>
<tr>
<td>First click</td>
<td>Gives all credit for the conversion to the first-clicked keyword</td>
<td>Most growth-oriented</td>
</tr>
<tr>
<td>Linear</td>
<td>Distributes the credit for the conversion equally across all clicks on the path</td>
<td>Moderate</td>
</tr>
<tr>
<td>Time decay</td>
<td>Gives more credit to clicks that happened closer in time to the conversion</td>
<td>Conservative</td>
</tr>
<tr>
<td>Position-based</td>
<td>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</td>
<td>Growth-oriented</td>
</tr>
<tr>
<td>Data-driven</td>
<td>Gives credit to clicked keywords based on how imperative they were in the conversion process</td>
<td>Based on account’s performance</td>
</tr>
</tbody>
</table>

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 $\Rightarrow$ 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
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:
  \[
  \hat{y}_i = p(y = 1 \mid x, z) = r(x_{m_i}, c^{i2v}, c^{c2v})
  \]

- Balance of Impression and Click:
  \[
  r(x_{m_i}, c^{i2v}, c^{c2v}) = r_{conv} \left( (1 - \lambda) \cdot c^{i2v} + \lambda \cdot c^{c2v} \right)
  \]

- \[
  \lambda = \frac{\exp [f_{\lambda}(x_{m_i}, c^{c2v})]}{\exp [f_{\lambda}(x_{m_i}, c^{i2v})] + \exp [f_{\lambda}(x_{m_i}, c^{c2v})]}
  \]

- Attentional Combination:
  \[
  c^{i2v} = \sum_{j=1}^{m_i} a_{ij}^{i2v} h_j
  \]
  \[
  c^{c2v} = \sum_{j=1}^{m_i} a_{ij}^{c2v} s_j
  \]

- Attribution Calculation:
  \[
  Attr_j = (1 - \lambda) \cdot a_{ij}^{i2v} + \lambda \cdot a_{ij}^{c2v}
  \]
Attention Implementation

\{ (Query, Key, Value) \}
Our Solution

Dual-attention Mechanism for Conversion Attribution

Attention as Attribution Credits (cont.)

Conversion Rate Estimation

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

Balance of Impression and Click

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

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

Attentional Combination

$$c^{i2v} = \sum_{j=1}^{m_i} a_{ij}^{i2v} h_j$$

$$c^{c2v} = \sum_{j=1}^{m_i} a_{ij}^{c2v} s_j$$

Attribution Calculation

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

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

∗(Shanghai Jiao Tong University)
The Usage of Attribution

Attribution of the $j$-th Touch Point

\[
\text{Attr}_j = (1 - \lambda) \cdot a_j^{i2v} + \lambda \cdot a_j^{c2v}.
\] (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.
Our Solution

The Usage of Attribution

Attribution of the $j$-th Touch Point

$$\text{Attr}_j = (1 - \lambda) \cdot a^{j2v}_j + \lambda \cdot a^{c2v}_j. \quad (1)$$

- 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 Attribution Guided Budget Allocation

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

\[
\text{Attr}(c_k | y_i) = \sum_{j=1}^{m_i} \text{Attr}_j \cdot 1(c_j = c_k)
\]  

(2)
Back Evaluation for Attribution Guided Budget Allocation

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

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

(2)

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

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

(3)
Our Solution

Back Evaluation for Attribution Guided Budget Allocation

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

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

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

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

Attribution guided Budget Allocation

For channel $c_k$, the allocated budget

$$b_k = \frac{\text{ROI}_{c_k}}{\sum_{v=1}^{K} \text{ROI}_{c_v}} \times B .$$  \hfill (4)
Our Solution

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 = $V_0 \cdot CN - \sum \text{cost}$
- 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**
Experiments

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 | Miaozen &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; | Criteo &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; |
|--------|---------------------------------------------------------------|
|        | AUC &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; | AUC &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; | Log-loss &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; | Log-loss &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; |
| LR     | 0.8418 &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; | 0.9286 &emsp; | 0.3496 &emsp; | 0.3981 |
| SP     | 0.7739 &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; | 0.6718 &emsp; | 0.5617 &emsp; | 0.5535 |
| AH     | 0.8693 &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; | 0.6791 &emsp; | 0.6791 &emsp; | 0.5067 |
| AMTA   | 0.8357 &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; | 0.8465 &emsp; | 0.1636 &emsp; | 0.3897 |
| ARNN   | 0.8914 &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; | 0.9793 &emsp; | 0.1610 &emsp; | 0.1850 |
| DARNN  | **0.9123** &emsp; &emsp; &emsp; &emsp; &emsp; &emsp; | **0.9799** &emsp; | **0.1095** &emsp; | **0.1591** |
Back Evaluation of Budget Allocation

<table>
<thead>
<tr>
<th>Models</th>
<th>$\frac{1}{2}$</th>
<th>$\frac{1}{4}$</th>
<th>$\frac{1}{8}$</th>
<th>$\frac{1}{16}$</th>
<th>$\frac{1}{32}$</th>
<th>$\frac{1}{2}$</th>
<th>$\frac{1}{4}$</th>
<th>$\frac{1}{8}$</th>
<th>$\frac{1}{16}$</th>
<th>$\frac{1}{32}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>31.79</td>
<td>29.47</td>
<td>29.77</td>
<td>27.83</td>
<td>27.46</td>
<td>8.022</td>
<td>6.938</td>
<td>4.386</td>
<td>3.238</td>
<td>1.954</td>
</tr>
<tr>
<td>AH</td>
<td>24.69</td>
<td>21.84</td>
<td>20.37</td>
<td>18.89</td>
<td>19.32</td>
<td>27.03</td>
<td>22.08</td>
<td>15.38</td>
<td>10.32</td>
<td>5.491</td>
</tr>
<tr>
<td>AMTA</td>
<td>24.71</td>
<td>21.91</td>
<td>20.43</td>
<td>18.89</td>
<td>19.41</td>
<td>27.01</td>
<td>21.96</td>
<td>15.29</td>
<td>10.33</td>
<td>5.446</td>
</tr>
<tr>
<td>ARNN</td>
<td>26.66</td>
<td>23.98</td>
<td>22.61</td>
<td>19.86</td>
<td>18.96</td>
<td>29.10</td>
<td>23.32</td>
<td>15.81</td>
<td>11.68</td>
<td><strong>7.010</strong></td>
</tr>
<tr>
<td>DARNN</td>
<td><strong>23.47</strong></td>
<td><strong>21.24</strong></td>
<td><strong>18.50</strong></td>
<td><strong>16.85</strong></td>
<td><strong>17.63</strong></td>
<td><strong>29.25</strong></td>
<td><strong>22.56</strong></td>
<td><strong>17.58</strong></td>
<td><strong>12.09</strong></td>
<td>6.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Models</th>
<th>Conversion Num.</th>
<th>CVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>576</td>
<td>0.0928</td>
</tr>
<tr>
<td>SP</td>
<td>452</td>
<td>0.1205</td>
</tr>
<tr>
<td>AH</td>
<td>1286</td>
<td>0.1120</td>
</tr>
<tr>
<td>AMTA</td>
<td>1285</td>
<td>0.1118</td>
</tr>
<tr>
<td>ARNN</td>
<td><strong>1527</strong></td>
<td><strong>0.1226</strong></td>
</tr>
<tr>
<td>DARNN</td>
<td>1315</td>
<td>0.1206</td>
</tr>
</tbody>
</table>

**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
Figure: Touch point level attribution statistics (Miaozhen).
Visualization of the Attribution

Channel Level

![Attribution of different models](image)

**Figure:** Attribution of different channels on Miaozhen.

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

(C) (Shanghai Jiao Tong University)
Visualization of the Attribution Preferences

Click-level v.s. Impression-level, \( \text{Attr}_j = (1 - \lambda) \cdot a_{j2v} + \lambda \cdot a_{c2v} \)

Figure: The distribution of \( \lambda \) over Criteo dataset.
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.
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.