CF-010

オンライン環境において公平な資源配分を実現するアルゴリズムに関する研究

山田 博瑛 * 小宮山 純平 † 阿部 拳之 ‡ 岩崎 敦 *
Hakuei Yamada Junpei Komiyama Kenshi Abe Atsushi Iwasaki

概要

本研究では，オンライン環境において公平な資源配分を実現するアルゴリズムを扱う．オンライン環境とは，災害時の避難所にどんな物資が届けられるかや，工場の勤務シフトを作成する際にどんな注文がやってくるかが，事前にはわからない状況を指す．資源配分における公平性には様々な概念があるが，本研究では参加者全員の効用の積であるナッシュ積を最大化することを目的とし，オンライン環境におけるフィッシャー市場の均衡解を求めるアルゴリズムを提案する．とくに，届けられる資源の価値が事前にはわからない環境で，アルゴリズムが事後的な最適解を漸近的に達成することを確認する．

1 はじめに

限られた財を効率的に配分する問題は，現代社会において非常に重要な課題である．また，効率性に加え，公平性も考慮した配分を求めるアルゴリズムの研究が多く行われている [1, 4]．そこで，本研究ではより実問題に対応させるべく，オンライン環境において公平な資源配分を実現するアルゴリズムを扱う．オンライン環境とは，財が逐次的に到着する状況を指しており，以下にオンライン環境の例を示す．

例えば，避難所へ送る救援物資の配分では，物資としては米やパンといった食料や，布団や毛布のような寝具といった相異なる種類の物資が送られてくると考えられる．毎日，何らかの物資が到着するが，どの種類の物資が到着するかは当日にならないと分からないうえ，被災者も物資を受け取って初めて，その必要度がわかることがある．つまり，被災者は各種類の物資に関して何らかの評価値をもつが，事前には自身でさえ明確にその値は分からないとする．このような状況では，限りある資源を効率的に配分するだけでなく，物資を受け取る人の間でのある種の公平性を満たすことも要求される．

次にクラウドソーシングの例を考える．このとき，財に対応するものはタスクであり，タスクが逐次的に到着するとみなすことができる．ここでタスクを発注するクライアントはワーカーにタスクを与えるが，ワーカーのタスクに対する成果やその質は実際にタスクを行ってもらった後でないと分からない．この状況下で，タスクが発生したときにどのようにワーカーに割り振りするかを考える．もちろんクライアントはなるべくタスクとの相性が良い有能なワーカーに割り振りたいが，特定の人にタスクが集中してしまい，ワーカーの新規参入を阻害してしまう恐れがある．そのため，単にタスクを効率的に割り当てるだけでなく，ワーカー間で公平に割り当てることも考慮する必要がある．

公平な資源配分のためには様々な概念があるが，本研究ではプレイヤ達の効用の積であるナッシュ積 [8] を最大化することを目的とする．ここで，オフラインもしくは静的環境におけるナッシュ積の最大化問題はフィッシャー市場の市場均衡として，凸計画問題に帰着できることが知られている [9]．しかし，本研究ではオンライン環境を対象としているため，双対平均化法 [11] に着目してこの凸計画問題をオンライン環境に拡張する．双対平均化法を用いたオンラインアルゴリズムには文献 [2] があるが，本研究はこれを拡張し，プレイヤの真の評価値が事前に分からないという設定を考える．具体的に多腕バンディットアルゴリズム [5, 7] を用いて，真の評価値をアルゴリズム内で推測しながら配分を決定するアルゴリズムを DA-Explore-then-Commit (DA-EtC) と DA-Upper Confidence Bound (DA-UCB) として提案する．これらのアルゴリズムは，バンディットフィードバックを通じて財の評価値を徐々に学習していき，その結果得られた評価の推定値を双対平均化法に適用し，割り当てを求めている．また，静的なフィッシャー市場による凸計画問題の最適解と提案手法による解との誤差をリグレットとして定義し，リグレットの最小化を目指すことで，静的な状況下での公平で効率的な配分に近づくことを目標としている．また，DA-EtC によるリグレットの上界を $\tilde{O}(T^{2/3}(nm)^{1/3})$，手法を問わないリグレットの下界を $\Omega(\sqrt{mT})$ で収束することを示した．さらに，ランダムに生成したデータと実データの両方において提案アルゴリズムの性能を評価した結果，長期的にリグレットが収束することが分かった．

2 モデル

本章ではオンライン環境での資源配分問題のモデル化を行う．まず，プレイヤの集合を $N = \{1, 2, \ldots, n\}$，財の種類の集合を $M = \{1, 2, \ldots, m\}$ とする．次に，プレイヤ i がある

* 電気通信大学大学院情報理工学研究科
† ニューヨーク大学
‡ 株式会社サイバーエージェント

財 j に対して持つ評価値を $v_{i,j} \in [0,1]$ と定義する．ここで，この評価値 $v_{i,j}$ は事前には分からないものとする．

財が与えられる回数を T 回とし，各ラウンド $t = \{1,2,\ldots,T\}$ に対して 1 つ財 $j(t) \in M$ が到着する場合を考える．このとき，m 種類から財を決定する確率の分布を $\mathcal{S} = \{s_1, s_2, \ldots, s_m\}$ と定義する．なお $\mathcal{S}$ は i.i.d. であると仮定する．ラウンド t で到着した財は一人のプレイヤ $i(t) \in N$ にのみ割り当てられ，財を受け取ったプレイヤ $i(t)$ は価値のノイズ付き評価 $u_i(t) = v_{i(t),j(t)} + \varepsilon_t$ を観測する．ここで，ε_t は平均 0，分散 σ^2 のサブガウシアン確率変数とする．

次に，T ラウンド終了後のプレイヤ i の累積効用 $U_i(T)$ を以下に定義する:

$$U_i(T) = \sum_{t:i(t)=i} u_i(t).$$

ここで財の評価値は加法的であると仮定する．さらに T ラウンド終了後のプレイヤ i のナッシュ積 NSW(T) を各プレイヤの効用の加重幾何平均として以下に定義する:

$$\mathrm{NSW}(T) = \prod_{i \in N} U_i(T)^{B_i}.$$

ここで B_i は所与のプレイヤ i の優先度を表しており，市場モデルでは予算として解釈する．また，$\sum_i B_i = 1$ と仮定する．

本研究ではこのナッシュ積の値を最大化する配分を見つけることが目的であり，提案手法のベンチマークとして，与えられる財の種類や評価値が事前に既知であるときの問題を Eisenberg-Gale (EG) 凸計画問題 [1, 4] とし，その最適解を用いる．ここでは財を分割可能と仮定し，種類 j の財をプレイヤ i に割り振る割合もしくは確率を $x_{i,j} \in [0,1]$ とする．1 ラウンド当たりのナッシュ積の最大化問題を以下に定義する:

$$\begin{aligned} \text{maximize}_{\{x_{i,j}\}} \quad & \prod_{i \in N} \left(\sum_{j \in M} s_j v_{i,j} x_{i,j} \right)^{B_i} \\ \text{subject to} \quad & \forall j \in M : \sum_{i \in N} x_{i,j} \leq 1, \\ & \forall i \in N,\ \forall j \in M : x_{i,j} \geq 0. \end{aligned} \tag{1}$$

次に，この最大化問題の解を ONSW とすると，T ラウンド分の ONSW と NSW(T) の差をリグレットとできる：

$$\mathrm{Regret}(T) = T \cdot \mathrm{ONSW} - \mathrm{NSW}(T). \tag{2}$$

本研究ではリグレットの最小化を目指す．なお，式 1 の最大化問題は $\sum_{i \in N} B_i \log \sum_{j \in M} v_{i,j} x_{i,j}$ の最大化問題に帰着できるので，これを目的関数としたとき，対応する双対問題は以下になる:

$$\begin{aligned} \text{minimize} \quad & \sum_{j \in M} s_j p_j - \sum_{i \in N} B_i \log \beta_i \\ \text{subject to} \quad & \forall i \in N,\ \forall j \in M :\ p_j \geq \beta_i v_{i,j}, \\ & \forall j \in M :\ p_j \geq 0. \end{aligned} \tag{3}$$

この双対問題における双対変数は p_j と β_i であり，p_j は財 j の「価格」を示している．ここでは双対性ギャップは存在せず, パラメータの値が有理数であるときに有理凸計画 (*rational convex program*) [9] に属する．したがって，内点法などのアルゴリズムを用いることで，多項式時間で求解可能な問題になっている [10].

しかし，本研究では真の評価値 $v_{i,j}$ が未知であると仮定しているため，この双対問題をそのまま解けばいいという訳ではなく，そのためのアルゴリズムを適用できない．そこで，我々は双対平均化法 [11] に着目し，漸近的に最適解に収束するアルゴリズムを提案する．

3 双対平均化法 (Dual Averaging)

本節では双対平均化法 [11] を概説する．これはある種の最適化問題をオンラインアルゴリズムに拡張する方法であり，Gao らはプレイヤがもつ真の評価値が既知であるときのオンラインアルゴリズムを，式 3 に示す双対問題を用いて構築している [2]．Algorithm 1 および 2 は双対平均化法の具体的な流れを示す．まず初めに各ラウンドがスタート時に財が確率分布 $\mathcal{S}$ から選ばれ，各プレイヤは到着した財に対する評価値を観測する．そしてラウンド t，評価値，累積平均効用 $\bar{u}_i^{\mathrm{DA}}$ を入力とした Algorithm 2 から誰に財を割り当てるかを決定する．Algorithm 2 における変数 $\beta_i \in \mathbb{R}^+$ は割当を決める際にどのプレイヤに優先して割り当てるかを表す優先度に対応しており，その定義域を $[B_i/(h(1+\delta_0)), (1+\delta_0)/l]$ とする．ただし，l, m, δ_0 は外生変数である．ラウンド t 時点で財が割り当てられたプレイヤ $i(t)$ の次ラウンド時点の $\beta_{i(t)}$ の値は減少し，割り当てられなかったプレイヤ $i'(\neq i(t))$ の $\beta_{i'}$ は増加する．

双対平均化法を実行するには，各ラウンドにおいて，財を割り当てる前に，その評価値がわかっていなければならない．しかし，本論文は事前に真の評価値が分からない状況を仮定しているため，Algorithm 2 に渡す $\{v_{i,j(t)}\}_i$ を推定するために Explore-then-Commit (EtC) [5] または Upper Confidence Bound (UCB) [7] と双対平均化法を合わせた 2 つの手法 DA-EtC および DA-UCB を提案する．

4 提案手法

4.1 DA-EtC

EtC は未知の値の探索とその利用をバランスよく行う手法であり，Algorithm 3 にこの EtC を双対平均化法に取り入れたアルゴリズム DA-EtC を示す．これは最初の $T_0 < T$ ラウンドまで評価値の推測値を探索し，各ラウンドで到着した財を一様ランダムに配分する．T_0 回終了後，Algorithm 3 の 7 行目に示す通り，推測値 $\hat{v}_{i,j}^{\mathrm{EtC}}$ を計算する．なお $N_{i,j}(t)$ は $t-1$ ラウンドまでにプレイヤ i に財 j を割り当てた回数であり，$\mathbb{I}[\cdot]$ は指示関数である．推測した $\hat{v}_{i,j}^{\mathrm{EtC}}$ は Algorithm 2 への入力である $\{\hat{v}_{i,j(t)}^{\mathrm{DA}}(t)\}_i$ として利用する．DA-EtC は T_0 回以降，$\hat{v}_{i,j}^{\mathrm{EtC}}$ を固定して更新しないものとする．

Algorithm 1 DA with true values $\{v_{i,j}\}$

1: Initialize utility $\bar{u}_i^{\mathrm{DA}} = 0$ for each i.
2: **for** $t = 1$ to T **do**
3: Observe the type $j(t)$ of an arriving item, drawn from $\mathcal{S}$.
4: Observe the value $v_{i,j(t)}$ of the item for each i.
5: Update the mean utilities $\{\bar{u}_i^{\mathrm{DA}}(t+1)\}_i = \text{DA-Iter}(t, \{v_{i,j(t)}\}_i, \{\bar{u}_i^{\mathrm{DA}}(t)\}_i)$
6: **end for**

Algorithm 2 DA-Iter

Require: $t, \{\hat{v}_{i,j(t)}^{\mathrm{DA}}(t)\}_i, \{\bar{u}_i^{\mathrm{DA}}(t)\}_i$
1: Define multiplier β_i as

$$\beta_i = \mathrm{Proj}_{[B_i/(h(1+\delta_0)),(1+\delta_0)/l]}(B_i/\bar{u}_i^{\mathrm{DA}}(t)) \quad (\delta_0 > 0),$$

where $\mathrm{Proj}_{[a,b]}(x) = \max(a, \min(b, x))$.
2: Agents bid $\beta_i \hat{v}_{i,j(t)}^{\mathrm{DA}}(t)$.
3: Winner is determined: $i(t) = \arg\max_i \beta_i \hat{v}_{i,j(t)}^{\mathrm{DA}}(t)$, where ties are broken arbitrarily.
4: Winner $i(t)$ pays $\beta_{i(t)} \hat{v}_{i,j(t)}^{\mathrm{DA}}(t)$.
5: Each agent receives a utility: $u_i^{\mathrm{DA}}(t) = \mathbb{I}\{i = i(t)\} \hat{v}_{i,j(t)}^{\mathrm{DA}}(t)$.
6: Update mean utility for each agent as

$$\bar{u}_i^{\mathrm{DA}}(t+1) = \frac{t-1}{t}\bar{u}_i^{\mathrm{DA}}(t) + \frac{1}{t}u_i^{\mathrm{DA}}(t)$$

7: **return** $\{\bar{u}_i^{\mathrm{DA}}\}_i$ (and $i(t)$ for DA-EtC, DA-UCB)

4.2 DA-UCB

UCB は未知の値を楽観主義の原則にしたがって，高確率である点を付け加える上限を評価して推測する手法であり，多腕バンディット問題などで用いられる．言い換えると，真の値より大きくなるような UCB 値を見積もり，その値にしたがって意思決定する．Algorithm 4 に示す DA-UCB はこの UCB が推測した評価値を用いて財を割り当てる．その 4 行目で計算した UCB 値を用いて，各ラウンドで最も高い UCB 値 ($= \max_i \hat{v}_{i,j(t)}^{\mathrm{UCB}}(t)$) をもつプレイヤに財を割り当てる．また，DA-EtC とは異なり，全てのラウンドにおいて評価値の推測値 $\hat{v}_{i,j}^{\mathrm{EtC}}$ を更新している．

5 性能評価

5.1 双対平均化法における収束

本節では，既存のアルゴリズムである双対平均化法の性質について述べる．[2] の定理 4 により，効用の誤差の期待値はの上界はすでに示されている．しかし，この定理が成立するためには，各プレイヤに対して $\sum_j s_j v_{i,j} = 1$ として評価値を正則化する必要があり，$v_{i,j}$ が未知の場合は適用することができない．そこで，新たに $l.h$ という変数を設定し，正則化の必要がない収束条件として以下に補題 1 を示す．

Algorithm 3 DA-EtC

1: **for** $t = 1$ to T_0 **do**
2: Observe the type $j(t)$ of an arriving item, drawn from $\mathcal{S}$.
3: Give item $j(t)$ to the agent who is chosen uniformly at random.
4: The agent $i(t)$ receives a utility $u_{i(t)}(t)$.
5: Update the cumulative utility: $U_i(t+1) = U_i(t) + \mathbb{I}[i = i(t)]u_{i(t)}(t)$ for each i.
6: **end for**
7: Fix the estimator:

$$\hat{v}_{i,j}^{\mathrm{EtC}} = \frac{\sum_{t:j(t)=j} u_{i(t)}(t)\mathbb{I}[i = i(t)]}{N_{i,j}(T_0+1)}$$

for each i, j.
8: Initialize the DA's utility $\bar{u}_i^{\mathrm{DA}}(1) = 0$ for each i.
9: **for** $t = T_0 + 1$ to T **do**
10: The type of item t is determined: $j(t) \sim \mathcal{S}$.
11: $t' = t - T_0$
12: $\{\bar{u}_i^{\mathrm{DA}}(t'+1)\}_i, i(t) = \text{DA-Iter}(t', \{\hat{v}_{i,j(t)}^{\mathrm{EtC}}\}_i, \{\bar{u}_i^{\mathrm{DA}}(t')\}_i)$
13: The agent $i(t)$ receives a utility $u_{i(t)}(t)$.
14: Update the cumulative utility: $U_i(t+1) = U_i(t) + u_{i(t)}(t)\mathbb{I}[i = i(t)]$ for each i.
15: **end for**

補題 1. *DA* を T ラウンド行い，財の評価値 $\{v_{i,j}^{\mathrm{DA}}\}$ は決定論的であり，全ての $i \in N$ に対して $l \le \sum_{j\in M} s_j v_{i,j}^{\mathrm{DA}} \le h$ を満たす l, h が存在すると仮定する．ラウンド T 終了後のプレイヤの平均効用を $\bar{u}^{\mathrm{DA}}(T) \in \mathbb{R}^n$ と定義し，$u^{*,\mathrm{DA}} \in \mathbb{R}^n$ を *EG* 問題の解と定義する．このとき，任意の $\delta_0 > 0$ に対して以下の不等式が成立する*:*

$$\mathbb{E}\left[\|\bar{u}^{\mathrm{DA}}(T) - u^{*,\mathrm{DA}}\|^2\right] \le C^{\mathrm{DA}} \frac{6 + \log T}{T}.$$

$\|\cdot\|$ は *l2* ノルムベクトルであり，C^{DA} と $\|v^{\mathrm{DA}}\|_\infty$ は，

$$C^{\mathrm{DA}} = \frac{\|v^{\mathrm{DA}}\|_\infty^2 (1+\delta_0)^6}{l^4 \left(\min_{i\in N} B_i\right)^4} \left(\frac{h^3 \|v^{\mathrm{DA}}\|_\infty^2}{l}\left(\frac{1}{\delta_0}\right)^2 + h^4\right)$$

$\|v^{\mathrm{DA}}\|_\infty = \max_{i\in N} \|v_i^{\mathrm{DA}}\|_\infty$.

次に，以下に示す DA-EtC のリグレットの上界と下界の証明のための補題を示す．

補題 2. *2* つのベクトルを $u^{(1)} = (u_1^{(1)}, \ldots, u_n^{(1)})$ と $u^{(2)} = (u_1^{(2)}, \ldots, u_n^{(2)})$ のように定義する．また，$u_i^{(1)} = \Theta(1)$, $u_i^{(2)} \ge$

Algorithm 4 DA-UCB

1: Initialize the DA's utility $\bar{u}_i^{\mathrm{DA}}(1) = 0$ for each i.
2: **for** $t = 1$ to T **do**
3: Observe the type $j(t)$ of an arriving item, drawn from $\mathcal{S}$.
4: Calculate the UCB value

$$\hat{v}_{i,j(t)}^{\mathrm{UCB}}(t) = \min\left(1, \hat{v}_{i,j(t)}(t) + \sqrt{\log t / 2N_{i,j(t)}(t)}\right) \quad \text{where } \min(1, +\infty) = 1.$$

5: $\{\bar{u}_i^{\mathrm{DA}}(t+1)\}_i, i(t) = \text{DA-Iter}(t, \{\hat{v}_{i,j(t)}^{\mathrm{UCB}}(t)\}_i, \{\bar{u}_i^{\mathrm{DA}}(t)\}_i)$
6: The agent $i(t)$ receives a utility $u_{i(t)}(t)$.
7: Update the cumulative utility: $U_i(t+1) = U_i(t) + u_{i(t)}(t)\mathbb{I}[i = i(t)]$ for each i.
8: Update the estimator:

$$\hat{v}_{i(t),j(t)}(t+1) = \frac{\sum_{\tau \le t : j(\tau) = j} u_{i(t)}(\tau)\mathbb{I}[i = i(t)]}{N_{i,j}(t+1)}$$

9: **end for**

$u_i^{(1)}(1 - r_i)$，$r = \max_i r_i$ に対して以下の不等式が成り立つ．

$$\prod_i (u_i^{(1)})^{B_i} - \prod_i (u_i^{(2)})^{B_i} \le O(r). \tag{4}$$

証明.

$$\prod_i (u_i^{(1)})^{B_i} - \prod_i (u_i^{(2)})^{B_i} \tag{5}$$
$$= \prod_i (u_i^{(1)})^{B_i} - \prod_i \left\{u_i^{(1)}(1 - r_i)\right\}^{B_i} \tag{6}$$
$$\le \prod_i (u_i^{(1)})^{B_i} - \prod_i \left\{u_i^{(1)}(1 - \max_{i'} r_{i'})\right\}^{B_i} \tag{7}$$
$$= \max_{i'} r_{i'} \prod_i (u_i^{(1)})^{B_i}. \tag{8}$$

□

補題 3. $T_0 = o(T)$ と仮定したとき，以下が成立する:

$$\mathbb{E}\left[\prod_i (u_i^{*,\mathrm{DA}})^{B_i} - \prod_i (\bar{u}_i^{\mathrm{DA}})^{B_i}\right] \le \tilde{O}\left(nC^{\mathrm{DA}}\sqrt{\frac{1}{T}}\right). \tag{9}$$

証明. $T' = T - T_0 = \Theta(T)$ と定義し，これは DA から見て，各財の評価値が $\{\hat{v}_{i,j(t)}\}_i$ として T' ラウンド行われるオンライン学習である．

補題 1 より:

$$\mathbb{E}\left[(u_i^{*,\mathrm{DA}} - \bar{u}_i^{\mathrm{DA}})^2\right] \le \frac{C^{\mathrm{DA}} \log T'}{T'},$$

マルコフの不等式より:

$$\Pr[|u_i^{*,\mathrm{DA}} - \bar{u}_i^{\mathrm{DA}}| \ge \epsilon'] \le \frac{1}{\epsilon'^2}\frac{C^{\mathrm{DA}} \log T'}{T'},$$

$\epsilon' = \epsilon u_i^{*,\mathrm{DA}}$ としたとき:

$$\Pr\left[\frac{|u_i^{*,\mathrm{DA}} - \bar{u}_i^{\mathrm{DA}}|}{u_i^{*,\mathrm{DA}}} \ge \epsilon\right] \le \frac{1}{(u_i^{*,\mathrm{DA}})^2\epsilon^2}\frac{C^{\mathrm{DA}} \log T'}{T'}, \tag{10}$$

式 10 の union bound をとると:

$$\Pr\left[\max_i \frac{|u_i^{*,\mathrm{DA}} - \bar{u}_i^{\mathrm{DA}}|}{u_i^{*,\mathrm{DA}}} \ge \epsilon\right] \le \frac{n}{\min_i (u_i^{*,\mathrm{DA}})^2\epsilon^2}\frac{C^{\mathrm{DA}} \log T'}{T'}. \tag{11}$$

ここで,

$$R := \max_i \frac{u_i^{*,\mathrm{DA}} - \bar{u}_i^{\mathrm{DA}}}{u_i^{*,\mathrm{DA}}} \le 1,$$

と定義したとき，全ての i において $\bar{u}_i^{\mathrm{DA}} \ge (1 - R)u_i^{*,\mathrm{DA}}$ となり:

$$\begin{aligned}
&\prod_i (u_i^{*,\mathrm{DA}})^{B_i} - \prod_i (\bar{u}_i^{\mathrm{DA}})^{B_i} \\
&\le \prod_i (u_i^{*,\mathrm{DA}})^{B_i} - \prod_i ((1 - R)u_i^{*,\mathrm{DA}})^{B_i} \\
&\le R\prod_i (u_i^{*,\mathrm{DA}})^{B_i}. \quad (\text{by } \prod_i (x)^{B_i} = x)
\end{aligned}$$

以上より:

$$\begin{aligned}
&\mathbb{E}\left[\prod_i (u_i^{*,\mathrm{DA}})^{B_i} - \prod_i (\bar{u}_i^{\mathrm{DA}})^{B_i}\right] \\
&\le \mathbb{E}\left[R\prod_i (u_i^{*,\mathrm{DA}})^{B_i}\right] \\
&\le \int_{1/\sqrt{T'}}^1 \prod_i (u_i^{*,\mathrm{DA}})^{B_i} \Pr[R \ge x]dx + \sqrt{\frac{1}{T'}}\prod_i (u_i^{*,\mathrm{DA}})^{B_i} \\
&= \int_{1/\sqrt{T'}}^1 \prod_i (u_i^{*,\mathrm{DA}})^{B_i} \Pr[R \ge x]dx + O\left(\sqrt{\frac{1}{T'}}\right) \\
&\le \int_{1/\sqrt{T'}}^1 \prod_i (u_i^{*,\mathrm{DA}})^{B_i} \frac{n}{\min_i (u_i^{*,\mathrm{DA}})^2 x^2}\frac{C^{\mathrm{DA}} \log T'}{T'}dx + O\left(\sqrt{\frac{1}{T'}}\right) \\
&= (u_i^{*,\mathrm{DA}})^{B_i}\frac{n}{\min_i (u_i^{*,\mathrm{DA}})^2}\frac{C^{\mathrm{DA}} \log T'}{T'}[-1/x]_{1/\sqrt{T'}}^1 + O\left(\sqrt{\frac{1}{T'}}\right) \\
&\le (u_i^{*,\mathrm{DA}})^{B_i}\frac{n}{\min_i (u_i^{*,\mathrm{DA}})^2}\frac{C^{\mathrm{DA}} \log T'}{\sqrt{T'}} + O\left(\sqrt{\frac{1}{T'}}\right) \\
&= \tilde{O}\left(nC^{\mathrm{DA}}\sqrt{\frac{1}{T'}}\right).
\end{aligned}$$

□

5.2 DA-EtC におけるリグレットの上界

本節では，提案手法の DA-EtC の上界に関する定理 1 と，その証明を示す．

定理 1. (DA-EtC におけるリグレットの上界)
全ての i, j において，$B_i = 1/n$, $s_j = 1/m$ とし，$l' \le v_{i,j} \le h'$ となるような $l', h' > 0$ が存在していると仮定する．その

時，$\sqrt{\frac{8\max(1,\sigma^2)nm\log(nmT)}{T_0}} \leq \frac{\min(1,l')}{2}$ が成立するような大きな T_0 において，*DA-EtC* におけるリグレットの上界は以下のように抑えられる:

$$\mathbb{E}[\mathrm{Regret}(T)] = \tilde{O}\left(T_0 + \sqrt{\frac{nm}{T_0}}T + nC^{\mathrm{DA}}\sqrt{T}\right).$$

また，$T_0 = T^{2/3}(nm)^{1/3}$ とした場合のリグレットの上界は $\mathbb{E}[\mathrm{Regret}(T)] = \tilde{O}(T^{2/3}(nm)^{1/3})$ となる．

証明. EtC では最初の T_0 ラウンドまではランダムに財を割り当てており，その時に得られる推定値 $\hat{v}_{i,j} = \hat{v}_{i,j}(T_0+1)$ は $N_{i,j}(T_0+1) \approx T_0/(nm)$ の標本に基づいて行われる．したがって．財を割り当てた回数 $N_{i,j}(T_0+1)$ 自身も確率変数である．そこで，2 つの事象 $\mathcal{A}, \mathcal{B}$ を以下のように定義する:

$$\mathcal{A} = \bigcap_{i,j}\left\{N_{i,j}(T_0+1) \geq \frac{T_0}{2nm}, \text{ and}\right\} \tag{12}$$

$$\mathcal{B} = \bigcap_{i,j}\left\{|\hat{v}_{i,j}(T_0+1) - v_{i,j}| \leq \sqrt{\frac{8\sigma^2 nm\log(nmT)}{T_0}}\right\}. \tag{13}$$

$N_{i,j}(T_0+1)$ は平均が $1/(nm)$ であるバイナリ変数を T_0 回足し合わせたものとしてみなすことができるため，乗法形式の Chernoff の不等式より，

$$N_{i,j}(T_0+1) < \frac{T_0}{2nm}$$

が最大で $\exp(-T_0/(8nm)) < 1/(nmT)$ の確率で成立する．また，事象 $\mathcal{A}$ は $i \in N, j \in M$ に関するブールの不等式を考慮することにより，少なくとも $1 - 1/T$ の確率で発生する．

次に，事象 $\mathcal{B}$ が $\mathcal{A}$ により高確率で発生することを示す．平均 $v_{i,j}$ と分散 σ^2 を持つ n' 個の独立同一分布からなる標本が与えられたとき，少なくとも $1 - 1/(nmT)^2$ の確率で

$$|\hat{v}_{i,j}(T_0+1) - v_{i,j}| \leq \sqrt{\frac{4\sigma^2\log(nmT)}{n'}}$$

が成立する．そして，$n' = N_{i,j}(T_0+1) \in [T]$ によるブールの不等式より

$$|\hat{v}_{i,j}(T_0+1) - v_{i,j}| \leq \sqrt{\frac{8\sigma^2 nm\log(nmT)}{T_0}} \,\middle|\, N_{i,j}(T_0+1) \geq \frac{T_0}{2nm}$$

となり，$\mathcal{A}$ が与えられたとき，$i \in N, j \in M$ に関するブールの不等式より，少なくとも $1 - 1/(nmT)$ の確率で $\mathcal{B}$ が成立する．これにより，$\mathcal{B}$ が成立しない確率は $O(1/T)$ であり，このことによるリグレットは高々 $O(T) \times O(1/T) = O(1)$ となり，無視できると判断できる．

次に，T_0 終了時点で事象 $\mathcal{B}$ が成立している時のリグレットの境界を以下のように変形する:

$$\begin{aligned}
&\frac{\mathrm{Regret}(T)}{T} \\
&= \prod_i (u_i^*)^{B_i} - \frac{1}{T}\prod_{i\in N} U_i(T)^{B_i} \\
&\leq \prod_i (u_i^*)^{B_i} - \frac{1}{T}\prod_{i\in N}((T-T_0)\bar{u}_i^{\mathrm{EtC}})^{B_i} \\
&= \prod_i (u_i^*)^{B_i} - \frac{T-T_0}{T}\prod_{i\in N}(\bar{u}_i^{\mathrm{EtC}})^{B_i} \qquad (\text{by } \sum_i B_i = 1) \\
&= \frac{T_0}{T}\prod_i (u_i^*)^{B_i} + \frac{T-T_0}{T}\left(\prod_i (u_i^*)^{B_i} - \prod_i (\bar{u}_i^{\mathrm{EtC}})^{B_i}\right) \\
&\leq \frac{T_0}{T}\prod_i (u_i^*)^{B_i} + \left|\prod_i (u_i^*)^{B_i} - \prod_i (\bar{u}_i^{\mathrm{EtC}})^{B_i}\right|. \qquad (14)
\end{aligned}$$

ここで，第 2 項を以下のように変形する:

$$\begin{aligned}
&\prod_i (u_i^*)^{B_i} - \prod_i (\bar{u}_i^{\mathrm{EtC}})^{B_i} \\
&= \left(\prod_i (u_i^*)^{B_i} - \prod_i (u_i^{*,\mathrm{DA}})^{B_i}\right) + \left(\prod_i (u_i^{*,\mathrm{DA}})^{B_i} - \prod_i (\bar{u}_i^{\mathrm{DA}})^{B_i}\right) \\
&+ \left(\prod_i (\bar{u}_i^{\mathrm{DA}})^{B_i} - \prod_i (\bar{u}_i^{\mathrm{EtC}})^{B_i}\right).
\end{aligned}$$

$u_i^{*,\mathrm{DA}}$ の定義より，任意の $\{x_{i,j}\}$ において $\prod_i (u_i^{*,\mathrm{DA}})^{B_i} \geq \prod_i \left(\sum_{j\in M} s_j \hat{v}_{i,j}(T_0+1)x_{i,j}\right)^{B_i}$ が成立する．したがって:

$$\prod_i (u_i^*)^{B_i} - \prod_i (u_i^{*,\mathrm{DA}})^{B_i} \leq \prod_i (u_i^*)^{B_i} - \prod_i (\hat{u}_i^{*,\mathrm{DA}})^{B_i}.$$

ここで，$\hat{u}_i^{*,\mathrm{DA}} = \sum_{j\in M} s_j \hat{v}_{i,j}(T_0+1)x_{i,j}^*$ と $\{x_{i,j}^*\}$ は真の評価値 $\{v_{i,j}\}$ による式 1 で表される凸計画問題の最適解とする．一方で，事象 $\mathcal{B}$ のもとで，

$$\begin{aligned}
|u_i^* - \hat{u}_i^{*,\mathrm{DA}}| &= \left|\sum_{j\in M} s_j x_{i,j}^*(v_{i,j} - \hat{v}_{i,j}(T_0+1))\right| \\
&\leq \sum_{j\in M} s_j x_{i,j}^* |v_{i,j} - \hat{v}_{i,j}(T_0+1)| \leq \tilde{O}\left(\sqrt{\frac{nm}{T_0}}\right)
\end{aligned}$$

これらの不等式と，補題 2 を組み合わせることにより，$u_i^* = \Theta(1)$ の仮定の下で，

$$\prod_i (u_i^*)^{B_i} - \prod_i (u_i^{*,\mathrm{DA}})^{B_i} \leq \tilde{O}\left(\sqrt{\frac{nm}{T_0}}\right). \tag{15}$$

さらに，事象 $\mathcal{B}$ と仮定より $l'/2 \leq \hat{v}_{i,j}(T_0+1) \leq h' + l'/2 \leq 2h'$ がいえるため，補題 3 に $l = l'/2, h = 2h'$ を適用して式 16 を得る．

$$\mathbb{E}\left[\prod_i (u_i^{*,\mathrm{DA}})^{B_i} - \prod (\bar{u}_i^{\mathrm{DA}})^{B_i}\right] \leq \tilde{O}\left(nC^{\mathrm{DA}}\sqrt{\frac{1}{T}}\right). \tag{16}$$

DA が各プレイヤに少なくとも $\Omega(T/(\sum_i (4h'/l')^2)) = \Omega((T/n)\times 1) = \Omega(T/n)$ 個の財を与える場合は，プレイヤ i_2

がプレイヤ i_1 よりも $(2h'/(l'/2))^2$ 倍多く財を受け取るときである．このとき，β_{i_2}/β_{i_1} の値は少なくとも $(2h'/(l'/2))$ となり，次のラウンドでどんな種類の財が到着したとしてもプレイヤ i_1 に割り当てが優先される．したがって，$\bar{u}_i^{\mathrm{DA}} = \Theta(1)$ である．このとき T_0 ラウンドでの標本の集中不等式を適用すると，少なくとも $1 - 1/T$ の確率で $\bar{u}_i^{\mathrm{EtC}} = \Theta(1), |\bar{u}_i^{\mathrm{EtC}} - \bar{u}_i^{\mathrm{DA}}| = \tilde{O}(\sqrt{nm/T_0})$ が成立する．これと補題 2 を用いると，

$$(\prod_i (\bar{u}_i^{\mathrm{DA}})^{B_i}) - \prod_i (\bar{u}_i^{\mathrm{EtC}})^{B_i}) = \tilde{O}\left(\sqrt{\frac{nm}{T_0}}\right). \tag{17}$$

つまり，

$$\mathbb{E}\left[\frac{\mathrm{Regret}(T)}{T}\right] = \frac{T_0}{T} + \tilde{O}\left(\sqrt{\frac{nm}{T_0}}\right) + nC^{\mathrm{DA}}\sqrt{\frac{1}{T}}. \tag{18}$$

□

5.3 リグレットの下界

定理 2. (Regret lower bound) 任意のアルゴリズムの期待リグレットの下界が $\mathbb{E}[\mathrm{Regret}(T)] = \Omega(\sqrt{mT})$ で抑えられるモデルが存在する．

証明. まず，財が到着する確率が一様に分布しているモデル ($s_j = 1/m$ for all $j \in [M]$, $B_i = 1/n$ for all $i \in [N]$) を考える．また，評価の行列 $\{v_{i,j}\}_{i,j} \in \mathbb{R}^{n\times m}$ をモデルとして扱う．ここではモデル v : $\forall_{i,j}\, v_{i,j} \in \{1/2, 1/2(1+\sqrt{m/T})\}^{nm}$ を考える．また，$v^{(0)} : \forall_{i,j}\, v_{i,j}^{(0)} = 1/2$ を基本モデルと呼ぶ．証明するうえで，[5] の補題 19 を変形した以下の補題 4 を示す．

補題 4. (任意の事象に対する下界) *2* つのモデル $v^{(1)}, v^{(2)}$ を定義する. また，$\mathbb{E}_{v^{(1)}}, \mathbb{E}_{v^{(2)}}$ をそれぞれのモデルに対する期待値，$\mathbb{P}_{v^{(1)}}, \mathbb{P}_{v^{(2)}}$ をそれぞれのモデルに対する確率変数と定義する．このとき，任意の事象 $\mathcal{E}$ に対して以下の不等式が成立する．

$$\sum_{i\in[n], j\in[m]} \mathbb{E}_{v^{(1)}}\left[N_{i,j}(T+1)\right] d(v_{i,j}^{(1)}, v_{i,j}^{(2)}) \geq d(\mathbb{P}_{v^{(1)}}(\mathcal{E}), \mathbb{P}_{v^{(2)}}(\mathcal{E})), \tag{19}$$

$d(p,q) = p\log(p/q) + (1-p)\log((1-p)/(1-q))$ は *2* つのベルヌーイ分布間の *KL* ダイバージェンスである．

ここで，$j_i = \arg\min_{j'} \mathbb{E}_{v^{(0)}}[N_{i,j'}(T+1)]$ を基本モデルのもとでエージェント i が T の間に受け取る回数が最も少ない財の種類と定義する．これより，$\sum_i N_{i,j_i}(T+1) \leq T/m$ が成立する．各 $i \in N$ に対して $v_{i,j_i} = (1+\sqrt{m/T})/2$ で，$k \neq j_i$ に対して，$v_{i,k} = 1/2$ であるモデル $v^{(a)}$ を新たに考える．このとき，

$$\sum_{i\in[n], j\in[m]} \mathbb{E}_{v^{(0)}}\left[N_{i,j}(T+1)\right] d(v_{i,j}^{(0)}, v_{i,j}^{(a)}) \tag{20}$$

$$\leq (T/m) d(1/2, 1/2(1+\sqrt{m/T})) \tag{21}$$

$$\leq (T/m) \times O(m/T) \quad (\text{by } d(1/2, 1/2+\alpha) = O(\alpha^2)) \tag{22}$$

$$= O(1). \tag{23}$$

が成立する．

事象

$$\mathcal{E} = \left\{\sum_i N_{i,j_i}(T+1) \leq 2T/m\right\}$$

を考えたときに，定義により $\mathbb{P}_{v^{(0)}}[\mathcal{E}] \geq 1/2$ となる．したがって，補題 4 と式 23 により，

$$d(\mathbb{P}_{v^{(0)}}[\mathcal{E}], \mathbb{P}_{v^{(a)}}[\mathcal{E}]) = O(1),$$

$$\mathbb{P}_{v^{(a)}}[\mathcal{E}] = \Omega(1).$$

これにより，代替モデル下での事象 $\mathcal{E}$ においてリグレットは少なくとも $\Omega(\sqrt{mT})$ となる．

□

5.4 DA-UCB

本節では DA-UCB の上界を与えることが困難である理由について述べる．理由として，DA-EtC においてリグレットの上界の証明に用いられた補題 1 には i.i.d. の仮定が必要だからである．DA-UCB では DA に適用する評価値の推定値に UCB 値を用いているため，推定値が財の与えた回数に依存することで i.i.d. ではなくなる．したがって，補題 1 は UCB 値の用いる DA-Iter に適用することができないために，先の証明を DA-UCB に用いることができない．さらに，双対平均化法における収束の証明のほとんどが [11] によると i.i.d. を仮定する必要があるため，証明の代替案を考えること，つまり DA-UCB の上界の導出は今後の課題である．

6 実験

本節では，DA-EtC と DA-UCB の性能を 3 つのデータセットを用いて評価する．全てのデータセットにおいて，B_i と財 j が選ばれる確率は一様であると仮定，つまり $B_i = 1/n$, $s_j = 1/m$ とする．次に Algolithm 2 における β_i の定義域を $[B_i/(1+0.95), (1+0.95)]$ とし，ノイズ ϵ_t はベルヌーイ分布に従って決定されるとした．つまり $\epsilon_t = 1$ が発生する確率を真の評価値 $v_{i,j}$ とし，$\epsilon_t = 0$ が発生する確率を $1 - v_{i,j}$ とした．DA-EtC の試行ラウンド T_0 は $T^{2/3}(nm)^{1/3}$ とし，プレイヤ数 (n)，財の種類 (m)，総ラウンド数 (T) を使用するデータセットに合わせて調整した．また，本実験はランダム性があるために，異なる試行を 20 回行った際のリグレットの平均値を求めた．

6.1 比較手法

本節では提案手法のベンチマークとして Random, UCB, DA-Grdy を紹介する．Random は到着した財をランダムに割り当てる．UCB は，双対平均化法を適用せずに，UCB 値が最も高いプレイヤに財を割り当てる．なお，UCB 値の更新は Algorithm 4 の 4 行目にしたがう．UCB は結果としてプレイヤの効用の和を最大化するので，ナッシュ積で評価すると性能が低くなることが予想される．最後に DA-Grdy は，DA-EtC とは異なり，各ラウンドでの評価値の推定値 $\hat{v}_{i,j}(t)$ の更新を停止せずに，DA-Iter 関数を実行した後推定値 $\hat{v}_{i,j}(t)$ を更新する．

6.2 データセット

本節では評価実験に使用する 3 つのデータセット *Uniform*，*Jester*，*Household* を概説する．まず，Uniform は評価値 $\{v_{i,j}\} \in [0,1]^{n\times m}$ をランダムに決定し，$n = 10$，$m = 10$ とした．また，総ラウンド数を $T = 100000$ とした．Jester は推薦システムや協調フィルタリングを研究するために構築された実データである [3]．これには，100 個のジョークに対する約 25000 人の評価が含まれており，その中で 7200 人が全てのジョークを評価している．その 7200 人から 10 人を，100 個のジョークから 50 個をランダムに選択し，$n = 10$，$m = 50$ とした．元のデータにおいて，その評価値の範囲は $[-10, 10]$ であったため，$[0, 1]$ に正規化した．また，総ラウンド数を $T = 300000$ とした．最後に，Household はオンラインレビューサイトから抽出した実データである [6]．これは 2876 人の 50 の家庭用品に対する支払い意思額を収集したデータであり，その額が $[0, 100]$ の範囲になる値をとるため，$[0, 1]$ に正規化した．今回は 2876 人から 10 人と 50 人をランダムに，50 の家庭用品を全て選択して，2 つのデータセット ($n = 10$，$m = 50$) と ($n = 50$，$m = 50$) を用意した．また，総ラウンド数を $T = 300000$ と設定した．

6.3 結果

図 1 に，3 つのデータセットそれぞれについて，5 つのアルゴリズムのリグレットを示す．また，図 2 に，図 1 と同設定において結果を見やすくするために，Random と UCB を除外した結果を示す．残りのデータセットである $n = 50$，$m = 50$ の Household の実験結果はまとめて図 3 に示す．

どのデータセットでも Random と UCB は同じ傾向を示している．実際，これらは双対平均化法を用いていないため，リグレットが大きくなる．また，UCB は Random よりもリグレットが大きい．これは UCB がナッシュ積ではなくプレイヤの効用の和を最大化した結果，不平等に財を割り当ててしまい，一部のプレイヤの効用が極端に少なくなってしまったためだと考えられる．

短期的にみると DA-EtC と DA-UCB は DA-Grdy と比べて低いリグレットを達成しているとは言えないが，長期的には低いリグレットを達成する．これは DA-Grdy において，各ラウンドで DA-Iter に評価値の推定値 $\hat{v}_{i,j}(t)$ を与えているため，探索が不足することがあるためである．例えば，あるラウンド t において $\hat{v}_{i,j}(t)$ が 0 であるプレイヤがいたとする．このとき双対平均化法では，今後種類 j の財をプレイヤ i に割り当てる可能性が低くなり，ラウンド t を過ぎると $\hat{v}_{i,j}$ を更新しなくなる．このような過小評価が起こるため，DA-Grdy は誤った配分を続けるようになり，そのリグレットが線形的に増加すると考えられる．

7 おわりに

本研究ではオンラインかつ，真の評価値が分からない状況下での資源配分アルゴリズム DA-EtC および DA-UCB を提案した．これらは長期的に見た場合，オフライン最適解をベンチマークとしたリグレットが他の一般的な手法と比べて優れた性能を示すことを計算機実験で確認した．さらに，DA-EtC におけるリグレットの上界を $\tilde{O}(T^{2/3}(nm)^{1/3})$，手法に問わないリグレットの下界を $\Omega(\sqrt{mT})$ で抑えられることを示した．

今後の課題として，DA-UCB のリグレットの上界の導出と，効用が加法的でない場合や財が到着する確率分布 $\mathcal{S}$ が独立かつ同一の分布で与えられていないケースを分析することなどが挙げられる．

参考文献

[1] E. Eisenberg and D. Gale. Consensus of subjective probabilities: The pari-mutuel method. *The Annals of Mathematical Statistics*, 30(1):165–168, 1959.

[2] Y. Gao, A. Peysakhovich, and C. Kroer. Online market equilibrium with application to fair division. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan eds., *Advances in Neural Information Processing Systems*, Vol. 34, pp. 27305–27318, 2021.

[3] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4:133–151, 2001.

[4] K. Jain and V. V. Vazirani. Eisenberg–gale markets: Algorithms and game-theoretic properties. *Games and Economic Behavior*, 70:84–106, 2010.

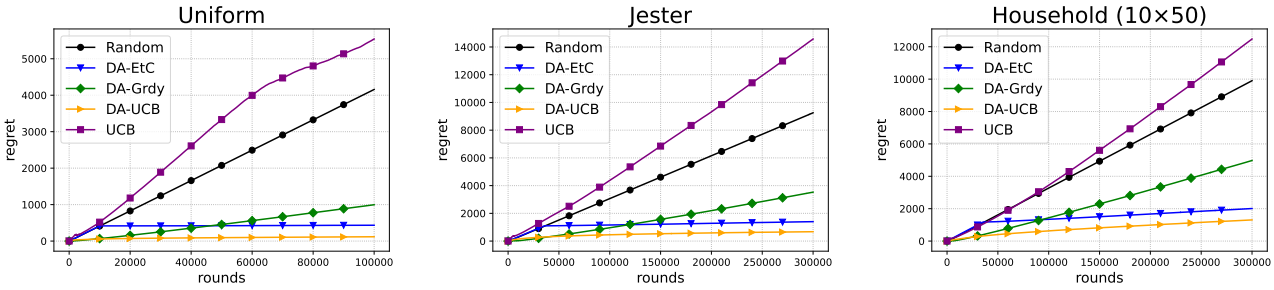
[5] E. Kaufmann, O. Cappé, and A. Garivier. On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17:1:1–1:42, 2016.

[6] C. Kroer, A. Peysakhovich, E. Sodomka, and N. E. S. Moses. Computing large market equilibria using abstractions. *Operations Research*, 70(1):329–351, 2021.

[7] T. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

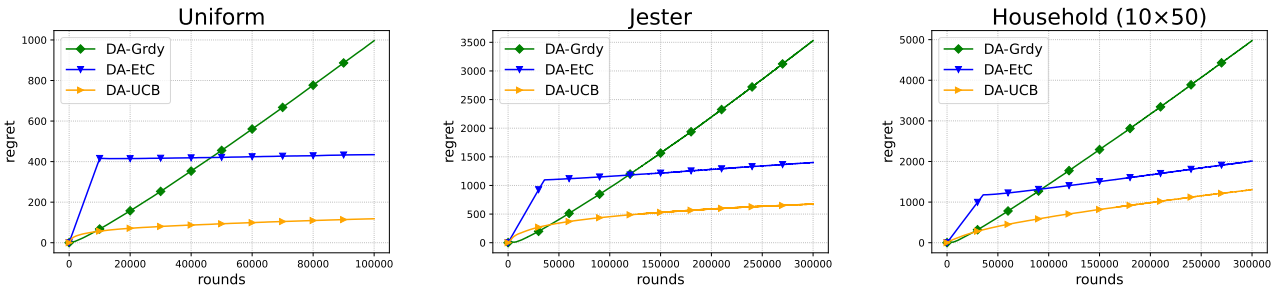
[8] J. F. Nash. The bargaining problem. *Econometrica*, 18(2):155–162, 1950.

[9] V. V. Vazirani. The notion of a rational convex program, and an algorithm for the arrow-debreu nash bargaining game. *Journal of the ACM*, 59(2), 2012.



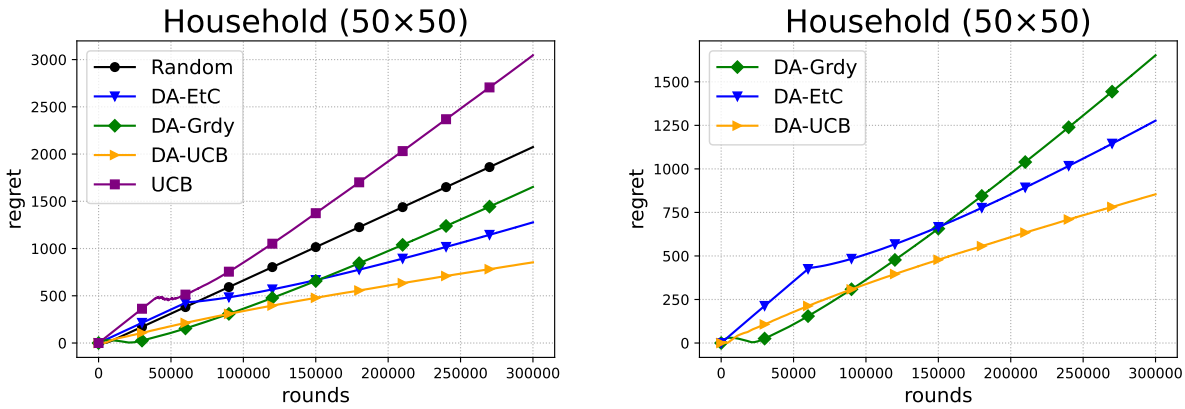
(a) Uniform (n=10 and m=10)　(b) Jester (n=10 and m=50)　(c) Household (n=10 and m=50)

図 1: 各手法におけるリグレットの推移



(a) Uniform (n=10 and m=10)　(b) Jester (n=10 and m=50)　(c) Household (n=10 and m=50)

図 2: 各手法におけるリグレットの推移 (Random と UCB を除外)



(a) All algorithms　(b) Three best algorithms

図 3: n=50 の Household におけるリグレットの推移

[10] N. K. Vishnoi. *Algorithms for Convex Optimization*. Cambridge University Press, 2021.

[11] L. Xiao. Dual averaging method for regularized stochastic learning and online optimization. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta eds., *Advances in Neural Information Processing Systems*, Vol. 22, 2009.