2T-01

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

1 はじめに

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，クラウドソーシングにおいてどんなタスクがやってくるかが，事前にはわからない状況を指す．例えば，避難所へ送る救援物資の配分を考えるとき，送ることができる物資に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ように複数種類が考えられる．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確実にはわからないうえ，被災者も物資が届いてからでないと，その必要度を事前に判断することはできない．つまり，被災者は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わからないとする．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，複数の避難所間での公平性も要求される．

一方で，クラウドソーシングでは，クライアントが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は実際にタスクを行ってもらわないとわからない．このとき，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．一般に，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したいが，特定の人にタスクが集中してしまう恐れが考えられる．そのため，タスク割当の効率性だけでなく，その公平性にも注意する必要がある．

資源配分における公平性には様々な概念があるが，本研究ではプレイヤ全員の効用の積であるナッシュ積 [3] を最大化することを目的とする．ここで，静的モデルとしての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のという視点から，凸計画問題に帰着できることが知られている [1]．そこで本研究は双対平均化法 [4] に着目して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 [1] があるが，本研究はこれをさらに一般化し，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．そこでバンディット問題 [2] に注目し，真の評価値をアルゴリズム内で予測しながら配分を決定するアルゴリズムを提案する．また，提案アルゴリズムの性能評価を静的なフィッシャー市場による凸計画問題の最適解を用いて行い，最適解におけるナッシュ積との差を評価する．

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 $N = \{1, 2, ..., n\}$，財の集合を $M = \{1, 2, ..., m\}$ とする．次に，プレイヤ i がある財 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 $v_{i,j}$ は事前にはわ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 $t = \{1, 2, ..., T\}$ に対して 1 つ財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（カテゴリカル分布）を $\mathcal{P} = \{p_1, p_2, ..., p_m\}$ と定義する．到着した財は一人のプレイヤ i_t にのみ割り当てされ，財を受け取ったプレイヤ i_t は価値のノイズあり評価 $u_i(t) = v_{i_t,j} + \epsilon_t$ を観測する．ここで，ϵ_t はある確率分布に従って発生するノイズである．最後に，プレイヤ i は 1 ラウンドあたり B_i の予算を持つとする．

次に，T ラウンド終了後のプレイヤ i のナッシュ積 $NSW(T)$ をそれぞれ以下のように定義する:

$$NSW(T) = \left(\prod_{i \in N} \left(\sum_{t:i_t=i} u_i(t) \right)^{B_i} \right)^{1/\|B\|_1} . \quad (1)$$

ここで財を分割可能と仮定し，真の評価値によるナッシュ積 $ENSW$ は，財 j をプレイヤ i に割り振る割合を $x_{i,j} \in [0, 1]$ とすると，以下の凸計画問題の目的関数の値として定義できる．

$$\begin{aligned} \text{maximize} \quad & \left(\prod_{i \in N} \left(\sum_{j \in M} p_j v_{i,j} x_{i,j} \right)^{B_i} \right)^{1/\|B\|_1} \quad (2) \\ \text{subject to} \quad & \forall j \in M : \sum_{i \in N} x_{i,j} = 1, \\ & \forall i \in N, \ \forall j \in M : x_{i,j} \geq 0. \end{aligned}$$

最後に，先に定義した NSW と $ENSW$ との差をリグレットとして定義し，この値の最小化を目指す:

$$Regret(T) = T \times ENSW - NSW(T). \quad (3)$$

3 双対平均化法 (Dual Averaging)

本節では、提案アルゴリズムの中心的アイデアである双対平均化法 [4] を概説する。まず, d 次元実数ベクトル $z \in Z$ ($Z \subseteq \mathbb{R}^d$)

* 電気通信大学情報理工学域
† ニューヨーク大学スターンスクールオブビジネス
‡ 株式会社サイバーエージェント
§ 電気通信大学大学院情報理工学研究科

に対して $f(\beta,z)$ を凸で微分可能な関数とする．次に強凸関数である Ψ を $\Psi := \{\beta \in \mathbb{R}^n : \Psi(\beta) < \infty\}$ とし、以下の最小化問題を考える．

$$\min_{\beta} \mathbb{E} f(\beta,z) + \Psi(\beta)$$

この最小化問題を解くプロセスは以下の手順で行われる．

1. 初期値を $\beta_1 \in \arg\min_{\beta} \Psi(\beta), \bar{g}^0 = 0, t = 1$ と設定
2. ラウンド t 時点で $f(\beta^t, z)$ を観察し，劣勾配 $g^t \in \partial f(\beta^t, z)$ を計算
3. 劣勾配の平均値である $\bar{g}^t$ を更新する
$$\bar{g}^t = \frac{t-1}{t}\bar{g}^{t-1} + \frac{1}{t}\bar{g}^t$$
4. β^{t+1} を計算する
$$\beta^{t+1} = \arg\min_{\beta}\{\langle \bar{g}^t, \beta\rangle + \Psi(\beta)\}$$
5. $t \leftarrow t+1$ として 2 から繰り返す

文献 [1] ではこの双対平均化法を式 2 に適用したオンラインアルゴリズムを提案しているが，プレイヤが各財に対してもつ真の評価値 $v_{i,j}$ が既知であると仮定している．次節では，真の評価値が未知と仮定したバンディット設定において，ϵ-貪欲法 [2] を用いた双対平均化法の拡張である DA-ϵ を提案する．

4 提案手法

本章ではオンラインアルゴリズムの提案手法である DA-ϵ を概説する．このアルゴリズムの設定では真の評価値 v が分からないため，ϵ-貪欲法によって評価値の予測値 $\hat{v}$ を求め，求めた予測値を基に双対平均化法を実行することで割り当てを行っている．

双対平均化法を実行するにあたり，評価の予測値を真の評価値に近づけるために ϵ-貪欲法によって以下の手順を行う．

1. 評価の予測値 $\hat{v}$ の初期値を 1 とする
2. ϵ の確率の時
   (a) 財 j をランダムに割り当てする
   (b) 割り当てた人 i_t はノイズあり評価値 $u_i(t)$ を観測し，予測値 $\hat{v}$ を更新
3. 1-ϵ の確率の時
   (a) 今までの予測値を用いて双対平均化法を実行
   (b) 双対平均化法によって割り当てられた人 i_t はノイズあり評価値 $u_i(t)$ を観測し，予測値 $\hat{v}$ を更新

以上の手順より，全ラウンド T の内 ϵT 回が予測値の計算のために使用され，$(1-\epsilon)T$ 回は予測値を基に双対平均化法を実行している．また，ϵT 回がランダムに割り当てすることで予測値を求めるためのデータ数の偏りを調整している．

次に，予測値 $\hat{v}$ を用いた双対平均化法の流れを概説する．まず, 初期値として $\beta_i^1 = 1+\delta_0\ (\delta_0 > 0), \bar{u}_i^t = 0, t = 1$ と定義する．ここで $\bar{u}_i^t$ は累積平均効用を表している．次に，各ラウンドで財 $j \in M$ が選ばれることで双対平均化法の $f(\beta^t, z)$ に対応する関数として $f(\beta^t, z) = \max_{i\in N} \beta_i^t \hat{v}_{i,j}$ を観測し，$i_t = \arg\max_i \beta_i^t \hat{v}_{i,j}$ としたときの劣勾配 $u^t = \hat{v}_{i_t,j}\mathbf{e}^{i_t} \in \partial f(\beta^t, j)$ を計算する．その後 u^t の平均値，すなわち累積平均効用 $\bar{u}^t$ を更新し，パラメータ β^{t+1} を更新する．ここまでの手順を以下に示す．

1. 与えられた財 j に従い，$i_t = \arg\max_i \beta_i^t \hat{v}_{i,j}$ を計算
2. 各プレイヤは効用 $u_i^t = \hat{v}_{i,j}\mathbb{I}\{i = i_t\}$ を得る
3. 各プレイヤの累積平均効用 $\bar{u}_i^t$ を更新
$$\bar{u}_i^t = \frac{1}{t}\sum u_i^\tau = \frac{t-1}{t}\bar{u}_i^{t-1} + \frac{1}{t}u_i^t$$
4. 各プレイヤの β_i^{t+1} を更新
$$\beta_i^{t+1} = \prod_{[B_i/(1+\delta_0),1+\delta_0]} (B_i/\bar{u}_i^t) := \min\{\max\{B_i/(1+\delta_0), B_i/\bar{u}_i^t\}, 1+\delta_0\}$$
5. $t \leftarrow t+1$ として繰り返す

β の値は累積平均効用が小さくなると大きくなるため，財が割り当てられない人に対して，次ラウンドの $\beta_i^{t+1}\hat{v}_{i,j}$ の値を大きくして割り当てされやすいように調整している．また，手順 4 の β の更新で β の取りうる範囲を $[B_i/(1+\delta_0), 1+\delta_0]$ としているが，これは真の評価値を正規化することで最適解での β の値が $[B_i/(1+\delta_0), 1+\delta_0]$ に抑えられるためである．正規化はすべての財が $1/m$ ずつ割り当てされたと仮定した場合の合計の効用が 1 になるようにしている．

5 おわりに

本研究ではオンライン環境において真の評価値が分からない状況下での資源配分を達成するオンラインアルゴリズムを提案した．今後このアルゴリズムの性能を確かめるために，式 2 の凸計画問題と DA-ϵ の実装をする．そして，各ラウンド終了時点での NSW と ENSW の誤差であるリグレットを計算し，リグレットの値が漸近的に収束することを確認する予定である．

参考文献

[1] Y. Gao, A. Peysakhovich, and C. Kroer. Online market equilibrium with application to fair division. *Advances in Neural Information Processing Systems*, 34:27305–27318, 2021.

[2] 本多淳也, 中村篤祥. バンディット問題の理論とアルゴリズム. 講談社, 2016.

[3] M. Kaneko and K. Nakamura. The Nash social welfare function. *Econometrica: Journal of the Econometric Society*, pp. 423–435, 1979.

[4] L. Xiao. Dual averaging method for regularized stochastic learning and online optimization. *Advances in Neural Information Processing Systems*, 22, 2009.