

MTA算法

参考自：[千川归因one Page](#)

MTA 的全称是 Mutil-Touch Attribution（多触点归因），是指一次转化的功劳不再完全归因至用户转化前的某一个广告触点上，而是按照一定的权重分配至发生转化前被触达的多个触点上。

DDA 的全称是 Data Driven Attribution（数据驱动的归因），是一种基于算法的多触点归因模型，主要通过算法或者模型预估来分析用户转化路径中，每个触点的存在或者顺序对转化的贡献。

MTA/DDA 为什么好？

目前普遍使用的 last touch 归因，简单直接的将价值归给最后一次触点，这种并不精准的归因方法极有可能高估最后一次触点的价值。这会带来：

- 客户预算向最后一次触点倾斜，但最终收益却不符合预期
- 模型将其他触点价值低估，减少其他触点的分发，从而导致客户转化变少

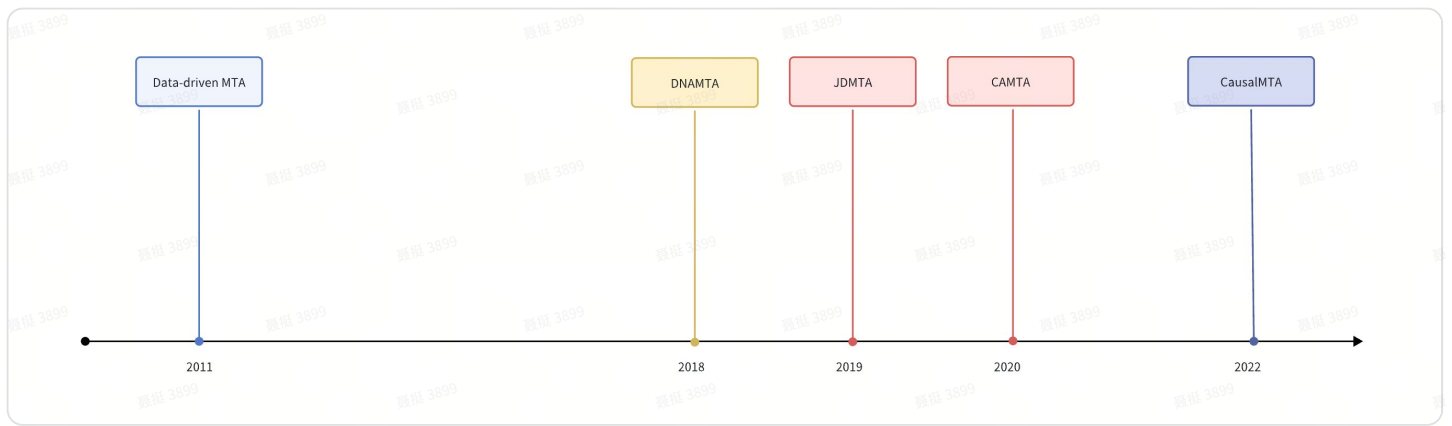
科学的 MTA/DDA 模型能够改变这种情况，公平合理的分配触点贡献，从而提升模型预估效果，为客户带来收益。

其他广告商的 MTA/DDA 服务 是什么样的？



有哪些 MTA/DDA 算法？它们是怎么做的？

迄今为止，有五篇代表性的工作。具体如下：



- **Data-driven MTA**: Data-driven Multi-touch Attribution Models. 2011
- **DNAMTA**: Deep Neural Net with Attention for Multi-channel Multi-touch Attribution. 2018
- **JDMTA**: Causally Driven Incremental Multi Touch Attribution Using a Recurrent Neural Network. 2019
- **CAMTA**: Causal Attention Model for Multi-touch Attribution. 2020
- **CausalMTA**: Eliminating the User Confounding Bias for Causal Multi-touch Attribution. 2022

一句话介绍

- **Data-driven MTA**
 - 是什么？
 - 第一次提出数据驱动的多触点归因模型。
 - 为什么？
 - 通过多触点归因，广告主可以量化评估各个渠道对转化的重要性，从而调整在各个渠道上的预算分配，实现全局角度的投放策略优化。
 - 怎么做？
 - 使用逻辑回归 + bagged 的组合方法。其中逻辑回归用于求每个触点的参数，bagged 方法能减少因高度相关的协变量而导致的估计变异性。
- **DNAMTA**
 - 是什么？
 - 利用 **触点序列信息** 的多触点归因模型。
 - 为什么？
 - 触点是一个序列，序列中触点的顺序、频率、发生时间以及触点间依赖性关系都会影响最终的结果。
 - 对触点序列的建模，可以捕捉到过去的信息是如何积累并影响最终结果的。
 - 怎么做？

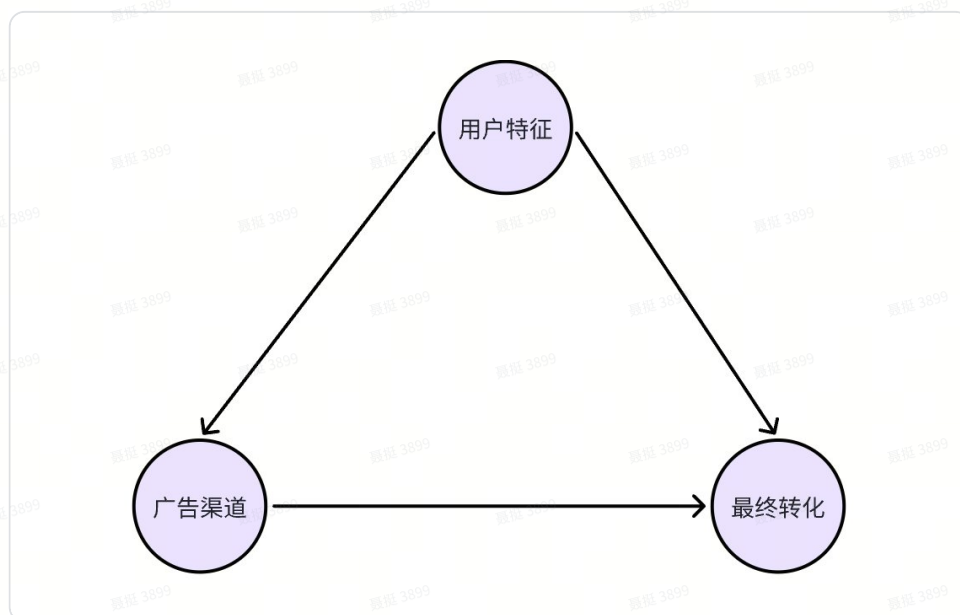
- LSTM 对触点序列进行学习，MLP 对用户静态信息进行学习，最终通过注意力机制获取触点权重。

• JDMTA

- 是什么？
 - 除了构建多触点模型，JDMTA 将购买时 **由广告产生的增量** 分配给每个广告触点。
- 为什么？
 - 由于用户自身的行为、偏好等等，本来就会有一定的购买倾向。
 - 极端来说，无论如何都会发生的购买不应该将其结果分配给广告触点。
 - eg：购买苹果手机后，需要购买充电器。
- 怎么做？
 - JDMTA 使用 **夏普利值** 在考虑触点顺序的情况下 将转化的整体增量 改进后 分配给触点序列。

• CAMTA

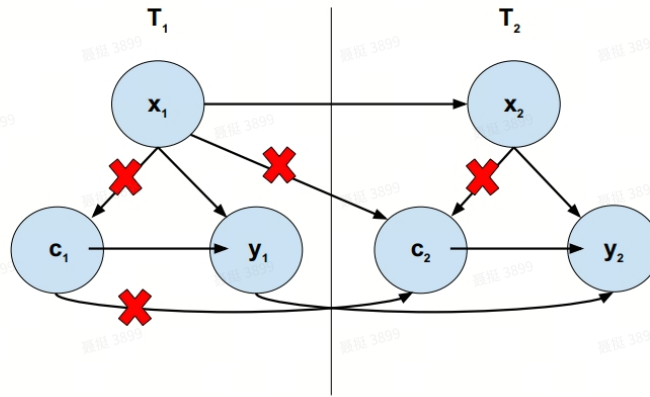
- 是什么？
 - 混杂问题：因果学习的概念。因为某种原因同时影响两个结果，并进而错误推断两个结果存在因果关系的问题被称为混杂。
 - 广告混杂问题： **广告场景下，用户的静态特征和历史行为也都会影响广告渠道和最终转化。**



- 举个例子：某个用户，对苹果手机非常喜欢，打算购置最新的一款。他搜索该款手机，点击搜索结果中该款手机的广告并最终下单。并不能得到因为该款手机在搜索渠道向用户展示，从而导致用户下单购买的结论。因为，真实的情况是，搜索渠道的广告曝光和用户下单购买都是因为用户对该款手机感兴趣，即使没有广告曝光，用户仍会购买该款手机。

- 为什么？

- 用户看到的广告不是随机的。在这种有偏的情况下，无法准确的衡量此广告带来的收益。也无法准确得到用户最终是否会转化。
- 理想无偏情况下，每个触点、每个广告位的用户分布应该是随机的。这样得到的转化能够更准确的衡量。
- 例子：如下图所示，T1 和 T2 时刻，用户上下文 x_1, x_2 作为混杂因子，既影响这两个时刻用户是否发生转化 (y)，也影响了用户看到的广告 (c)



- 下文会举例子详细解释这些概念。

怎么做？

- 在对触点做建模的时候，采用两个 loss function
- 渠道预估的分类器经过 softmax 函数得到各渠道概率的预估值。
- 点击预估的分类器 Sigmoid 函数得到点击概率的预估值。
- 为了学到无偏估计，需要最大化渠道预估分类器的损失函数。
- 为了准确预估广告是否被点击，需要最小化点击预估分类器的损失函数。

CausalMTA

是什么？

- 和 CAMTA 一样，CausalMTA 认为渠道序列中的混杂问题需要考虑。
- CausalMTA 从静态属性和动态特征两个角度，系统性的消除了用户偏好这一混杂偏差。

为什么？

- CausalMTA 认为广告中静态属性和动态特征的影响是相互独立的。其中，静态属性决定用户对于物品的兴趣，动态特征决定用户购买的可能性。
- eg:
 - 大学生喜欢买时尚的手机，而中年人更喜欢高性价比的手机。这就是静态属性（年龄，性别，职业等等）。
 - 看到广告的停留时间，反映了用户的购买意图。这就是动态特征。

怎么做？

- 对于静态属性，论文使用 变分循环自编码器 作为渠道序列生成模型 获取它的无偏分布。之后，基于无偏分布和逆概率加权的方法对每个转化路径重加权，从而消除静态属性引起的选择偏差。
- 对于动态特征，其和 CAMTA 类似，也是借鉴 CRN，通过循环神经网络和域对抗训练，生成用户历史的无偏表征，从而消除动态特征引起的选择偏差，得到无偏的转化率预估模型。
- 最后，基于转化率预估模型，采用反事实分析计算各渠道的夏普利值作为归因权重

具体是怎么做的？

Data-driven MTA

方法一

bagging 方法训练多个逻辑回归模型，每个模型训练对应的自变量和数据都是随机抽取的。最终，每个自变量的参数是多个模型参数的平均值。这里的参数代表每个渠道的权重。

方法二

衡量渠道贡献的方法：

- 对于数据集，论文计算概率如下：

$$P(y|x_i) = \frac{N_{positive}(x_i)}{N_{positive}(x_i) + N_{negative}(x_i)}$$

$$P(y|x_i, x_j) = \frac{N_{positive}(x_i, x_j)}{N_{positive}(x_i, x_j) + N_{negative}(x_i, x_j)}$$

其中： $i \neq j$ ， y 是一个转化的结果变量，取值为 0 或 1，表示有无转化。 x_i 表示 n 个不同的广告路径。 $N_{positive}$ 表示经过了渠道 x_i 且转化了的用户个数， $N_{negative}$ 表示没有转化。

- 计算渠道 i 贡献如下：

$$C(x_i) = p(y|x_i) + \frac{1}{2N_{j \neq i}} \sum_{j \neq i} p(y|x_i, x_j) - p(y|x_i) - p(y|x_j)$$

公式很好理解，计算一个渠道的总贡献。

DNAMTA

算法

- 是什么？（问题建模）
 - MTA 中有两种事件，分别为转化事件和触点事件
 - 客户的访问路径为 $P_i = \{x_1, x_2, \dots, x_T\}$ ，其中用 x_t 表示路径中用户接触到的第 t 个事件。
 - 客户的静态信息如性别、年龄和注册日期等控制变量用 C_i 表示
 - 若客户路径以转化结束，则视为正路径 ($Y_i = 1$)，否则，为负路径 ($Y_i = 0$)。

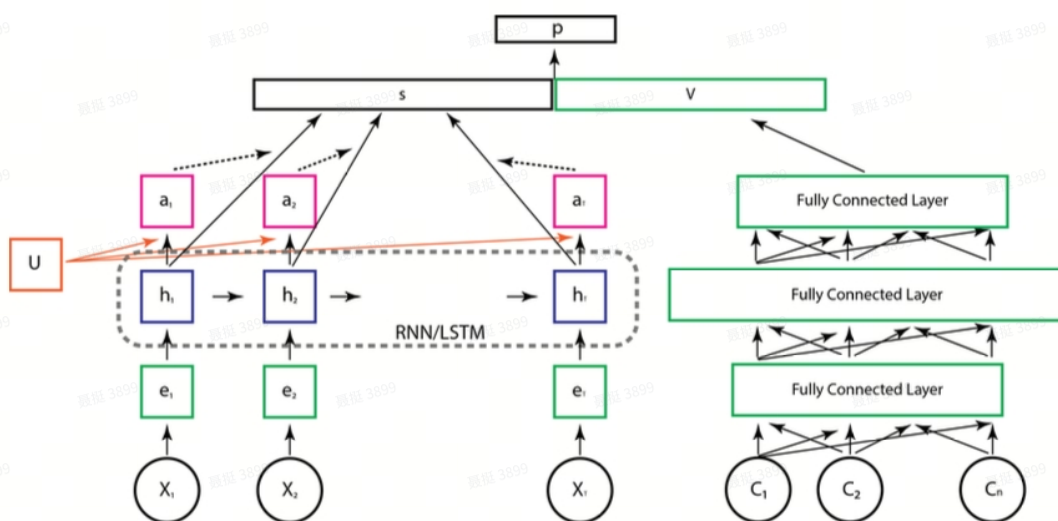
- 而多触点归因的目标就是：

- 预估每个触点 x_t 的贡献值 a_t ，且满足 $\sum_{i=1}^n a_i = 1$

- 怎么样？

- 论文利用概率推理来解释客户的转化决策。
 - 即找到一条路径，其中有一系列触点 P_i 及其相应的控制变量 C_i ，转化的概率为 $P(Y_i|P_i, C_i)$
 - DNAMTA 将目标变为求 $P(Y_i|P_i, C_i)$ 的值，而每个触点的归因权重通过模型的中间层输出得出。

- 最终建模：



- 模型结构

- 触点输入层：输入一个触点序列，每个触点用 one-hot 编码。
 - 触点嵌入层：将 one-hot 编码转变为稠密向量。操作相当于求 embedding，其中蕴含了触点之间隐藏的上下文。
 - LSTM 层：使用 LSTM 获得触点的另一层表示，该表示中蕴含了上下文信息。上下文信息对客户的转化比较重要，因为触点顺序、频率以及触点之间的依赖关系都会影响最终的结果。
 - 时间衰减下的注意力层：触点序列中每个触点产生的时间都是不同的。该层结合时间衰减信息，使用注意力机制。其中，上下文向量 u 作为触点归因权重的领域知识，论文将其作为模型参数的一部分，通过训练来学习。发生时间与结束时间之间的时间间隔差定义为 T_t 。最终得到归因权重：

$$v_t = \tanh(W_v h_t + b_v)$$

$$a_t = \frac{\exp(v_t^T u - \lambda T_t)}{\sum_t \exp(v_t^T u - \lambda T_t)}$$

$$s = \sum_t a_t h_t$$

- 融合层：以上结构建立了触点序列和转化结果的联系，但忽略了客户自身的静态信息。比如：性别，年龄等可能会影响最终的转化率。融合部分使用多个密集的全连接层来学习这些控制变量。
- 之后，我们将两个部分联合起来，应用于之后的操作。
- 最终分类层，触点序列最终转化为正例，没有转化为负例。作为学习结果的依据。

JDMTA

写在前面

- 夏普利值 (Shapley Value)

- 合作博弈理论中的一个概念，用于公平的分配合作收益。
- 它按照成员对联盟的边际贡献率将利益进行分配，即成员 i 所分得的利益等于该成员为它所参与联盟创造的边际利益的平均值。
- 成员 i 的夏普利值计算公式如下：

$$x_i = \sum_{s \in N, i \in s} w(|s|) [v(s) - v(s \setminus i)], i = 1, 2, \dots, n$$

$$w(|s|) = \frac{(n - |s|)! (|s| - 1)!}{n!}$$

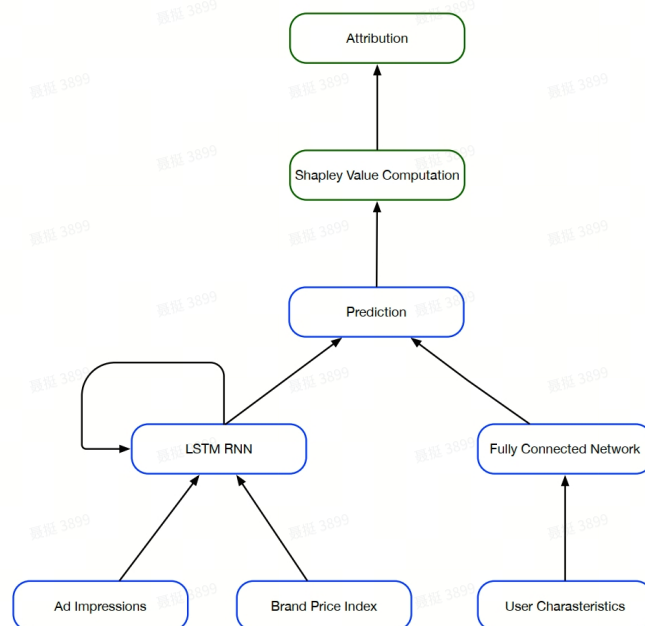
- 其中 s 是 N 中包含 i 的所有子集， $|s|$ 是子集 s 中的元素数目 (人数)， $w(|s|)$ 是加权因子， $v(s)$ 是 s 的贡献值。

- 夏普利值加权因子推导

- 联盟规模为 s ，取值范围为 $1 \sim n$ 。所有规模可能性相同，则选择一个特定规模联盟的概率为 $1/n$
- 为了构成规模为 s 的联盟，从剩余 $n-1$ 个参与人中选择参与人 i 的 $s-1$ 个同伴，是一个排列组合 C_{n-1}^{s-1}
- $1/C_{n-1}^{s-1}$ 为任意一种选择的概率，该数字乘以 $1/n$ 则得到一个规模为 s 的特定联盟 S 包含成员 i 的概率。

算法

- JDMTA 算法



- JDMTA 整体的算法结构分为两个部分：

- 第一部分（图中蓝色框）和 **DNAMTA 算法类似**，JDMTA 算法也将目标转化为下单率预测，使用 LSTM 挖掘广告曝光序列的信息，并另外通过一个全连接网络对用户特征进行建模，最后结合广告曝光序列隐状态和用户特征隐向量预测下单率。
- 第二部分（图中绿色框）量化分析归因权重，和上述算法通过注意力机制、使用模型中间层输出作为归因权重不同，JDMTA 算法引入合作博弈论。

- 第一部分在 DNAMTA 中已经说过，接下来重点说说第二部分。第二部分通过量化分析某品牌某天在某广告位上对用户的广告曝光，得到最终下单的归因权重。

- 在论文中首先定义广告位集合对下单的期望增益 $v(A_{ibT})$ 。令 $Y_{ibT} \in (0, 1)$ 表示用户 i 在第 T 天是否对品牌 b 下订单，1 为下单，0 为未下单。则：

$$v(A_{ibT}) = E[Y_{ibT}|A_{ibT}] - E[Y_{ibT}|\emptyset_b]$$

- 其中， $E[Y_{ibT}|A_{ibT}]$ 表示有广告曝光下的用户 i 对品牌 b 的下单概率， $E[Y_{ibT}|\emptyset_b]$ 表示无广告曝光下的用户 i 对品牌 b 的下单概率，即期望增益是因广告曝光吸引用户 i 兴趣后对最终下单品牌 b 的概率的提升。

- 而所有广告位的归因权重之和，等于期望增益，也就等于最终下单概率的提升。

- 论文然后定义如何分配 $v(A_{ibT})$ ，将其划分到某个粒度的广告曝光上，作为该粒度每组广告曝光的归因权重。

- 分配的方式即是夏普利值，接下来我们看下具体的例子。

- 举例：

- 假设品牌数为 2 ($B=2$)，广告位数为 3 ($K=3$)，归因窗口为 3 天 ($T=3$)
- 用户 i 第一天 没有浏览广告

$$x_{i11} = (0, 0, 0)$$

$$x_{i21} = (0, 0, 0)$$

- 第二天 在广告位 1 浏览了 4 次品牌 1 的广告，在广告位 2 浏览了 7 次品牌 1 的广告，在广告位 3 浏览了 10 次品牌 2 的广告

$$\begin{aligned}x_{i12} &= (4, 7, 0) \\x_{i22} &= (0, 0, 10)\end{aligned}$$

- 第三天 在广告位 1 上浏览了 4 次品牌 1 的广告，并且第 3 天对品牌 1 下单

$$\begin{aligned}x_{i13} &= (4, 0, 0) \\x_{i23} &= (0, 0, 0)\end{aligned}$$

- 则 我们可以得到 访问序列如下

$$\begin{aligned}x_{i1} &= (0, 0, 0, 0, 0, 0) \\x_{i2} &= (4, 7, 0, 0, 0, 10) \\x_{i3} &= (4, 0, 0, 0, 0, 0) \\x_{i,1:T} &= (0, 0, 0, 0, 0, 0, 4, 7, 0, 0, 0, 10, 4, 0, 0, 0, 0, 0)\end{aligned}$$

- 用户 i 浏览品牌 1 广告的广告位、日期二元组为：

$$N_{iT} = \{\{1, 2\}, \{2, 2\}, \{1, 3\}\}$$

- 假设计算用户 i 第 3 天在广告位 1 上浏览品牌 1 广告的归因权重，即 $\{k, t\} = \{1, 3\}$ ，则 $N_{iT}/\{k, t\} = \{\{1, 2\}, \{2, 2\}\}$ ，其相应的幂集为 $\{\emptyset, \{1, 2\}, \{2, 2\}, \{\{1, 2\}, \{2, 2\}\}\}$
- 对于幂集中的每个子集 S ，需要计算 $w(S \cup \{k, t\})$ 和 $w(S)$
- 以 $S = \{\{1, 2\}\}$ 举例，需要将品牌 1 在二元组 $\{2, 2\}$ 和 $\{1, 3\}$ 上的广告曝光次数设置为 0，则：

$$x_{i,1:T}^{(1,S)} = (0, 0, 0, 0, 0, 0, 4, 0, 0, 0, 0, 10, 0, 0, 0, 0, 0, 0)$$

- 对于 $S \cup \{1, 3\}$ ，即假设用户 i 第 2 天在广告位 1 上浏览了品牌 1 的广告的基础上，第 3 天又在广告位 1 上浏览了品牌 1 的广告。相应的访问序列变化为：

$$x_{i,1:T}^{(1,S \cup \{1,3\})} = (0, 0, 0, 0, 0, 0, 4, 0, 0, 0, 0, 10, 4, 0, 0, 0, 0, 0)$$

- 分别将上述两个访问序列输入模型得到用户 i 下单品牌 1 的概率预测值。
- 如此对每个子集进行处理，最终即可根据夏普利值计算公式得到 $\{1, 3\}$ 对于用户 i 下单品牌 1 的归因权重。
- 归因权重公式为：

$$\varrho_{k,t}(\mathcal{N}_{iT}) = \sum_{S \in \mathcal{N}_{iT} \setminus \{k,t\}} \frac{|S|!(|\mathcal{N}_{iT}| - |S| - 1)!}{|\mathcal{N}_{iT}|!} \left[\sigma \left(\mathbf{x}_{i,1:T}^{(b, S \cup \{k,t\})}, \mathbf{p}_{1:T}, \mathbf{d}_i; \hat{\Omega} \right) - \sigma \left(\mathbf{x}_{i,1:T}^{(b,S)}, \mathbf{p}_{1:T}, \mathbf{d}_i; \right) \right]$$

CAMTA

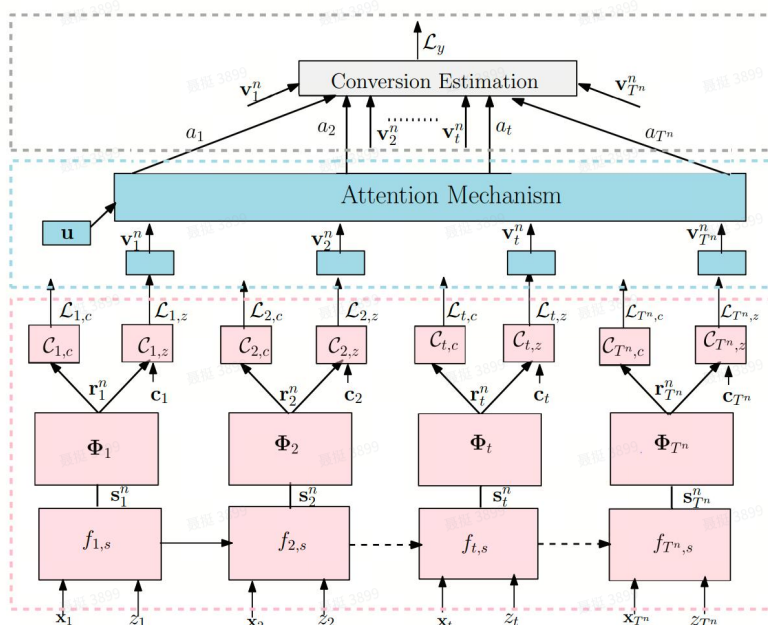
算法

- 问题定义

- 有数据集 $D = \{u^n, \{x_t^n, c_t, z_{t+1}^n\}, y^n\}_{n=1}^N$ ，包含了 N 个用户的转化路径的数据。
- 对于用户 u^n 的转化路径数据，其中有 T^n 个触点。
- 对于第 t 个触点，其包含当时的渠道 c_t 、上下文特征向量 x_t^n （其中包含用户信息、广告信息和上下文信息等），以及用户最终是否点击的二值结果 z_{t+1}^n 。
 - 注：渠道在论文中指邮件、app、网页等等
- 渠道 c_t 是 one-hot，即 $c_t = [c_t(1), \dots, c_t(k), \dots, c_t(K)]$ 。其中，若第 t 个触点在渠道 k ，则 $c_t(k) = 1$ ，其他位为 0。
- y^n 代表是否发生转化的二值结果。

模型结构

- 粉色虚线框代表因果循环网络
- 蓝色虚线框代表注意力机制
- 灰色框代表转化估计组件



因果循环网络

- 使用循环神经网络对用户历史触点进行建模，得到用户历史隐状态表征。

$$s_t^n = f_{t,s}(x_t^n, x_{t-1}^n, \dots, x_1^n, c_{t-1}, \dots, c_1, z_t^n, z_{t-1}^n, \dots, z_2^n)$$
- s_t^n 作为历史隐状态表征，会被用在第 t 个触点时的广告投放决策中，之后影响触点对应的渠道选择。如果它是有偏的，则结果也会有偏。
 - 注：上文说的混杂因子概念
- 因此，在 s_t^n 基础上，会学习一个历史状态的无偏表征 r_t^n 。具体操作是通过一个线性映射矩阵 Φ_t ，将 s_t^n 转化为 r_t^n 。

$$\Phi_t : s_t^n \rightarrow r_t^n$$

- 之后，将 r_t^n 作为后续的两个 loss function 输入进行域对抗训练。
- loss function（最大最小损失）
 - 渠道预估的分类器 $C_{t,c}$ 为两层 MLP 结构，最后一层的输出为 K 维向量，该向量再经过 softmax 函数得到各渠道概率的预估值。
 - 点击预估的分类器 $C_{t,z}$ 也为两层 MLP 结构，最后一层的输出为 1 维标量，该标量再经过 Sigmoid 函数得到点击概率的预估值。
 - 为了学到无偏估计，需要最大化渠道预估分类器的损失函数。
 - 注：说明和原始不相符，隐态表征无偏
 - 为了准确预估广告是否被点击，需要最小化点击预估分类器的损失函数。
 - 注：说明预测足够精准
 - 综上，最终需要最小化如下损失函数：

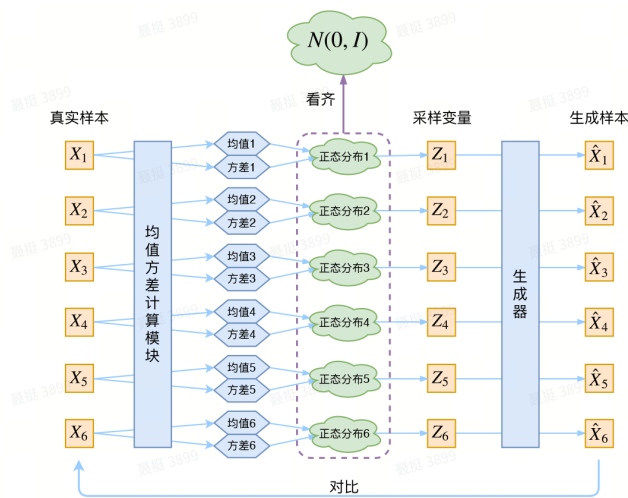
$$L_r(\Phi, \lambda) = \sum_t L_{t,z}(r_t^n) - \lambda L_{t,c}(r_t^n)$$

- 注意力机制
 - 在因果循环网络获取用户隐式表征的基础上，进一步通过注意力机制计算每个触点对最终转化的归因权重。
- 转化预估
 - 对各触点点击的隐状态表征基于归因权重进行加权求和，得到用户历史的隐状态表征
 - 之后，通过一层 MLP 和 sigmoid 函数得到最终的转化率预估值

CausalMTA

写在前面

- 变分自编码器
 - 参考自：[苏神博客](#)
 - 变分自编码器（Variational Autoencoder, VAE）是一种生成式模型。
 - 它由编码器和解码器两部分组成。编码器将输入数据映射到一个潜在空间，这个潜在空间通常具有较低的维度，并遵循特定的概率分布，一般假设为正态分布。**编码器的输出是该分布的参数**，例如均值和方差。通过对这些参数进行采样，可以得到潜在变量。解码器则将潜在变量映射回原始数据空间，试图重构输入数据。
 - 对比只能生成和原始输入相似结果的自编码器，这样做的好处在于可以从原始数据集产生新的数据。



• 循环变分自编码器：

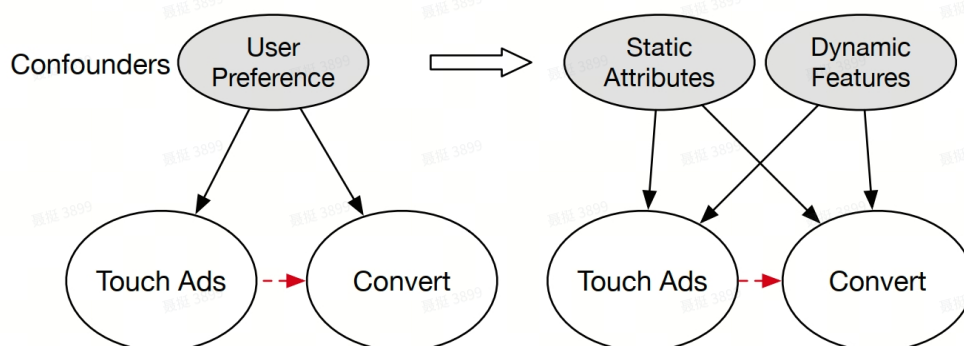
- 参考自：[Variational Recurrent Auto-Encoders](#)
- 循环变分自编码器在 2014 年提出，它利用 RNN 对时间序列的循环记忆能力，将变分自编码器和 RNN 融合，从而将时间序列映射到潜在表示。

• IPW (Inverse Probability Weighting) ，逆概率加权

- 逆概率加权是因果推断中一种常见的技术，主要目的是处理样本中的选择偏倚或实验的不随机分配。
- 算法通过赋予每个样本一个权重来调整样本的代表性，这些权重根据每个样本被选中的概率的倒数进行计算。

算法

- 问题的背景和 CAMTA 一致，常见的 MTA 归因方法并没有考虑到触点序列的混杂因子。
- CausalMTA 归因方法的优势点在于：
 - 它从静态属性和动态特征两个角度，系统性的消除了用户偏好这一混杂偏差。
 - CausalMTA 认为广告中静态属性和动态特征的影响是相互独立的。其中，静态属性决定用户对于物品的兴趣，动态特征决定用户购买的可能性。



(a) Structure casual model of MTA

- eg:

- 大学生喜欢买时尚的手机，而中年人更喜欢高性价比的手机。这就是静态属性（年龄，性别，职业等等）。
- 先前访问的广告和停留时间，反映了用户的购买意图。这就是动态特征。
- 论文方法概述

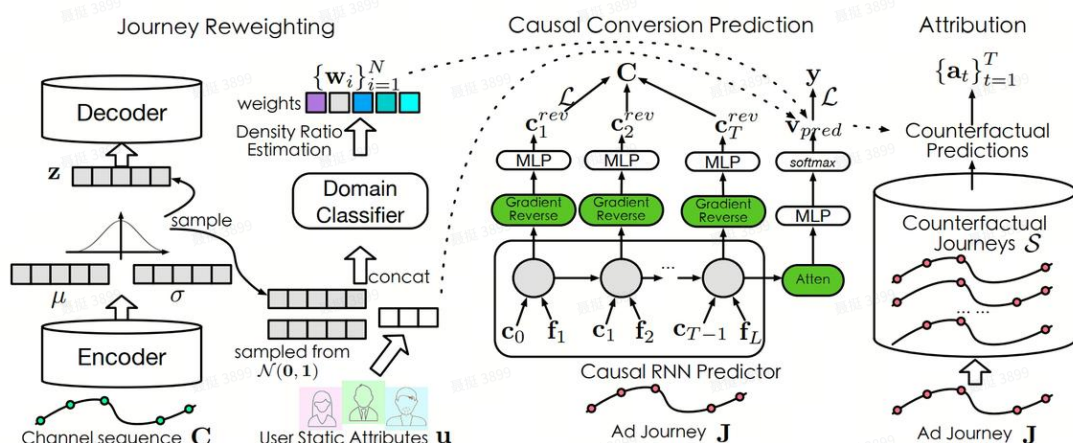


Figure 2: Architecture of CausalMTA.

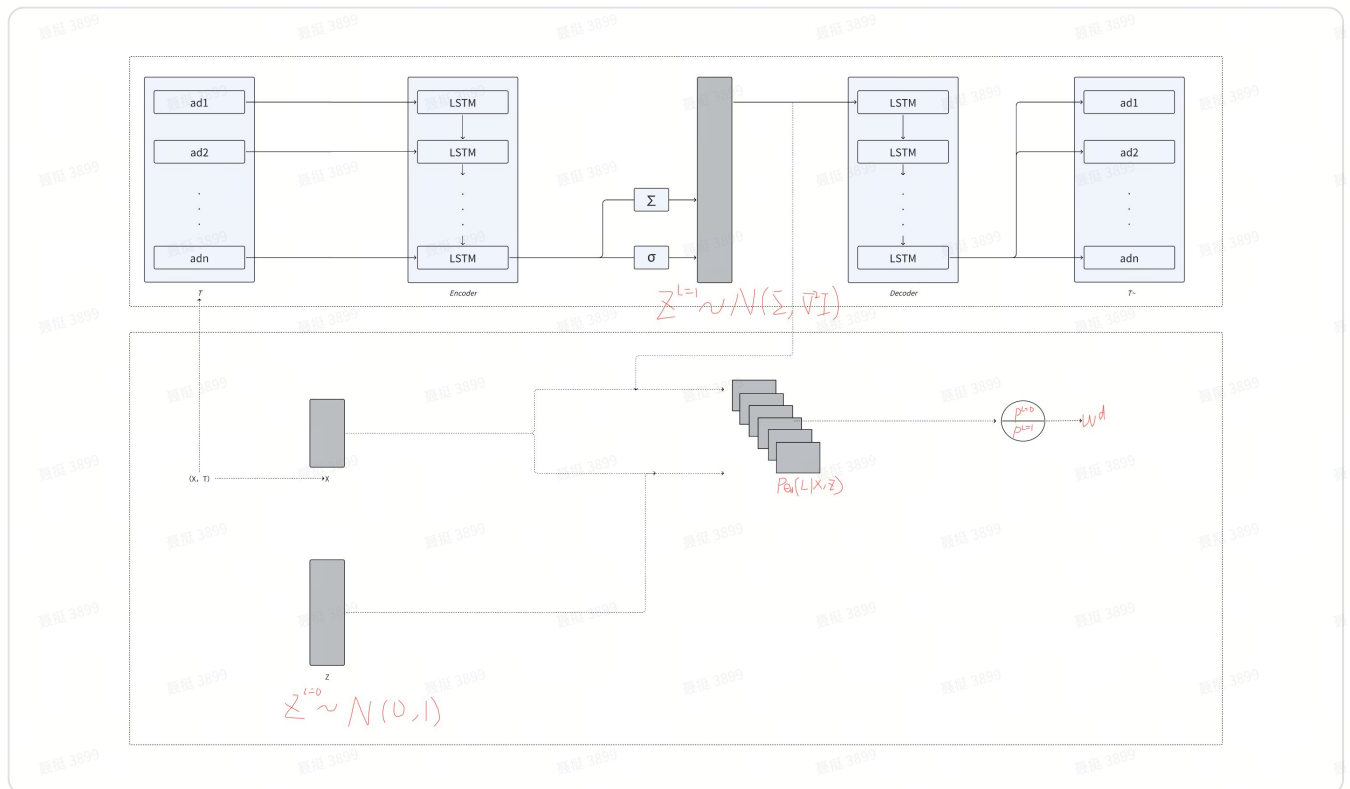
- 方法总共分为三个部分，转化路径加权（Journey Reweighting）、因果转化率预估（Causal Conversion Prediction）、归因权重计算（Attribution）。
- 转化路径加权（Journey Reweighting）。对于静态属性，论文使用 变分循环自编码器 作为渠道序列生成模型 获取它的无偏分布。之后，基于无偏分布和逆概率加权的方法对每个转化路径重加权，从而消除静态属性引起的选择偏差。
- 因果转化率预估（Causal Conversion Prediction）。对于动态特征，其和 CAMTA 类似，也是借鉴 CRN，通过循环神经网络和域对抗训练，生成用户历史的无偏表征，从而消除动态特征引起的选择偏差，得到无偏的转化率预估模型。
- 归因权重计算（Attribution）。基于转化率预估模型，采用反事实分析计算各渠道的夏普利值作为归因权重，即对各渠道，使用转化率预估模型分别预估有无该渠道时的转化率，因引入该渠道带来的转化率提升即该渠道对转化的边际期望增益，也就是该渠道的夏普利值，被作为该渠道的归因权重。
- 论文方法详述
 - 问题定义

- D 表示广告触点数据集，其中包含 U 个用户的 N 个转化路径
- 每条转化路径可用三元组表示，为 (u^i, J^i, y^i)
- 其中， u^i 表示用户静态属性， J^i 表示触点序列， y^i 表示最终是否发生转化(0, 1 标志)
- 每个触点序列包含多个触点，即 $\{p_t^i\}_{t=1}^{T^i}$
- 每个触点进一步由两部分组成，即 $p_t^i = (c_t^i, f_t^i)$ 。其中， c_t^i 表示渠道，用 one-hot 编码，共有 K 个渠道。即 $c_t^i \in \{c_1, \dots, c_k, \dots, c_K\}$ ， f_t^i 表示用户动态特征（即当前触点前用户已浏览的广告）。

- 多触点归因的目标即对于转化路径进行建模，计算每个渠道的归因权重。

● 论文模块详解

○ 转化路径加权 (Journey Reweighting)



- 静态属性代表用户的偏好，用户的偏好会影响转化的类型。
- 而理论上当数据足够多时，循环变分自编码器生成的分布是随机的，不受用户静态属性的影响。但此时结果还是在用户偏好的影响下的。
- 为了解决这种影响，论文采用了重加权的方法。具体的：
 - 重加权由逆概率加权的方式得到，学得的样本权重应当满足公式：
$$W_T(u, C) = p(C)/p(C|u)$$
 - 如上图方框所示，在变分循环自编码器中已经得到的样本渠道序列隐变量的分布。那么对于每个样本的权重，可以通过如下推断得到：

$$\begin{aligned}
w^i &= W_T(\mathbf{u}^i, \mathbf{c}^i) = \frac{p(\mathbf{c}^i)}{p(\mathbf{c}^i | \mathbf{u}^i)} \\
&= \frac{p(\mathbf{c}^i)}{\int_{\mathbf{z}} p(\mathbf{c}^i | \mathbf{z}) p(\mathbf{z} | \mathbf{u}^i) d\mathbf{z}} = \frac{1}{\int_{\mathbf{z}} \frac{p(\mathbf{c}^i | \mathbf{z})}{p(\mathbf{c}^i)} p(\mathbf{z} | \mathbf{u}^i) d\mathbf{z}} \\
&= \frac{1}{\int_{\mathbf{z}} \frac{p(\mathbf{z} | \mathbf{c}^i)}{p(\mathbf{z})} p(\mathbf{z} | \mathbf{u}^i) d\mathbf{z}} = \frac{1}{\int_{\mathbf{z}} \frac{p(\mathbf{z} | \mathbf{u}^i)}{p(\mathbf{z})} p(\mathbf{z} | \mathbf{c}^i) d\mathbf{z}} \\
&= \frac{1}{\int_{\mathbf{z}} \frac{p(\mathbf{z}, \mathbf{u}^i)}{p(\mathbf{z}) p(\mathbf{u}^i)} p(\mathbf{z} | \mathbf{c}^i) d\mathbf{z}} = \frac{1}{\int_{\mathbf{z}} \frac{1}{W_Z(\mathbf{u}^i, \mathbf{z})} p(\mathbf{z} | \mathbf{c}^i) d\mathbf{z}} \\
&= \frac{1}{\mathbb{E}_{\mathbf{z} \sim q_\phi(\mathbf{z} | \mathbf{c}^i)} \left[\frac{1}{W_Z(\mathbf{u}^i, \mathbf{z})} \right]},
\end{aligned}$$

- 也就是说，每个样本的权重可以通过**渠道分布的采样和用户同时出现的概率**获得。
- 渠道分布和用户的正例我们得到了，负例该怎么获得呢？
- 论文将**标准分布的采样和用户**来作为负例。
- 如上图，我扭曲的字体所示。
- 之后，论文搭建了一个领域分类器。分类器为常规的 Embedding+MLP 结构，最后一层的输出经过 Sigmoid 函数得到正样本概率。
- 最终训练好分类器后，作者计算 $W_z(u, z)$ 方式如下。这部分权重计算方式作者并未给出具体推导方法，我猜测是个经验值。在带回上述公式，计算 w^i 。
 - 注：负例 $L=0$ / 正例 $L=1$

$$W_z(u, z) = \frac{p(u, z | L=0)}{p(u, z | L=1)} = \frac{p(L=0 | u, z)}{p(L=1 | u, z)}$$

◦ 因果转化估计

- 和 CAMTA 类似，通过循环神经网络和域对抗训练，生成用户历史的无偏表征。
- 该模块将渠道序列、用户动态特征序列和用户静态属性通过 Embedding 层得到相应的 Embedding 向量，然后逐步将序列中每步的渠道 Embedding 向量和用户动态特征 Embedding 向量拼接在一起，通过 LSTM 输出每步的用户历史的表征 out。
- 该步骤不再详述。
- 之后，该模块的损失函数为下式。转化路径加权计算的 w_i 在该算式中也有体现。

$$\mathcal{L}_p = \gamma \sum_{i=1}^N \sum_{t=1}^{T_i} CE(\mathbf{c}_t^{rev}, \mathbf{c}_t) + \delta \sum_{i=1}^N \mathbf{w}_i \cdot CE(\mathbf{v}_{pred}^i, \mathbf{y}^i), \quad (4)$$

◦ 归因权重计算

- 和 JDMTA 类似，采用反事实分析计算各渠道的夏普利值作为归因权重。

- 即对各渠道，使用转化率预估模型分别预估有无该渠道时的转化率，因引入该渠道带来的转化率提升即该渠道对转化的边际期望增益，也就是该渠道的夏普利值，被作为该渠道的归因权重。