



# 3rd Party Cookie 規制後の広告配信技術

Privacy Sandbox の紹介とその課題

福島大祐

MicroAd inc. システム開発本部

人工知能学会全国大会（第 38 回）インダストリアルセッション

- ① オンライン広告における機械学習の活用と 3rd Party Cookie の規制
- ② Privacy Sandbox
- ③ Attribution Reporting API (ARAPI) によって生じる課題の検討



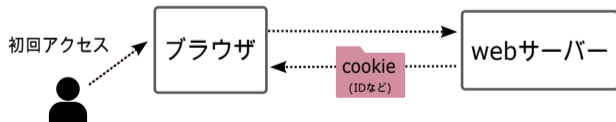
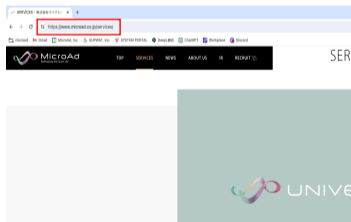
# 1st Party Cookie と 3rd Party Cookie

## Cookie (クッキー) とは

- WEBサイトを閲覧したときに、訪問者が訪れたサイトや入力したデータ・利用環境などの情報が記録されたファイル(仕組み)
- ログイン情報の保持などにも活用

**1st Party Cookie** : ブラウザ URL 欄に表示されているドメイン (サイト) が発行

**3rd Party Cookie** : ブラウザ URL 欄のドメインとは異なる他ドメインが発行



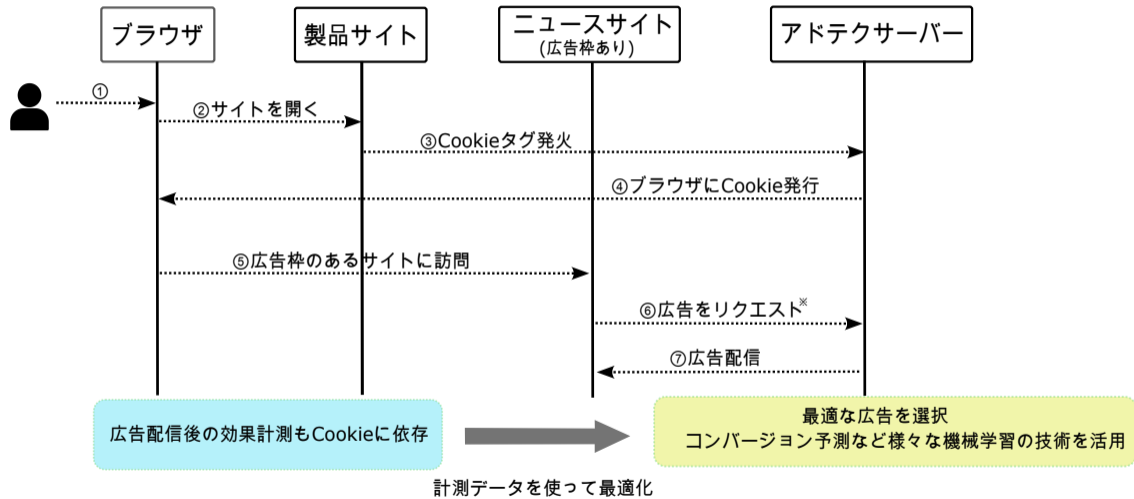
3rd partyのcookieタグが埋まっていれば, 3rd party cookieも発行される

## 3rd Party Cookie はドメイン横断でデータの管理が可能

### 広告配信における 3rd Party Cookie の活用例

- **効果計測**
  - ▶ 広告のクリックから商品を購入したという一連の流れを計測する仕組み
- **ターゲティング**
  - ▶ オーディエンスの趣向に合わせた広告を配信する仕組み
- **フリークエンシー制御**
  - ▶ 同一の広告が執拗に表示されたりといったことが起きないように表示頻度を制御する仕組み

# 機械学習の活用例



\* RTB (Real-Time-Bidding)など

# ユーザーデータ保護の動き

## 一部抜粋して紹介

- 2016年5月: GDPR EU 一般データ保護規則制定
- 2017年9月: safari ITP<sup>1</sup> 1.0 開始
  - ▶ 3rd party Cookie の有効期間の制限
- 2020年1月: Google が Chrome における 3rd party Cookie の段階的な廃止を発表
  - ▶ 2025年初頭を目安に段階的に廃止予定 (2度目の延期)
- 2020年3月: safari の 3rd party cookie を制限 (ITP 2.3)
- 2020年7月: CCPA カリフォルニア州消費者プライバシー法適用開始

※ ブラウザの国内シェアは Chrome と Safari が 8割を占める (2022時点)

ユーザーデータを保護しつつ、広告配信を最適化できる仕組みが必要 !!

---

<sup>1</sup>Intelligent Tracking Prevention

- ① オンライン広告における機械学習の活用と 3rd Party Cookie の規制
- ② Privacy Sandbox
- ③ Attribution Reporting API (ARAPI) によって生じる課題の検討



# Privacy Sandbox

## Privacy Sandbox

Google の提唱するプライバシー保護と広告の効率化を両立するための代替技術の総称

具体的には, Chrome で実装されるいくつかの API から構成

- **Protected Audience API**

- ▶ デバイス上で広告オークションを実施 (ユーザー情報を外部へ送信しないリターゲ広告)

- **Attribution Reporting API**

- ▶ 3rd party Cookie を利用しない効果計測を実施

- **Topics API**

- ▶ 追跡を公開しない興味・関心トピックの取得

etc.

現在 30 個以上の API が発表及び提供されている





# Attribution Reporting API

2種類のレポートが利用可能

- Aggregatable Report
- Event-level Report : 機械学習の学習に利用可能な想定

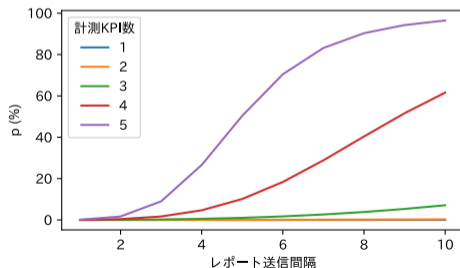
Attribution Reporting API のレポートには**一定確率でノイズが混入** (プライバシー保護のため)  
レポートタイミングも一定間隔ごとになる

Event-level Report の場合

ノイズの発生確率  $p$

$$p = \frac{k}{k + \exp(\epsilon) - 1}$$

- $k$ : とりうるパターン数<sup>1</sup>
- $\epsilon$ : プライバシー保護係数



対象の計測地点を増やすほどにノイズ確率が上昇

<sup>1</sup>設定値によって変化  
Fukushima Daisuke (MicroAd inc.)

- ① オンライン広告における機械学習の活用と 3rd Party Cookie の規制
- ② Privacy Sandbox
- ③ Attribution Reporting API (ARAPI) によって生じる課題の検討



# CVR 予測の例 (配信する広告の選択などで利用)

配信ログを用いて、広告を表示した際にコンバージョンに至る確率を予測

データセット ( 配信ログ ) 例

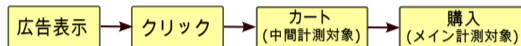
広告ID	URL	ユーザーID		曜日	クリック	コンバージョン
34345	http://aaa.co.jp/hoge	XXXXXXXXXX	...	金曜日	0	0
11671	http://bbb.com/fuga	YYYYYYYYYY	...	木曜日	1	1
31221	http://bbb.com/hoge	ZZZZZZZZZ	...	木曜日	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮

## 既存の主要な課題と研究

- 正例 (conversion=1) が極端に少ない (コンバージョンに至るのは 0.1%以下)
    - ▶ 学習が不安定になる
- ⇒ 中間計測地点を定義したマルチタスク学習
- 遅れコンバージョン (クリックからコンバージョンに至るまでにタイムラグが存在)
    - ▶ 将来的には正例でも、計測されるまでは負例として扱われてしまう
- ⇒ コンバージョンの遅れをモデル化

# [既存研究] 中間 KPI のマルチタスク学習

コンバージョンに至るまでには複数のステップが存在



- 本来達成したいコンバージョンはメイン計測対象
  - メインの計測対象のコンバージョン数は少数
    - ▶ 正例が少なすぎる場合, 学習が不安定
- ⇒ 中間計測対象を含めたマルチタスク学習

左図はよくあるマルチタスク学習の構成

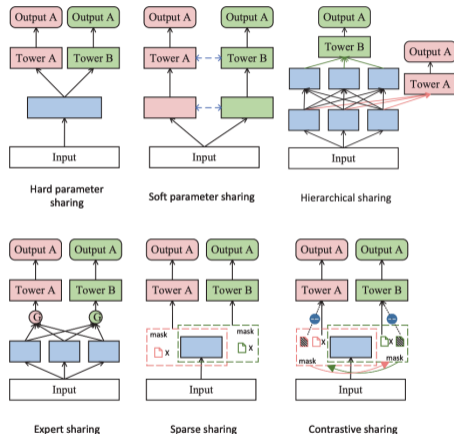
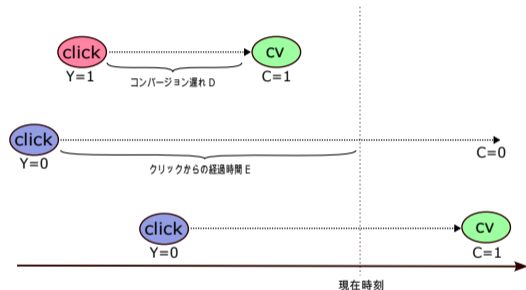


Figure: [MINGZHU et al., 2023]<sup>2</sup> から引用

<sup>2</sup>Advances and Challenges of Multi-task Learning Method in Recommender System: A Survey

# [既存研究] Delayed Feedback Model (DFM)

## 遅れコンバージョン問題



以下のことが成立

- $Y = 0 \Leftrightarrow C = 0 \quad \text{or} \quad E < D$
- $Y = 1 \Rightarrow C = 1$

DFM [Olivier Chapelle, 2014]<sup>a</sup>では、遅れ分布を別のパラメトリックモデルで定義

コンバージョン発生確率

$$\Pr(C = 1|X = \mathbf{x}) = p(\mathbf{x}) = \frac{1}{1 + \exp(-\mathbf{w}_c^\top \cdot \mathbf{x})}$$

遅れ分布

$$\Pr(D = d|X = \mathbf{x}, C = 1) = \lambda(\mathbf{x}) \exp(\lambda(\mathbf{x})d)$$

このとき

$$\lambda(\mathbf{x}) = \exp(\mathbf{w}_d^\top \cdot \mathbf{x})$$

# [既存研究] Delayed Feedback Model (DFM)

コンバージョンが観測される確率

$$\begin{aligned}\Pr(Y = 1, D = d | X = \mathbf{x}) &= \Pr(C = 1, D = d | X = \mathbf{x}) \\ &= \Pr(D = d | X = \mathbf{x}, C = 1) \Pr(C = 1 | X = \mathbf{x}) \\ &= \lambda(\mathbf{x}) \exp(-\lambda(\mathbf{x})d) p(\mathbf{x})\end{aligned}$$

コンバージョンが観測されない確率 (式変形省略)

$$\begin{aligned}\Pr(Y = 0 | X = \mathbf{x}, E = e) &= \Pr(Y = 0 | C = 0, X = \mathbf{x}, E = e) \Pr(C = 0 | X = \mathbf{x}) \\ &\quad + \Pr(Y = 0 | C = 1, X = \mathbf{x}, E = e) \Pr(C = 1 | X = \mathbf{x}) \\ &= \dots \\ &= 1 - p(\mathbf{x}) + p(\mathbf{x}) \exp(-\lambda(\mathbf{x})e)\end{aligned}$$

上記から尤度関数を定義して EM アルゴリズムなどで最適化

※ 現在では, NN でより柔軟に遅れ分布をモデリングするなどの派生もあり

# ARAPIによる計測によって生じる問題

計測地点を増やすほどに大幅にノイズ確率が上昇

⇒ マルチタスク学習のための中間計測に制限がかかる

レポートのタイミングが一定間隔ごとに (間隔を短くするほどノイズ確率が上昇)

⇒ クリックからコンバージョンまでの正確な経過時間が不明

(正確な遅れ分布のモデル化が難しくなる)

ノイズによる精度悪化の懸念に加え、既存の研究にも大きな影響

- 近い将来, 主要なブラウザで 3rd Party Cookie が使えなくなる
- Chrome では, 代替手法として **Privacy Sandbox** が登場<sup>3</sup>
- ARAPI を活用すれば 3rd Party Cookie に依存しない効果計測は可能であるが, ノイズ混入の問題が発生
- 今までの手法も影響を受けるため, **ノイズを考慮したアプローチが必要**

---

<sup>3</sup>Post Cookie 対応としては, ID solution など他のアプローチも存在



ご清聴ありがとうございました

ブース出展してるので、興味があればぜひ細かい話聞きに来てください