F-025

シンクホーンアルゴリズムを用いた公平ランキング問題の高速求解

上原 祐輝 *† 池田 春之介 ‡ 西村 直樹 § 鮏川 矩義 ¶ 高野 祐一 ‖
筑波大学 † 筑波大学 ‡ 株式会社リクルート § 法政大学 ¶ 筑波大学 ‖

1 はじめに

オンラインフリーマーケットなどの双方向市場では，商品購入者の目的に合わせた商品のランキングを提供する推薦システムが取引を促進する重要な役割を果たしている．多くのランキング手法は，商品購入者の満足度を高めることに焦点を当てているが，双方向市場では，商品提供者の活動を促進するために，商品購入者の満足度と商品提供者が得る利益のバランスを取る必要がある．

各商品が得る影響度（クリック数や売上額など）の公平性を保証するため，文献 [1] では，ナッシュ社会的厚生を最大化するランキング手法が提案された．しかし，この手法は大規模な制約付き非線形最適化問題を解く必要があり，実用的な規模の推薦システムに適用するのは困難である．

本研究では，影響度の公平ランキング問題に対する高速解法を提案する．提案手法では，影響度の公平ランキング問題を制約なし最適化問題に変換し，シンクホーンアルゴリズムを繰り返し実行する勾配上昇法を設計する．

2 影響度の公平ランキング問題

まず，商品購入者に対して個別化された商品のランキングを作成するための影響度の公平ランキング問題 [1] について説明する．

商品購入者 $u \in U$ に対して，商品 $i \in I$ が k 番目の位置 $(k \in [m])$ に表示される確率を x_{uik} とする．ただし，$[m] \coloneqq \{1, 2, \ldots, m\}$ とする．

購入者 $u \in U$ に対する確率的ランキングを二重確率行列 $\boldsymbol{X}_u \coloneqq (x_{uik})_{(i,k)\in I\times[m]} \in \mathbb{R}^{|I|\times m}$ として定式化すると，$\boldsymbol{X}_u$ は以下の制約を満たす．

$$\sum_{k\in[m]} x_{uik} = 1 \quad (i \in I), \tag{1}$$

$$\sum_{i\in I} x_{uik} = 1 \quad (k \in [m-1]), \tag{2}$$

$$x_{uik} \geq 0 \quad (i \in I,\ k \in [m]). \tag{3}$$

ここで，制約 (1)–(2) は商品と表示位置に対する確率制約を表す．また，商品の数が表示位置の数よりも多いため，m 番目の表示位置をダミーとする．

このとき，$\boldsymbol{X} \coloneqq (\boldsymbol{X}_u)_{u\in U} \in \mathbb{R}^{|U|\times|I|\times m}$ による商品への影響度は次のように定義される:

$$\mathrm{Imp}_i(\boldsymbol{X}) \coloneqq \sum_{u\in U}\sum_{k\in[m-1]} r(u,i)\cdot e(k)\cdot x_{uik} \quad (i \in I)$$

ここで，$r(u,i)$ は商品購入者 u と商品 i の関連度を表し，$e(k)$ は k 番目の表示位置の閲覧確率を表す．公平性を考慮した目的関数は，すべての商品に対する影響度の積によって次のように定義される:

$$F(\boldsymbol{X}) \coloneqq \log\prod_{i\in I}\mathrm{Imp}_i(\boldsymbol{X}) = \sum_{i\in I}\log\mathrm{Imp}_i(\boldsymbol{X}).$$

最終的に，公平ランキング問題は次のように定式化される:

$$\begin{array}{ll} \underset{\boldsymbol{X}\in\mathbb{R}^{|U|\times|I|\times m}}{\text{maximize}} & F(\boldsymbol{X}) \\ \text{subject to} & \text{式 (1)–(3)} \quad (u \in U). \end{array} \tag{4}$$

Fast solution to the fair ranking problem using the Sinkhorn algorithm

† Yuki Uehara, University of Tsukuba
‡ Shunnosuke Ikeda, University of Tsukuba
§ Naoki Nishimura, Recruit Holdings Co., Ltd.
¶ Noriyoshi Sukegawa, Hosei University
‖ Yuichi Takano, University of Tsukuba

3 シンクホーンアルゴリズムを用いた勾配上昇法

問題 (4) を解くための提案手法を説明するため，$u \in U$ に対するエントロピー正則化項付き最適輸送問題を説明する [2].

$$\begin{aligned} \underset{\boldsymbol{X}_u \in \mathbb{R}^{|I| \times m}}{\text{minimize}} \quad & \sum_{i \in I} \sum_{k \in [m]} c_{uik} x_{uik} + \varepsilon H(\boldsymbol{X}_u) \\ \text{subject to} \quad & \text{式 (1)–(3)}. \end{aligned} \tag{5}$$

ここで，$\boldsymbol{C}_u := (c_{uik})_{(i,k) \in I \times [m]} \in \mathbb{R}^{|I| \times m}$ は $u \in U$ に対する輸送費用行列であり，$\varepsilon > 0$ は正則化パラメータ，$H(\boldsymbol{X}_u) := \sum_{i \in I} \sum_{k \in [m]} x_{uik}(\log x_{uik} - 1)$ はエントロピーである．この最適輸送問題は，シンクホーンアルゴリズムを用いて効率的に解くことができる [2]．シンクホーンアルゴリズムによって算出した問題 (5) の最適解を $\boldsymbol{X}_u^\star(\boldsymbol{C}_u)$ とする．このとき，勾配 $\nabla \boldsymbol{X}_u^\star(\boldsymbol{C}_u)$ は連鎖律を用いて計算することができる．

次に，公平ランキング問題 (4) に対して最適となる解 $\boldsymbol{X}^\star(\boldsymbol{C}) := (\boldsymbol{X}_u^\star(\boldsymbol{C}_u))_{u \in U}$ が得られるような輸送費用 $\boldsymbol{C} := (\boldsymbol{C}_u)_{u \in U}$ を求める問題を考え，制約付き最適化問題 (4) を以下の制約なし最適化問題に変換する．

$$\underset{\boldsymbol{C} \in \mathbb{R}^{|U| \times |I| \times m}}{\text{minimize}} \quad F(\boldsymbol{X}^\star(\boldsymbol{C})). \tag{6}$$

ここで，問題 (4) と問題 (6) の実行可能領域は，次の意味で等価である:

定理 1. 式 (1)–(3) を満たす任意の $\boldsymbol{X}_u$ に対して，$\boldsymbol{X}_u = \boldsymbol{X}_u^\star(\boldsymbol{C}_u)$ となるような $\boldsymbol{C}_u$ が存在する．

証明. 問題 (5) の目的関数が凸であるため，その最適性条件は $(i,k) \in I \times [m]$ に対して $c_{uik} + \varepsilon \log x_{uik} = 0$ である．したがって，任意の $\boldsymbol{X}_u$ に対して，すべての $(i,k) \in I \times [m]$ で $c_{uik} = -\varepsilon \log x_{uik}$ とすることで，$\boldsymbol{X}_u$ が問題 (5) の最適解（すなわち $\boldsymbol{X}_u = \boldsymbol{X}_u^\star(\boldsymbol{C}_u)$）となる． □

問題 (6) を解くための勾配上昇法は以下のように記述できる．

Algorithm 1 問題 (6) に対する勾配上昇法

Input: 収束判定の閾値 $t > 0$.
Initialize: $\boldsymbol{C} := (\boldsymbol{C}_u)_{u \in U} \in \mathbb{R}^{|U| \times |I| \times m}$,
$\boldsymbol{X} := (\boldsymbol{X}_u)_{u \in U} \in \mathbb{R}^{|U| \times |I| \times m}$.
while $\|\nabla F(\boldsymbol{X})\| > t$ **do**
  for $u \in U$ **do in parallel**
    シンクホーンアルゴリズムを用いて $\boldsymbol{X}_u = \boldsymbol{X}_u^\star(\boldsymbol{C}_u)$ を計算する．
  end
  勾配上昇方向 $\nabla_{\boldsymbol{C}} F(\boldsymbol{X}^\star(\boldsymbol{C}))$ に $\boldsymbol{C}$ を更新する．
end
Output: 二重確率行列 $\boldsymbol{X}_u$ $(u \in U)$.

4 数値実験

提案手法の性能を人工データおよび実データに対する評価指標と，商品数および購入者数を変化させた際の実行時間で検証する．実験に使用したデータおよび結果の詳細は当日報告する．

参考文献

[1] Saito, Y., Joachims, T.:Fair ranking as fair division:Impact-based individual fairness in ranking. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 1514-1524 (2022)

[2] Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. Advances in Neural Information Processing Systems, 26 (2013)