因果推断在哈啰智能营销的应用

董彦燊 哈啰出行 算法工程师

目 录

PART 01

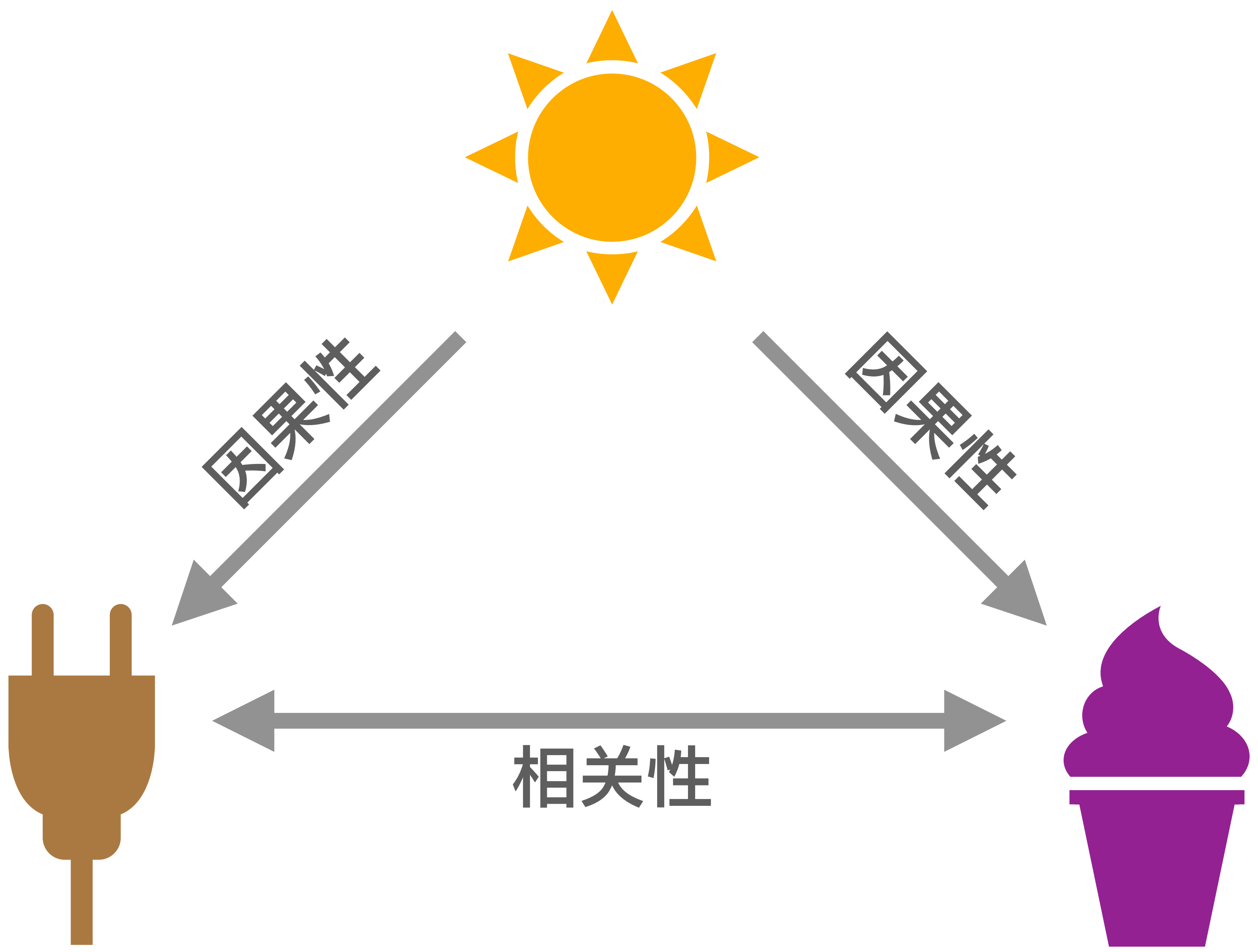
第一部分

因果推断简介

- 相关性vs因果性
- 因果推断研究问题&研究方法

因果推断

因果性&相关性



天气与用电量有因果关系：气温上升导致用电量上升

天气与冰淇淋销量有因果关系：气温上升导致冰淇淋销量上升

用电量与冰淇淋销量有相关性：没有因果关系

因果推断

研究问题&研究方法

2个主要问题
- Causal Discovery：因果发现
- Causal Effect：因果效应

2个研究框架
- Structural Causal Models：结构因果模型
- Potential Outcome Framework：潜在结果框架

PART 02

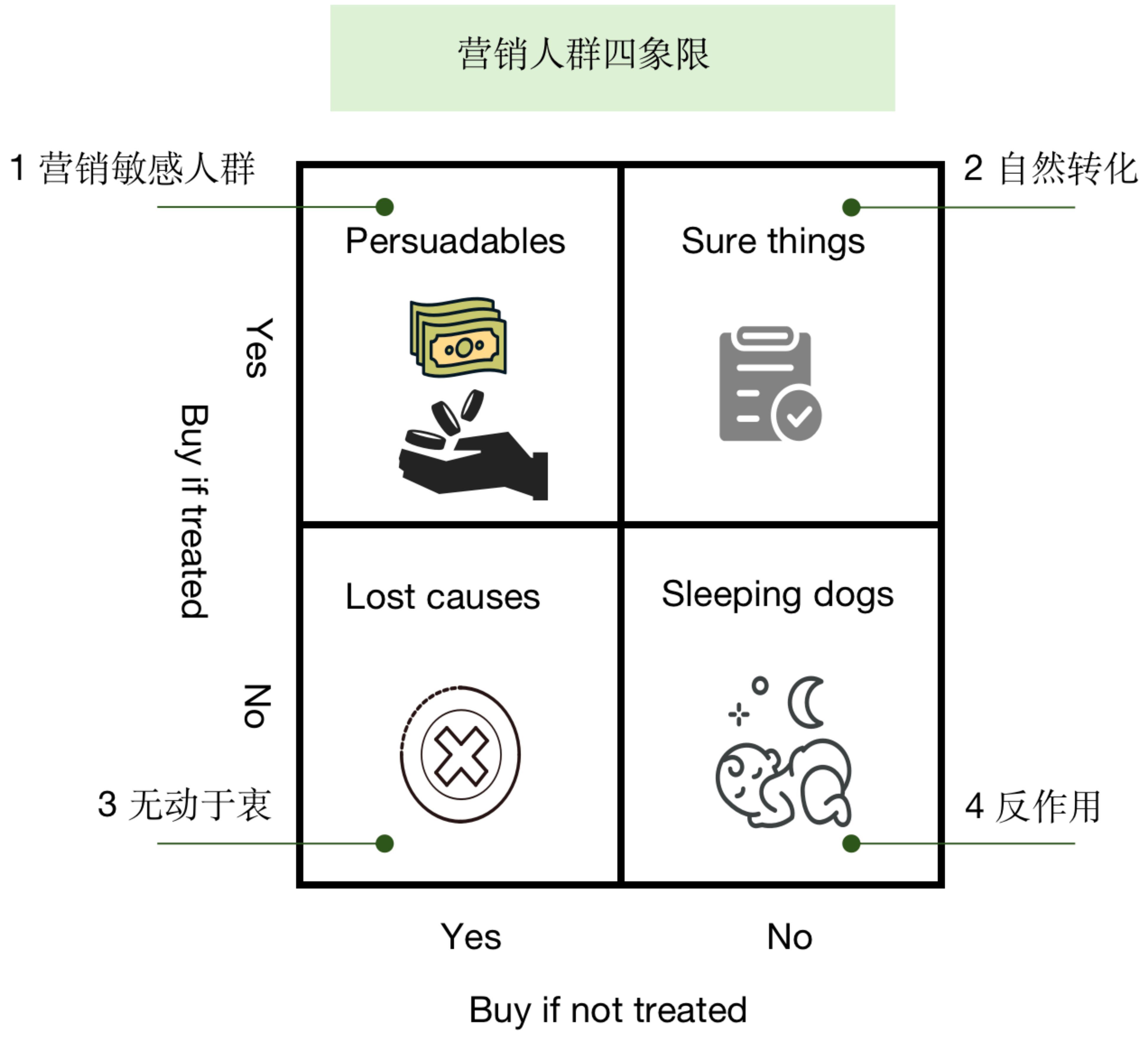
第二部分

因果推断&智能营销

- 传统方法
- 因果效应

因果推断&智能营销

营销4象限



营销敏感人群：无券不购买，发券才购买
自然转化：无论有没有发券，都会购买
无动于衷：无论发不发券，都不会购买
反作用：无券会购买，发券反而不购买

通过发放优惠券，促使用户转化

因果推断&智能营销

智能营销-传统方法

特征 → 模型 → 转化率

X $\quad$ $u(X)$ $\quad$ $P(y|x) = u(x)$

	发券购买率	无券购买率
用户1	1.3%	0.8%
用户2	1.5%	1.4%

根据模型预测结果决定是否给用户发券

机器学习模型预估发放优惠券时用户的转化率

因果推断&智能营销

传统方法是否最优解?

给用户2发券是否为最优解?

	发券购买率	无券购买率
用户1	1.3%	0.8%
用户2	1.5%	1.4%

相关性模型预测结果

1.5% > 1.3%
1不发2发?

假设用户1和用户2各10000人,
不发券价格10元、发券后价格8元,
如何收益最大化?

	发券	不发券
用户1	1040	800
用户2	1200	1400

发券和不发券分别带来的收益

都不发券：10000*0.8%*10+10000*1.4%*10=2200
都发券：10000*1.3%*8+10000*1.5%*8=2240
1不发2发：10000*0.8%*10+10000*1.5%*8=2000
1发2不发：10000*1.3%*8+10000*1.4%*10=2400

因果推断&智能营销

因果效应

Treatment Effect：干预因素变化对目标结果的影响。

问题设定：研究给用户发放优惠券对用户转化率的影响，那么是否发放优惠券就是treatment，假设treatment有两种 $t \in \{0,1\}$，对于用户 i ，目标结果为转化率 y，y^t 表示treatment为 t 时的转化率。

Individual Treatment Effect：$ITE_i = \tau_i = y_i^1 - y_i^0$

Average Treatment Effect：$ATE = E[\tau_i] = E[y_i^1 - y_i^0] = \frac{1}{n}\sum_{i=1}^{n}(y_i^1 - y_i^0)$

Conditional Average Treatment Effect：$CATE = E_{x_i \in X}[\tau_i]$

因果推断&智能营销

传统模型VS因果推断

相关性 $\neq$ 因果性

Response Model：$P(Y=1 \mid X)$，用户购买概率，有label

Uplift Model：$\tau(X)$，干预对用户购买概率的影响，无label

因果推断在营销场景的应用主要是基于Uplift Model预测营销干预带来的增益

PART 03

第三部分

因果推断：Uplift建模

- 建模方法

因果推断：Uplift建模

Uplift

$$\tau(x) = P(y\,|\,x, t = 1) - P(y\,|\,x, t = 0)$$

Uplift模型就是学习给定一个干预（发券）后，结果（转化率）的增量，即对因果效应进行建模。

因果推断：Uplift建模

Uplift建模方法：meta learner

T-Learner

$$u_0 = f(X_0, Y_0)$$
$$u_1 = f(X_1, Y_1)$$
$$\tau(x) = u_1(x) - u_0(x)$$

S-Learner

$$u = f(X, T, Y)$$
$$\tau = u(x, T=1) - u(x, T=0)$$

X-Learner

$$u_0 = f(X_0, Y_0),\ \ u_1 = f(X_1, Y_1)$$
$$D_0 = u_1(X_0) - Y_0,\ \ D_1 = Y_1 - u_0(X_1)$$
$$\tau_0 = f(X_0, D_0),\ \ \tau_1 = f(X_1, D_1)$$
$$e(x) = P(T = 1 \mid X = x)$$
$$\tau(x) = e(x)\tau_0(x) + (1 - e(x))\tau_1(x)$$

R-Learner
Tree-based Learner
...

因果推断：Uplift建模

Uplift建模方法：tree model

Uplift-Tree

$$gain(A) = D(P^T(Y) : P^C(Y)|A) - D(P^T(Y) : P^C(Y))$$

$$D(P^T(Y) : P^C(Y)|A) = \sum_a \frac{N(a)}{N} D(P^T(Y|a) : P^C(Y|a))$$

$$D(P : Q) = KL(P : Q) = \sum_i p_i log\frac{p_i}{q_i}$$

$$\tau(x) = \frac{1}{N^T|x \in l} \sum_{i|x \in l} Y_i - \frac{1}{N^C|x \in l} \sum_{i|x \in l} Y_i$$

Propensity-Tree

估计 $P(T|x)$，使叶子节点中的样本的propensity score一致

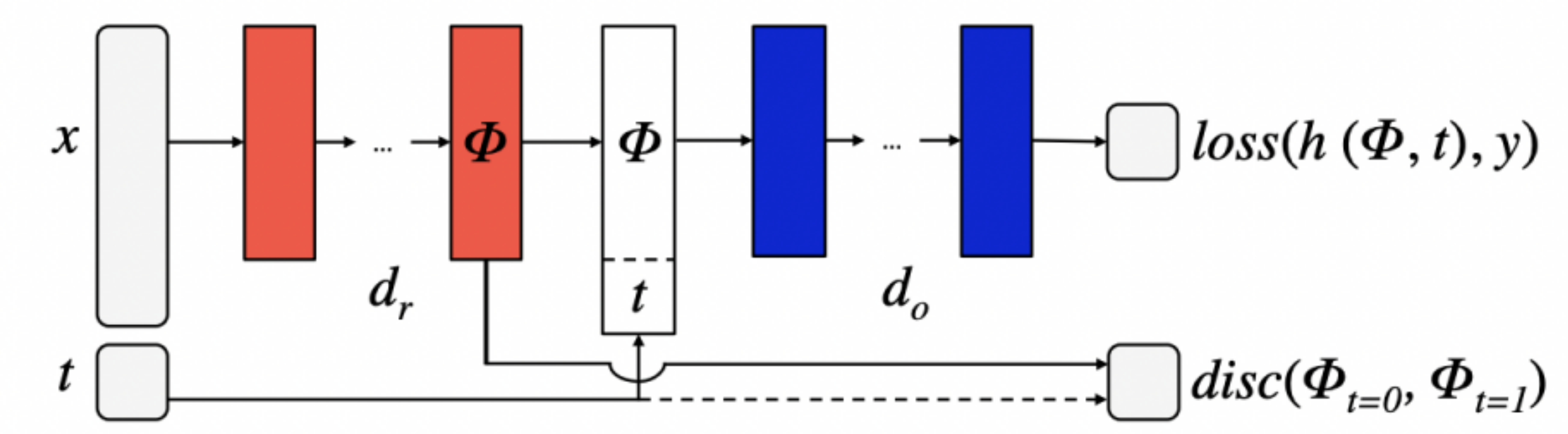
$$\tau(x) = \frac{1}{N^T|x \in l} \sum_{i|x \in l} Y_i - \frac{1}{N^C|x \in l} \sum_{i|x \in l} Y_i$$

Causal Forest
Double-Sample Tree
…

因果推断：Uplift建模

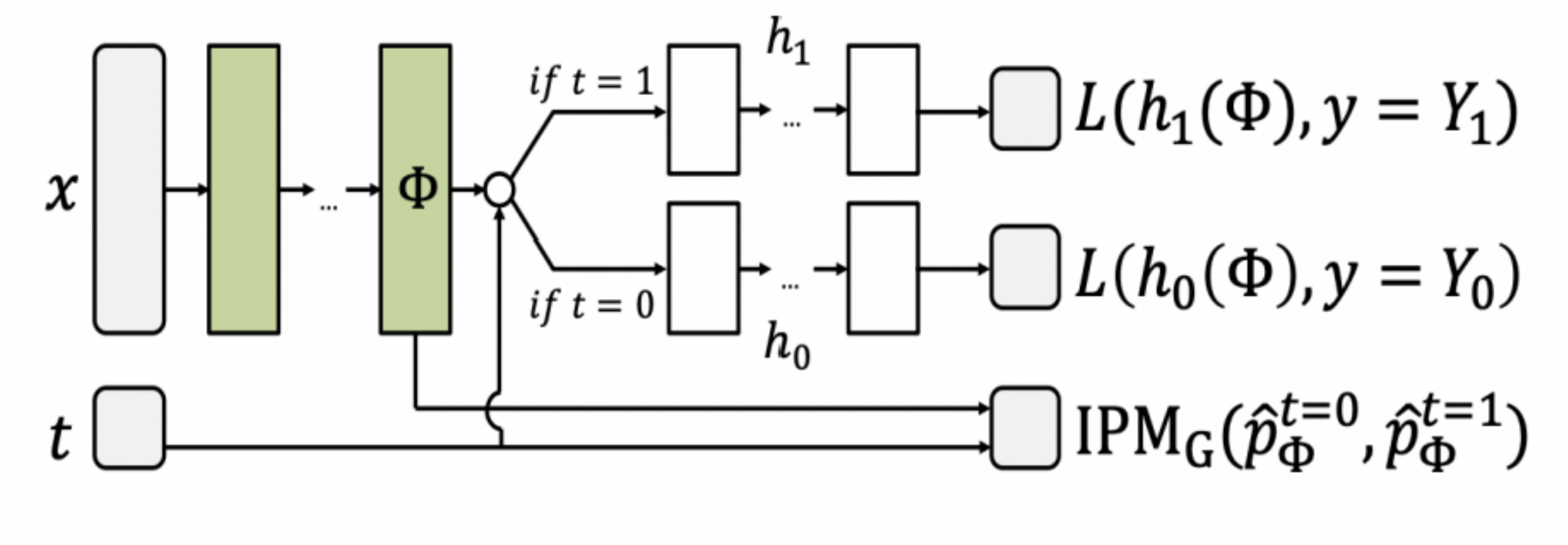
Uplift建模方法：deep learning　　通过表征网络来解决unobserved confounder的问题，使事实分布和反事实分布一致

BNN



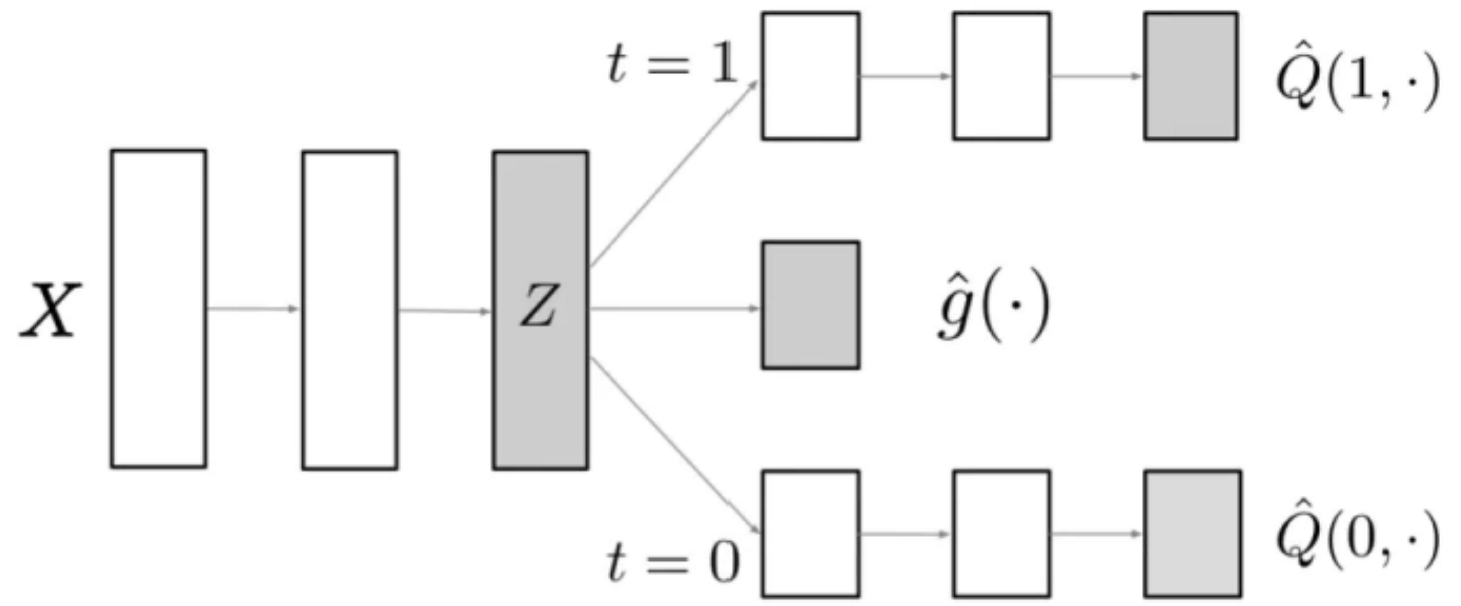
$$B_{\mathcal{H},\alpha,\gamma}(\Phi,h) = \frac{1}{n}\sum_{i=1}^{n}|h(\Phi(x_i),t_i) - y_i^F| + \alpha\,\mathrm{disc}_{\mathcal{H}}(\hat{P}_\Phi^F,\hat{P}_\Phi^{CF}) + \frac{\gamma}{n}\sum_{i=1}^{n}|h(\Phi(x_i),1-t_i) - y_{j(i)}^F|\,,$$

TarNet



$$\min_{\substack{h,\Phi\\ \|\Phi\|=1}} \frac{1}{n}\sum_{i=1}^{n} w_i \cdot L\left(h(\Phi(x_i),t_i)\,,y_i\right) + \lambda\cdot\mathfrak{R}(h) + \alpha\cdot\mathrm{IPM_G}\left(\{\Phi(x_i)\}_{i:t_i=0},\{\Phi(x_i)\}_{i:t_i=1}\right),$$

DragonNet



$$\hat{R}(\theta;\mathbf{X}) = \frac{1}{n}\sum_i\left[(Q^{\mathrm{nn}}(t_i,x_i;\theta)-y_i)^2 + \alpha\mathrm{CrossEntropy}(g^{\mathrm{nn}}(x_i;\theta),t_i)\right],$$

$$\hat{\theta},\hat{\varepsilon} = \operatorname*{argmin}_{\theta,\varepsilon}\left[\hat{R}(\theta;\mathbf{X}) + \beta\frac{1}{n}\sum_i\gamma(y_i,t_i,x_i;\theta,\varepsilon)\right].$$

PART 04

第四部分

因果推断在哈啰的应用

- 树模型应用
- 深度模型探索

Uplift模型在哈啰智的应用

Uplift Tree

1. 假设分裂前数据集为θ，计算分裂前发券组和无券组样本人均收益的差异：

$$D(G^T : G^C) = (\frac{1}{N^T}\sum_i G^T - \frac{1}{N^T}\sum_i G^T)^2$$

2. 根据某个特征 f 将数据集 θ 分成左右两个子集θ_1, θ_2，计算分裂后的人均收益差异：

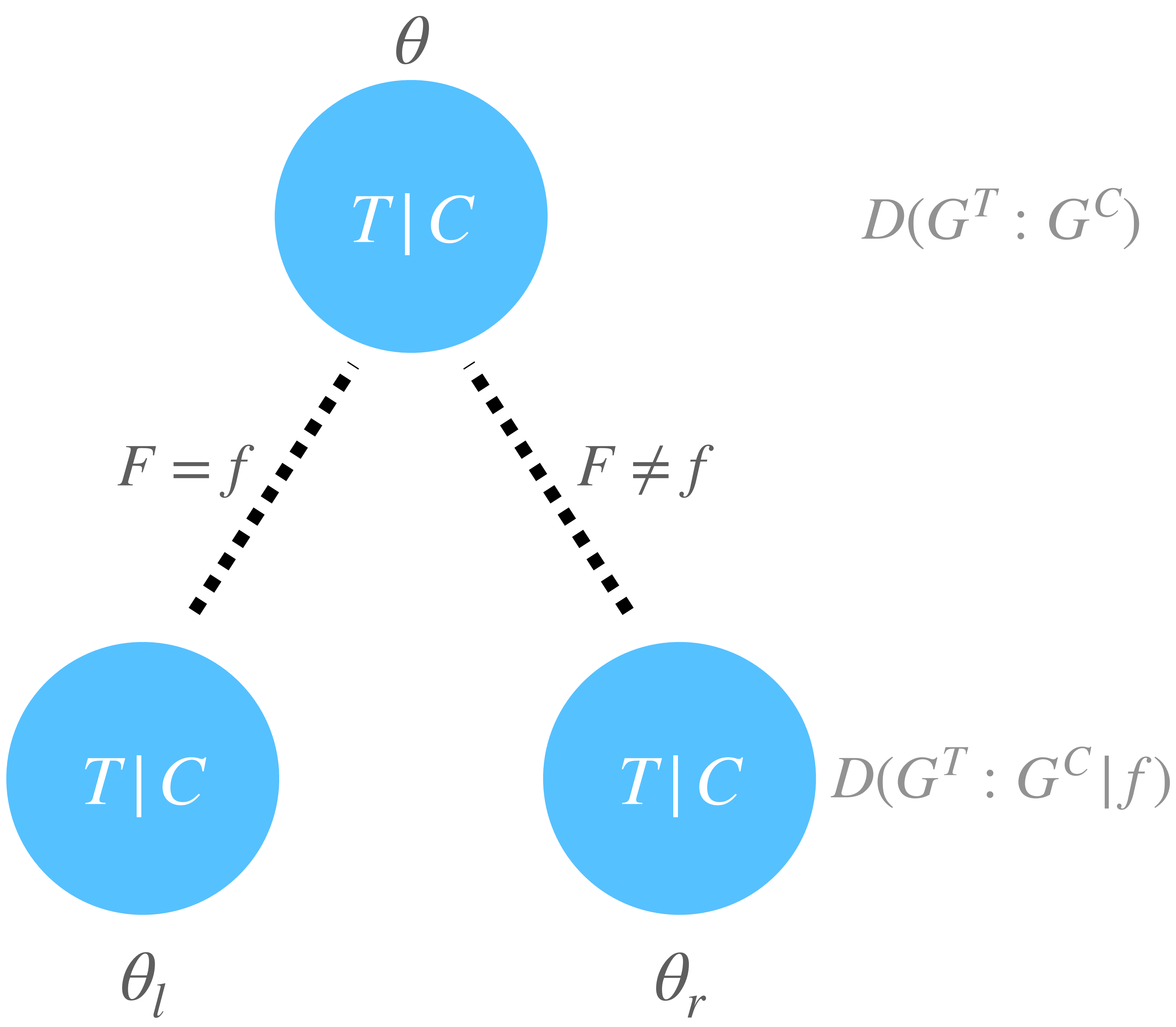
$$D(G^T : G^C | f) = \sum_{a \subseteq \{\theta_l, \theta_r\}} \frac{N(a)}{N} D(G^T(a) : G^C(a))$$

3. 分裂后的人均收益差异减去分裂前的人均收益差异，得到D_{gain}：

$$D_{gain} = D(G^T : G^C | f) - D(G^T : G^C)$$

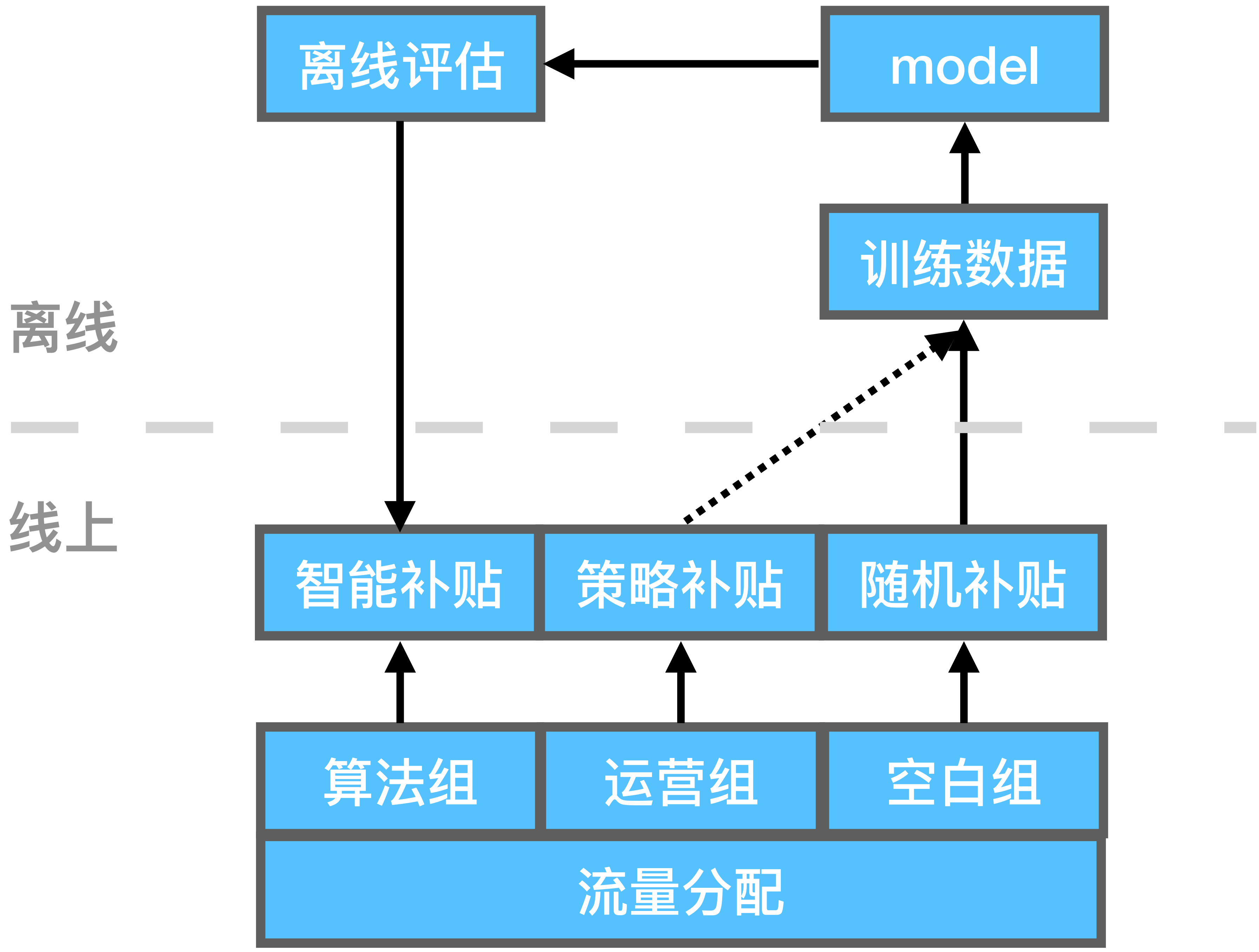
4. 遍历数据集 θ 中所有特征值，重复2、3步，取D_{gain}最大时对应的特征值作为节点分裂值，将数据集分为θ_1, θ_2。

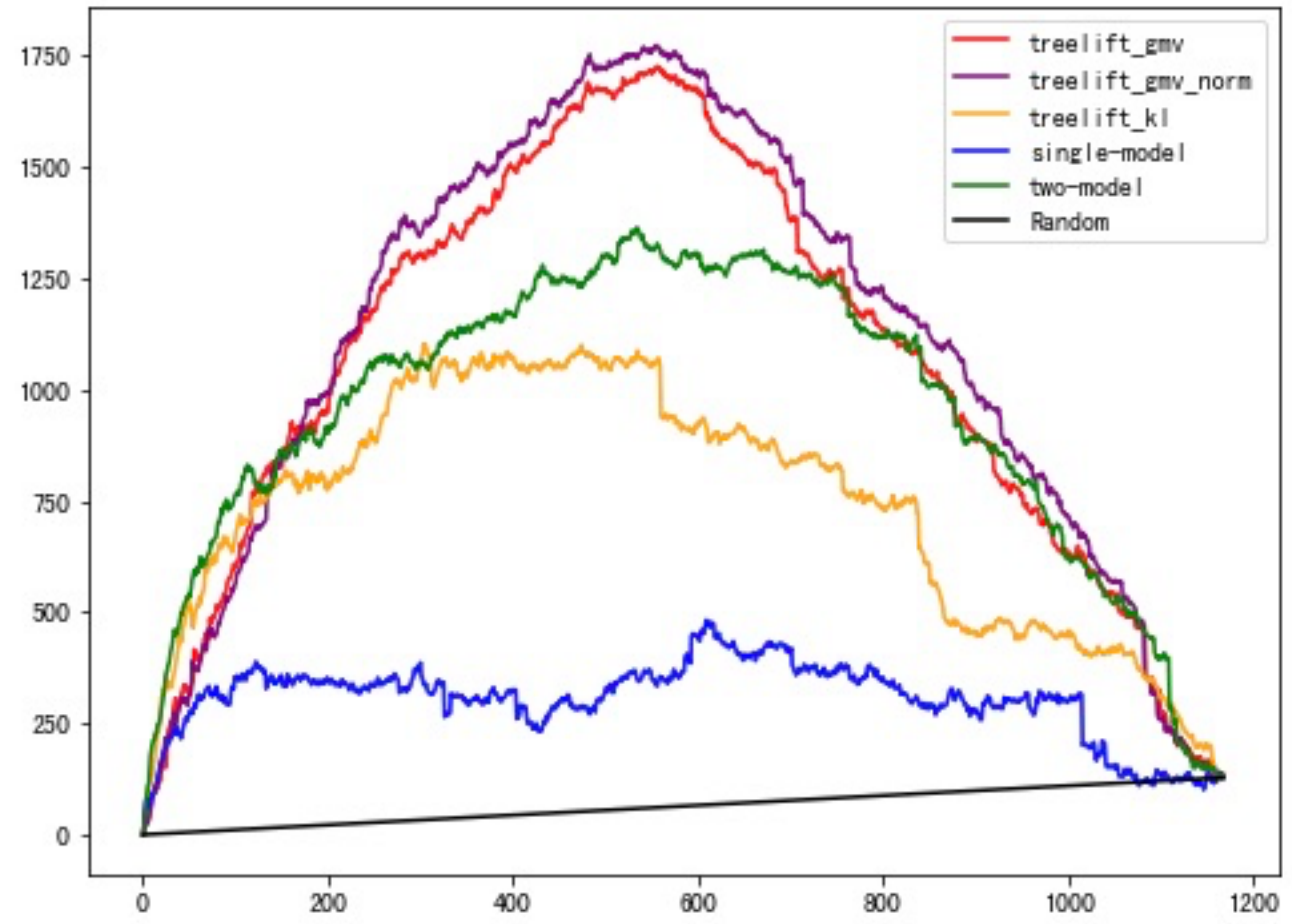
5. 对数据集 θ_1, θ_2 分别重复上述步骤，生成树。



Uplift模型在哈啰智的应用

Uplift Tree





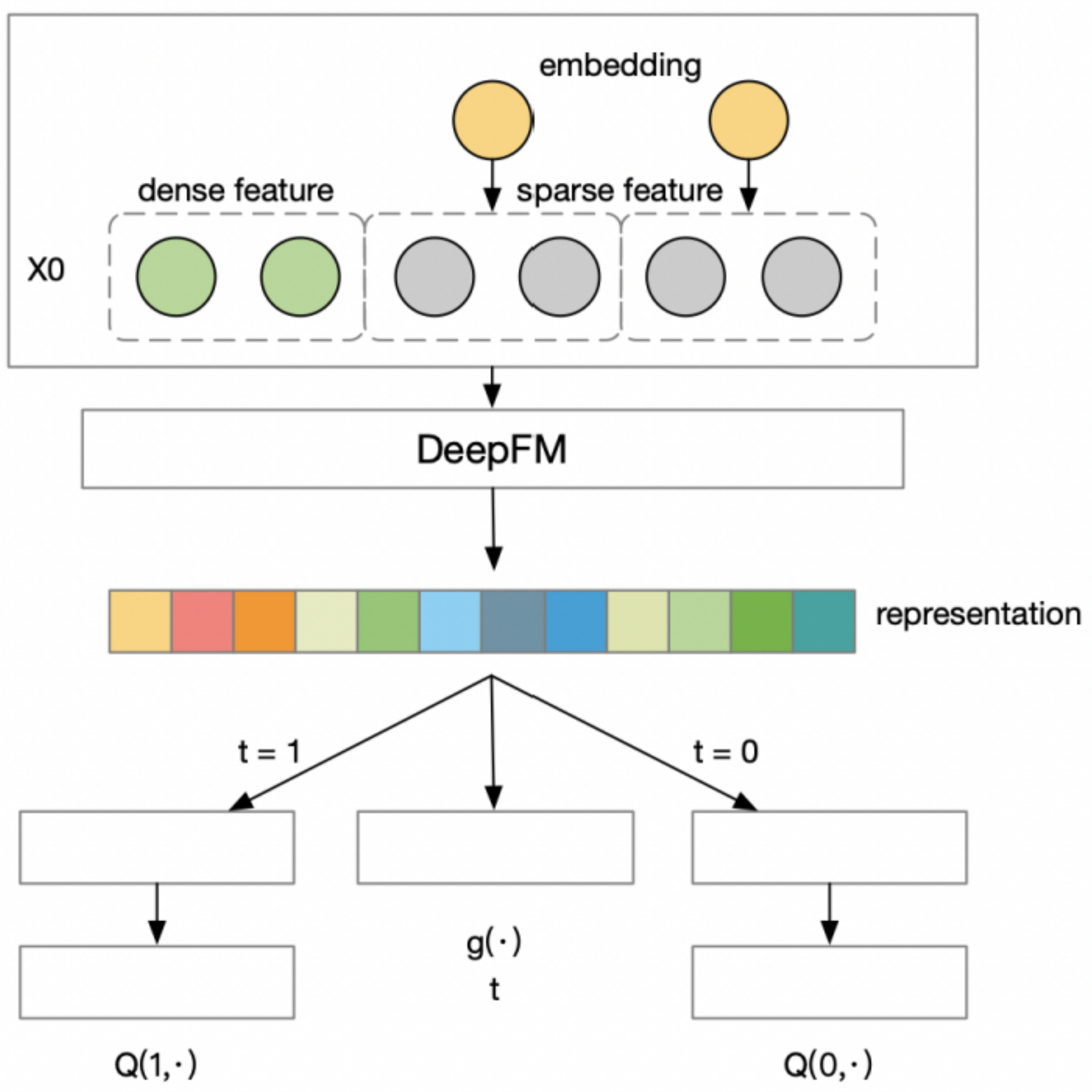
模型离线对比，uplift tree最优

分流	人均收益	实验组提升幅度
实验组	xx	-
Response Model	xx	2.3%
策略组	xx	4.7%

线上AB测试结果，uplift tree最优

Uplift模型在哈啰智的应用

DragonNet



$$\hat{R}(\theta;\boldsymbol{X})=\frac{1}{n}\sum_{i}\left[(Q^{\mathrm{nn}}(t_i,x_i;\theta)-y_i)^2+\alpha\mathrm{CrossEntropy}(g^{\mathrm{nn}}(x_i;\theta),t_i)\right],$$

$$\hat{\theta},\hat{\varepsilon}=\underset{\theta,\varepsilon}{\mathrm{argmin}}\left[\hat{R}(\theta;\boldsymbol{X})+\beta\frac{1}{n}\sum_{i}\gamma(y_i,t_i,x_i;\theta,\varepsilon)\right].$$

哈啰技术

微信扫描二维码，关注我的公众号

THANKS
Q&A

评估方法

反事实

无增量标签

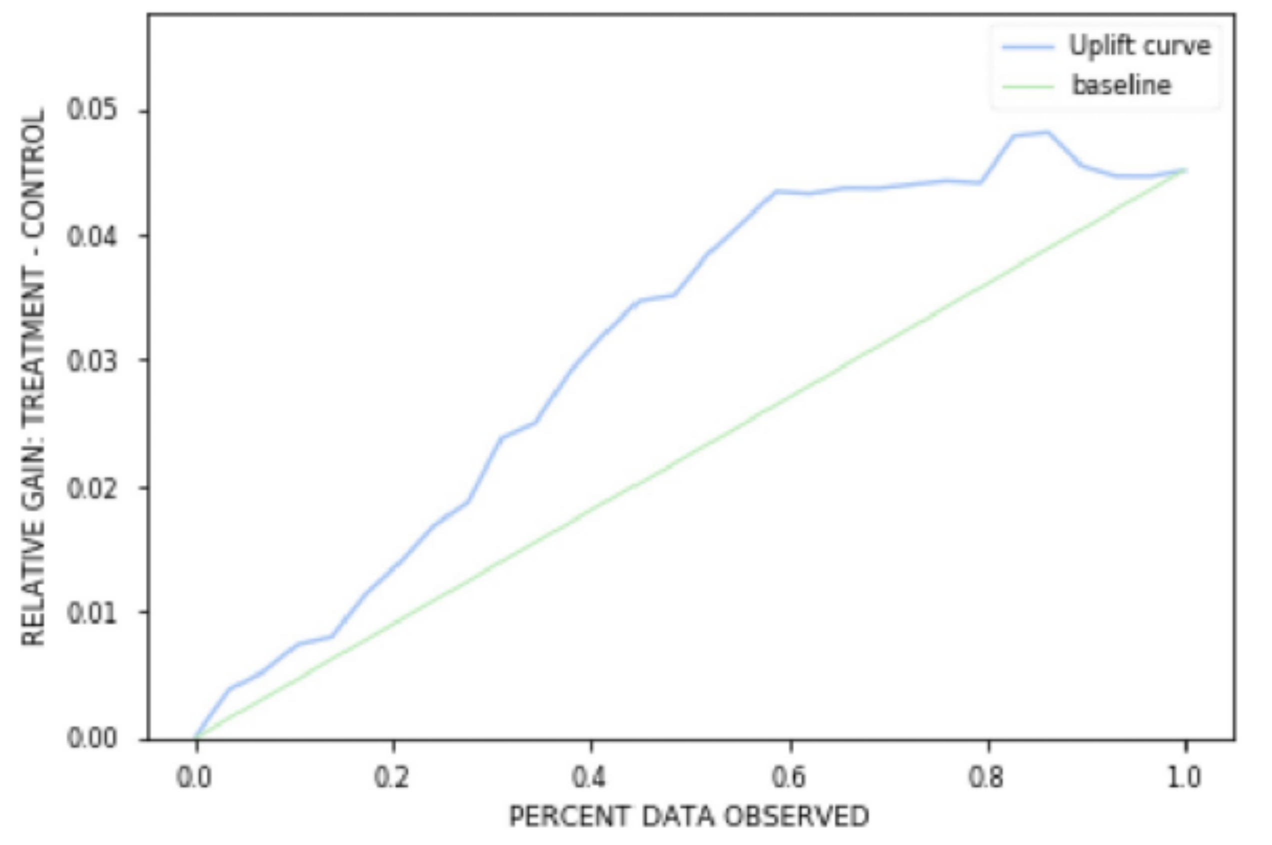
离线：AUUC

线上：AB对照

AUC
准确率
RMSE

$$f(t) = (\frac{Y_t^T}{N_t^T} - \frac{Y_t^C}{N_t^C})(N_t^T + N_t^C)$$

$$AUUC = \sum^{N} f(t)$$



AUUC计算流程

1. 测试集进入模型输出uplift score
2. 所有样本按照uplift score降序排序
3. 分桶，桶序号记为t
4. uplift curve: $f(t) = (\frac{Y^T}{N^T} - \frac{Y^C}{N^C})(N^T + N^C)$,
   Y表示分组正例数量， N 表示分组总量，
   $.^T$表示Treatment组， $.^C$表示Control组。
5. $AUUC = \sum_{t=0}^{N} f(t)$