Threshold Influence Model for Allocating Advertising Budgets

宮内 敦史[1]　岩政 勇仁[2]　福永 拓郎[3]　垣村 尚徳[2]

[1] 東京工業大学

[2] 東京大学

[3] 国立情報学研究所

2015 年 8 月 3 日

河原林巨大グラフプロジェクト
ERATO Kawarabayashi Large Graph Project

ICML 2015

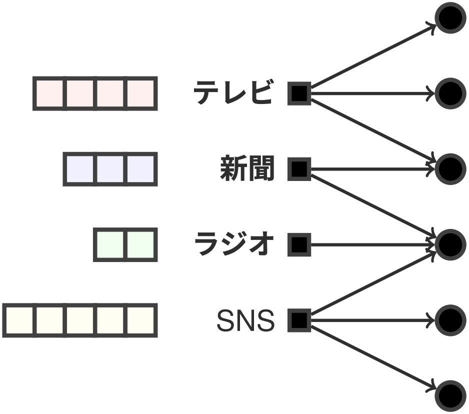
The 32nd International Conference on Machine Learning

- **期間：**7 **月** 6 **日から** 11 **日**
- **場所：**Lille Grand Palais**（リール，フランス）**

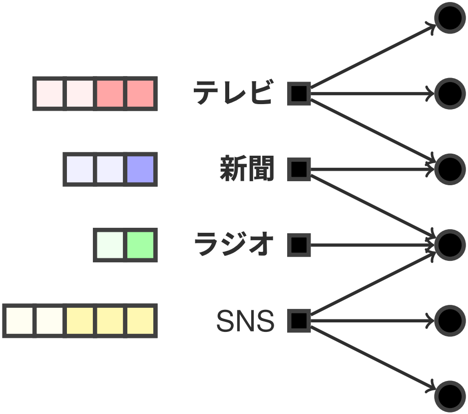
今年の統計

- 1,600 **人以上が参加**
- 270 **本の採択論文** (**採択率：**26%)

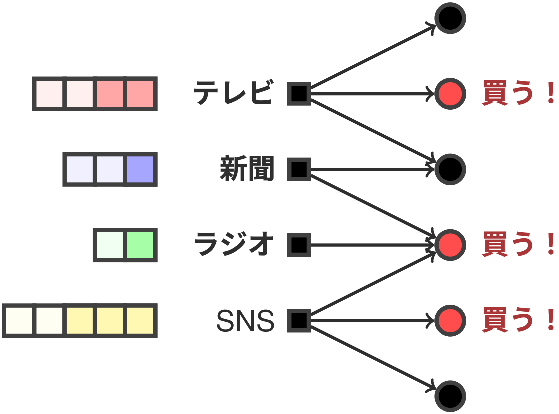
広告予算配分



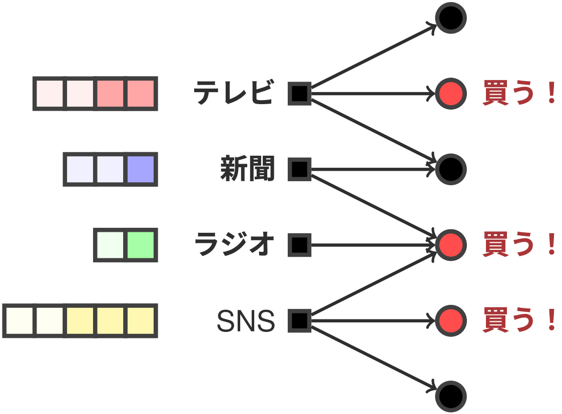
広告予算配分



広告予算配分



広告予算配分



疑問点

- **予算配分が顧客に及ぼす影響はどのようにモデル化できる？**
- **最適化すべき指標は？**

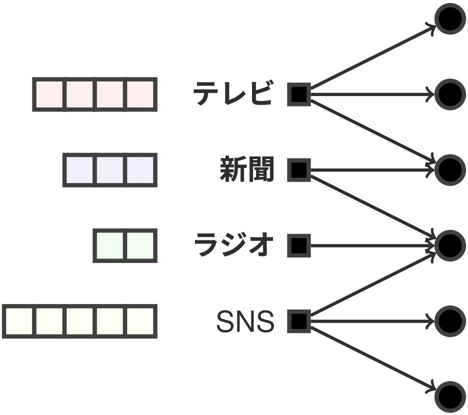
本研究の成果

閾値影響モデル

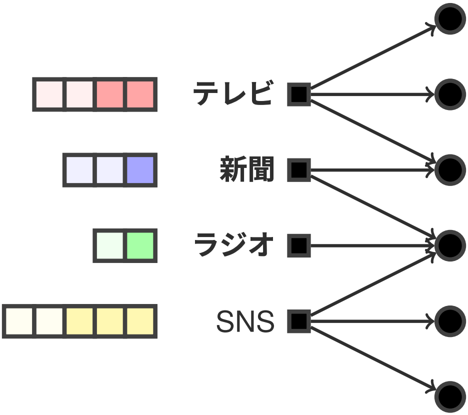
各顧客が活性化するか否かを閾値的な振る舞いとして表現

1. 予算制約付き影響最大化
   - 2 種類の貪欲解法
   - 影響関数が劣モジュラの場合
2. コスト効率性最大化
   - 高速な精度保証付き近似解法
   - 線形計画法を用いた精度保証付き近似解法

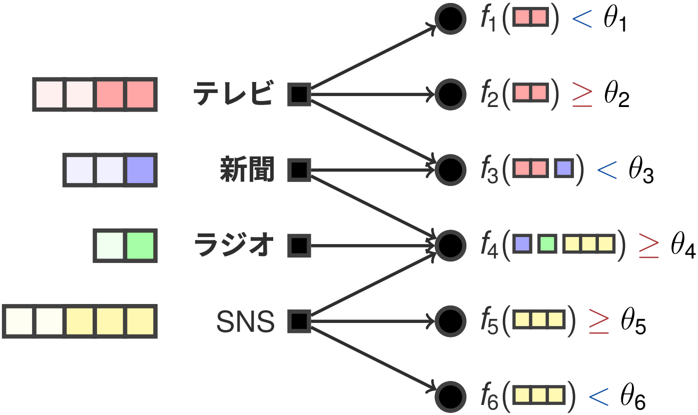
閾値影響モデル（概要）



閾値影響モデル（概要）

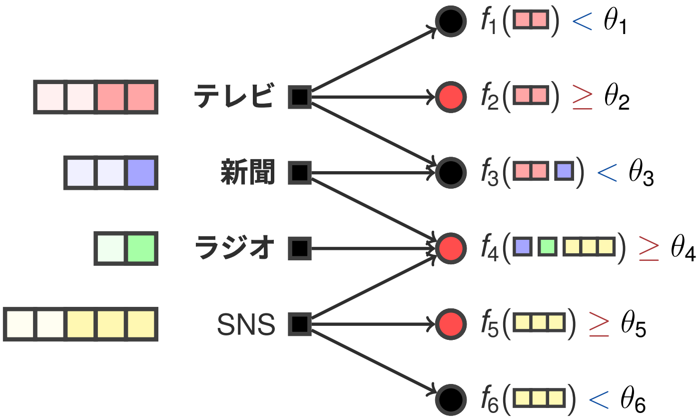


閾値影響モデル（概要）



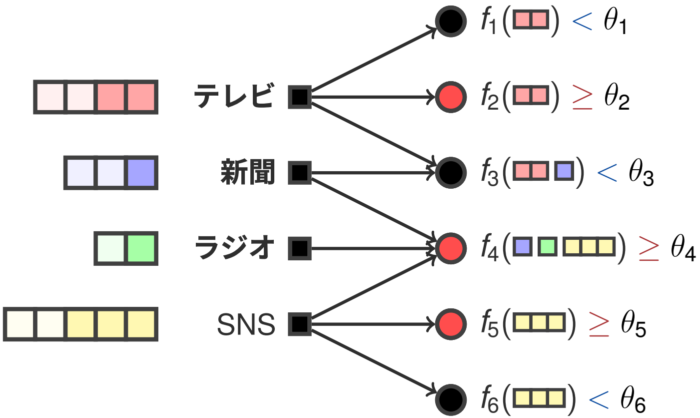
テレビ
新聞
ラジオ
SNS
$f_1(\ ) < \theta_1$
$f_2(\ ) \geq \theta_2$
$f_3(\ ) < \theta_3$
$f_4(\ ) \geq \theta_4$
$f_5(\ ) \geq \theta_5$
$f_6(\ ) < \theta_6$

閾値影響モデル（概要）



テレビ
新聞
ラジオ
SNS
$f_1(\ ) < \theta_1$
$f_2(\ ) \geq \theta_2$
$f_3(\ ) < \theta_3$
$f_4(\ ) \geq \theta_4$
$f_5(\ ) \geq \theta_5$
$f_6(\ ) < \theta_6$

閾値影響モデル（概要）



予算配分は整数格子点として表現できる e.g., (2, 1, 1, 3)

閾値影響モデル（定義）

二部グラフ $G = (S, T; E)$

- **各頂点 $s \in S$:**
  - **配分可能な予算の上限 $u_s \in \mathbb{Z}_+$**
- **各頂点 $t \in T$:**
  - **単調影響関数 $f_t : \mathbb{Z}_+^{\Gamma(t)} \to \mathbb{R}_+$ ($\Gamma(t)$: $t \in T$ の隣接点の集合)**
  - **閾値 $\theta_t \in \mathbb{R}_+$**

予算配分 $\boldsymbol{x} \in \mathbb{Z}_+^S$ に対して，$f_t(\boldsymbol{x}_t) \geq \theta_t$ のとき，頂点 $t \in T$ は活性化

- **$\boldsymbol{x} \in \mathbb{Z}_+^S$ の影響：** $f_G(\boldsymbol{x}) = |\{t \in T \mid f_t(\boldsymbol{x}_t) \geq \theta_t\}|$
- **$\boldsymbol{x} \in \mathbb{Z}_+^S$ のコスト：** $c(\boldsymbol{x}) = \sum_{s \in S} x_s$

単位価格 $c_s \in \mathbb{R}_+$ や重要度 $w_t \in \mathbb{R}_+$ をもつモデルも定義できる

本研究の成果

閾値影響モデル

各顧客が活性化するか否かを閾値的な振る舞いとして表現

1. 予算制約付き影響最大化
   - 2 種類の貪欲解法
   - 影響関数が劣モジュラの場合
2. コスト効率性最大化
   - 高速な精度保証付き近似解法
   - 線形計画法を用いた精度保証付き近似解法

本研究の成果

閾値影響モデル

各顧客が活性化するか否かを閾値的な振る舞いとして表現

1. **予算制約付き影響最大化**
    - 2 種類の貪欲解法
    - 影響関数が劣モジュラの場合
2. **コスト効率性最大化**
    - 高速な精度保証付き近似解法
    - 線形計画法を用いた精度保証付き近似解法

予算制約付き影響最大化

予算 $B \in \mathbb{Z}_+$

Maximum General-Thresholds Coverage Problem

$$\begin{array}{lll} \text{Maximize} & f_G(\boldsymbol{x}) & \\ \text{subject to} & c(\boldsymbol{x}) \leq B & \text{(予算制約)} \\ & \boldsymbol{x} \leq \boldsymbol{u} & \text{(上限制約)} \\ & \boldsymbol{x} \in \mathbb{Z}_+^S & \text{(予算配分)} \end{array}$$

特殊ケース

- **最大被覆問題**
- **最密 k-部分グラフ問題** (Kortsarz & Peleg ’93)
- Target-side influence model (Alon et al. ’12)

センサ配置，文書要約，特徴選択，などにも応用をもつ

2 種類の貪欲解法

Incremental Greedy:
最大被覆問題に対する貪欲解法を一般化 (Hochbaum '97)

1: **空な配分** $\boldsymbol{x} = (0, \ldots, 0)$ **をとる**
2: **影響の増加分が最大の** $s \in S$ **に予算を配分**

- $O(\beta \sum_{s \in S} u_s |E| + |T|)$ **時間** (β: **関数** f_t **の値の計算に要する時間**)

Decremental Greedy:
最密 k**-部分グラフ問題に対する "最小次数頂点除去法" を一般化**
(Asahiro et al. '00)

1: **上限の配分** $\boldsymbol{x} = (u_1, \ldots, u_{|S|})$ **をとる**
2: **"貢献度" が最小の** $s \in S$ **の予算を削る**

- $O(\sum_{s \in S} u_s + |T| + \beta \sum_{t \in T} \sum_{s \in \Gamma(t)} u_s)$ **時間**
  $\longrightarrow$ u_s **が定数であれば，** $O(|S| + |T| + \beta|E|)$ **時間**

影響関数 f_t が劣モジュラの場合

整数格子上の関数 $f : \mathbb{Z}_+^n \to \mathbb{R}_+$ **が劣モジュラ**

$$\overset{\text{def}}{\iff} \quad f(\boldsymbol{x}) + f(\boldsymbol{y}) \geq f(\boldsymbol{x} \vee \boldsymbol{y}) + f(\boldsymbol{x} \wedge \boldsymbol{y}) \qquad (\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{Z}_+^n)$$

($\boldsymbol{x} \vee \boldsymbol{y}$ **は成分ごとの最大，** $\boldsymbol{x} \wedge \boldsymbol{y}$ **は成分ごとの最小**)

仮定

- f_t: **値域が** $[0, 1]$ **の単調劣モジュラ関数**
- θ_t: $[0, 1]$ **内の一様ランダムな値**

$$\text{Maximize} \quad \sum_{t \in T} f_t(\boldsymbol{x}_t) \quad \text{（影響の期待値）}$$

$$\text{subject to} \quad c(\boldsymbol{x}) \leq B,\ \boldsymbol{x} \leq \boldsymbol{u},\ \boldsymbol{x} \in \mathbb{Z}_+^S$$

定数近似解法

1: 次の条件を満たすインスタンスに帰着
   - 予算の上限 $u_s = 1$ $(\forall s \in S)$
   - 目的関数が単調劣モジュラ
   - 擬多項式サイズ

2: 集合関数に対する $(e/(e-1))$-近似解法を適用 (Sviridenko ’04)

$\longrightarrow$ 擬多項式時間で $(e/(e-1))$-**近似解**が得られる

- Soma et al. (ICML 2014) の別証明
- 整数格子上の単調劣モジュラ関数を扱う他の問題にも適用可能

本研究の成果

閾値影響モデル

各顧客が活性化するか否かを閾値的な振る舞いとして表現

1. **予算制約付き影響最大化**
    - 2 種類の貪欲解法
    - 影響関数が劣モジュラの場合
2. コスト効率性最大化
    - 高速な精度保証付き近似解法
    - 線形計画法を用いた精度保証付き近似解法

本研究の成果

閾値影響モデル

各顧客が活性化するか否かを閾値的な振る舞いとして表現

1. 予算制約付き影響最大化
   - 2 種類の貪欲解法
   - 影響関数が劣モジュラの場合
2. コスト効率性最大化
   - 高速な精度保証付き近似解法
   - 線形計画法を用いた精度保証付き近似解法

コスト効率性最大化

予算配分 $\boldsymbol{x} \in \mathbb{Z}_+^S$ のコスト効率性：

$$d(\boldsymbol{x}) = \frac{f_G(\boldsymbol{x})}{c(\boldsymbol{x})}$$

$\boldsymbol{x}$ の影響

$\boldsymbol{x}$ のコスト

Cost-Effective General-Thresholds Coverage Problem

$$\begin{aligned} \text{Maximize} \quad & d(\boldsymbol{x}) & \\ \text{subject to} \quad & \boldsymbol{x} \le \boldsymbol{u} & \text{(上限制約)} \\ & \boldsymbol{x} \in \mathbb{Z}_+^S & \text{(予算配分)} \end{aligned}$$

- **広告予算配分の予算額の目安を知る上で有用**
- NP **困難**

精度保証付き近似解法

Fast Approximation:
本質的には Decremental Greedy **と同じ**

1: **予算** $B = 0$ **として** Decremental **を実行し，予算配分の列を得る**
2: **その列のなかで，コスト効率性が最大のものを出力**

- $\max_{t \in T} \sum_{s \in \Gamma(t)} u_s$-**近似**
- Decremental Greedy **と同じ実行時間**

LP-based Approximation:
ある特殊ケース (NP **困難**) **に対してのみ適用可能**

1: **線形計画問題の最適解から予算配分の列を得る**
2: **その列のなかで，コスト効率性が最大のものを出力**

- $\max_{t \in T} (\sum_{s \in \Gamma(t)} u_s - \theta_t + 1)$-**近似**
- **擬多項式時間**

数値実験 (CPU: Intel Core i7 2.4 GHz, Memory: 16 GB RAM)

提案解法 Incremental **と** Decremental **の性能評価**

グラフ

- **実グラフ** (Yahoo! Bidding Data)
  - **"検索キーワード" と "アカウント" のグラフ**
  - $(|S|, |T|, |E|) \approx (1\text{k}, 10\text{k}, 53\text{k})$
- **人工グラフ**
  - S **に属す頂点の次数分布がべき則**
  - $(|S|, |T|, |E|) \approx (20\text{k}, 200\text{k}, 2000\text{k})$

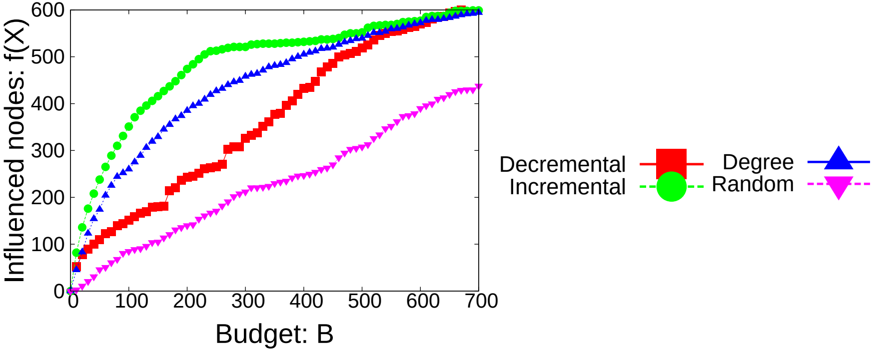
比較手法

- Degree (**次数が大きい頂点** $s \in S$ **から予算を配分**)
- Random (**ランダムに頂点** $s \in S$ **を選択して予算を配分**)

実グラフ $(|S|, |T|, |E|) \approx (1k, 10k, 53k)$

設定 (1)

- f_t: **劣モジュラ**
- θ_t: **区間** [0, 1] **から一様ランダム**

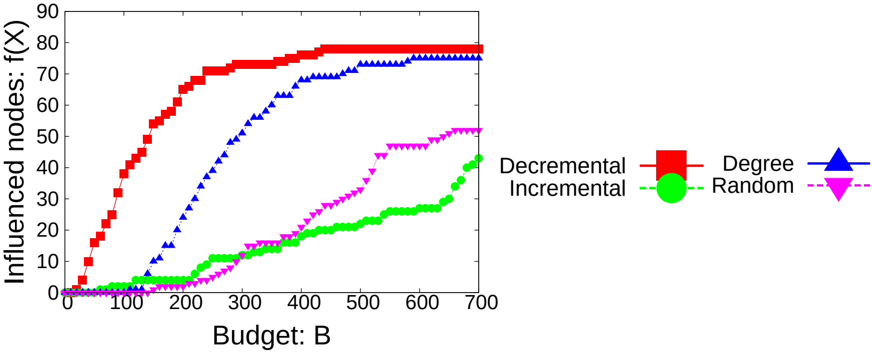


- **どの予算額でも** Incremental **が最大の影響**

実グラフ $(|S|, |T|, |E|) \approx (1k, 10k, 53k)$

設定 (2)

- f_t: 劣モジュラでない
- θ_t: 区間 [0.5, 1] から一様ランダム

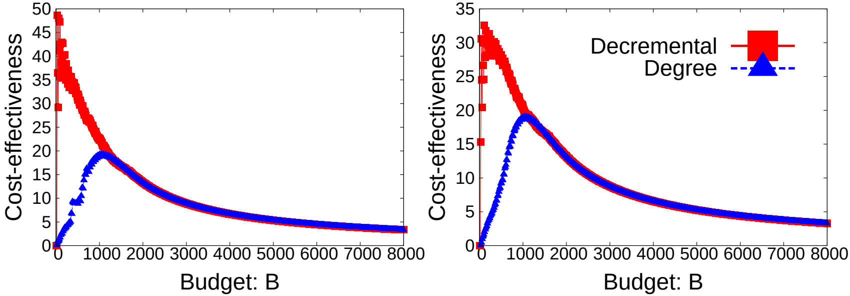


- どの予算額でも Decremental が最大の影響
- Decremental の実行時間は 1 秒程度

人工グラフ $(|S|, |T|, |E|) \approx (20k, 200k, 2000k)$

設定

- f_t: 左図では劣モジュラ，右図では劣モジュラでない
- $\theta_t = 0.8$ $(\forall t \in T)$



- Decremental がより高いコスト効率性
- Decremental の実行時間は数分程度

本研究の成果（再掲）

閾値影響モデル

各顧客が活性化するか否かを閾値的な振る舞いとして表現

1. 予算制約付き影響最大化
   - 2 種類の貪欲解法
   - 影響関数が劣モジュラの場合
2. コスト効率性最大化
   - 高速な精度保証付き近似解法
   - 線形計画法を用いた精度保証付き近似解法