Computer Security Symposium 2021
26 - 29 October 2021

マッチングにおける個人主義的公平性とその侵害検知

中村 徹[1,a)]　新田 修也[1]　磯原 隆将[1]

概要：本稿では，安定結婚問題における男女マッチングの問題設定をベースとして，これまでに考えられてきた全体最適的な意味合いでの公平性や男女間の公平性とは異なる観点で，個人主義的公平性を新たに定義する．本稿で考える個人主義的公平とは，能力や評価がより高い者はより希望通りになるという考え方である．次に，全体最適なマッチングや安定なマッチングが、個人主義的な不公平を生じる事例を示す．さらに，提示されているものとは異なる方針によってマッチングを実施されることを検知する手法について提案する．

キーワード：マッチング，公平性，方針違反検知

Individual Fairness on Matching and Detection of Its Violation

Toru Nakamura[1,a)]　Naoya Nitta[1]　Takamasa Isohara[1]

Abstract: This paper proposes a new definition on fairness for matching, called individual fairness, based on the existing setting of stable marriage problem. The new definition of individual fairness means that a wish of a person who has better capability or evaluation takes a priority than that of worse person. This definition is different from the existing definitions of fairness such as ”overall optimization” or ”gender equality”. Next, this paper provides examples that overall optimized matching and stable matching cause individual unfairness. This paper also proposes a method to detect policy violation for matching.

Keywords: Matching, Fairness, Policy Violation Detection

1. はじめに

意思決定を支援するツールは，レコメンドやマッチングなどの様々な領域のサービスとして普及している．一方で，自動的に不公平であったり不利益をもたらす意思決定が行われているかもしれないという不信も生じている．特に近年は AI の公平性や透明性を含む AI のトラストに関する議論が盛んである．

本稿では特に，ジョブマッチングやオンラインデートのようなマッチングサービスに注目して考える．マッチングの参加者は，ジョブマッチングの場合は求職者と求人を出す会社，オンラインデートの場合は男性参加者と女性参加者，となり，マッチングサービス提供者によりそれぞれの参加者に対しマッチング結果やレコメンドが提供される．ここで，それぞれの参加者，及びマッチングサービス提供者の利害関係は必ずしも一致しない [1]．例えばジョブマッチングについて考えたとき，能力の高い求職者は能力に応じて条件の良い職が得られるように期待しており，条件の良い求人を出している会社は条件の高さに応じて能力の高い求職者が得られると期待しているかもしれない．しかしながら，マッチングサービス提供者が，全体としての満足度の最適化やマッチング成立率などを重視した場合，必ずしも参加者の期待に沿った結果にはならない．さらに悪質な場合には，あえてマッチングを成立させずに参加者をサービスに滞留させることが，マッチングサービス提供者にとって最大の利益となる方針として採用されることも考えられる．

どのような方針でマッチングを決めることが公平である

1　株式会社 KDDI 総合研究所
KDDI Research, Inc.
a)　tr-nakamura@kddi-research.jp

のかは一概に決めることはできない．透明性の観点からは，どのような方針でマッチングを行っているのかを参加者に提示することが望ましい．マッチングに AI を活用している場合においては，説明可能な AI 技術の適用により参加者に結果の根拠を示すことで透明性が確保できると期待する向きもあるが，その説明の妥当性の担保をどのように行うかは課題となる [2]．また，実際に適用されたモデルやモデルの元となる学習データにアクセスできない参加者が，信頼していない事業者から与えられた根拠が正当なものであると確信できるわけではない．

本稿では，安定結婚問題における男女マッチングの問題設定 [3] をベースとして，公平性及び透明性に関する議論を行う．この問題設定では，各参加者がマッチング相手について予め好みの順番（選好順序）を決めている前提が置かれており，近年よく議論されている機械学習における公平性 [4], [5], [6] とは異なる設定であることに注意されたい．我々は，これまでに考えられてきた全体最適的な意味合いでの公平性 [7] や男女間の公平性 [8] とは異なる観点で，個人主義的公平性を新たに定義する．本稿で考える個人主義的公平とは，能力や評価がより高い者はより希望通りになるという考え方である．次に，全体最適なマッチングや安定なマッチングが、個人主義的な不公平を生じる事例を示す．また，個人主義的公平性の定義を緩和する拡張により，公平性を定量的に評価可能にする．さらに，例えば個人主義的に公平なマッチングを実施していると提示しておきながら，実際には全体最適なマッチングや安定なマッチングなど，異なる方針によってマッチングを実施されることを検知する手法について提案する．

本稿は以下の構成からなる．2 章で本稿で考えるマッチングのモデルについて説明する．3 章で本稿で新たに定義するマッチングにおける個人主義的公平性の概念と定義を与える．4 章で，本章では，具体的な選好順序の例を挙げ，個人主義的な公平マッチング例，及び全体最適や安定なマッチングではあるが個人主義的公平ではないマッチング例を示す．また，個人主義的公平性の定義を緩和する拡張を示す．5 章で，異なる方針によってマッチングを実施されることを検知する提案手法を示す．6 章で関連研究の紹介を行い，7 章でまとめる．

2. 男女マッチング

本章では，本稿で考えるマッチングのモデルについて説明する．本稿では，安定結婚問題における男女マッチングの問題設定 [3] をベースとしたモデルを用いる．

2.1 モデル

同サイズ n の男性集合 M と女性集合 W があり，それらのマッチングを $\mathcal{M}$ とする．ここでは簡単のため，必ず 1 対 1 のペアを n 組作る場合のみ考える．各人それぞれ，もう一方の性別のすべてのメンバについての全順序である**選好順序**を持つ．マッチング $\mathcal{M}$ において男性 m と女性 w がペアであるとき，$\mathcal{M}(m)=w, \mathcal{M}(w)=m$ とする．男性 m と女性 w について、$P_m(w)$ を，m の選好順序における w の順位とする． 同様に $P_w(m)$ を，w の選好順序における m の順位とする．ある人物 p が異性 b よりも異性 a を好むとき，$a \succ_p b$ と表す．

2.2 全体最適なマッチング

定義 1. 以下のコスト関数 C を最小にするマッチング $\mathcal{M}$ を，**全体最適なマッチング**と呼ぶ．

$$C(\mathcal{M}) = \sum_{(m,w)\in\mathcal{M}} (P_m(w) + P_w(m))$$

これは重み付き最小完全二部マッチング問題となる．

2.3 安定なマッチング

以下の安定性を持つマッチングを求める問題は，安定結婚問題として知られる [3]．安定性とは，直感的には浮気が発生する可能性がない性質である．

定義 2. 以下の条件を満たすペア (m', w') を**ブロッキングペア**と呼ぶ．

(1) $(m', w') \in \mathcal{M}$ かつ $m' \in M, w' \in W$

(2) $w' \succ_{m'} \mathcal{M}(m')$

(3) $m' \succ_{w'} \mathcal{M}(w')$

定義 3. ブロッキングペアが存在しないマッチング $\mathcal{M}$ を，**安定なマッチング**と呼ぶ．

任意の選好順序の条件において，少なくとも 1 つの安定なマッチングが存在することが知られており，例えば Gale-Shapley アルゴリズムを用いることで $\mathcal{O}(n^2)$ で安定なマッチングを見つけることが可能である [3]．Gale-Shapley アルゴリズムは，男性集合に対し最適であり，女性集合に対し最低となるマッチングを見つけるアルゴリズム（女性に最適，男性に最低にすることは可能）であることが知られており，公平性の概念に拡張した変種の問題が提案されている [7], [8]．例えば以下のコスト関数 D を最小にする安定なマッチングを見つける問題を**男女平等な安定マッチング問題**と呼ぶ．

$$D(\mathcal{M}) = |\sum_{(m,w)\in\mathcal{M}} P_m(w) - \sum_{(m,w)\in\mathcal{M}} P_w(m)|$$

3. 男女マッチングにおける個人主義的公平

本章では，本稿で新たに定義するマッチングにおける個人主義的公平性の概念と定義を与える．

3.1 概念

本稿で考える個人主義的公平とは，能力や評価がより高い者はより希望通りになるという考え方である．男女マッ

チングにおいては，異性からより早い選好順序に指定されている傾向にある人物を，評価の高い人物とみなすこととする．本稿では，全ての異性からの選好順序の総和を持って，人物の定量的な評価を行い，同性他者との比較を行う．あるマッチングにおける任意の人物について，その人物よりも評価の低い同性他者とペアとなっている相手の順位が，その人物とペアとなっている相手の順位よりも常に下であるとき，このマッチングは個人主義的公平であるとする．

3.2 定義

$G(m) = \sum_{w_i \in W} P_{w_i}(m)$ を，m の**優秀度**とする．優秀度は小さいほど異性からの評価が高いことを表す．同様に，$G(w) = \sum_{m_i \in M} P_{m_i}(w)$ を，w の優秀度とする．$O(m)$ を，男性集合を優秀度の順にソートした場合の男性 m の順位とし、これを**優秀順位**と呼ぶ．同様に，$O(w)$ を，女性集合を優秀度の順にソートした女性 w の優秀順位とする．ある集合 H について，p よりも優秀順位の低い集合を $H_{<p}$ とする．

定義 4. 全ての $m \in M$ 及び $w \in W$ について，

(1) 全ての $m' \in M_{<O(m)}$ について

$$P_m(\mathcal{M}(m)) < P_{m'}(\mathcal{M}(m'))$$

(2) 全ての $w' \in W_{<O(w)}$ について

$$P_w(\mathcal{M}(w)) < P_{w'}(\mathcal{M}(w'))$$

が成り立つとき，このマッチング $\mathcal{M}$ を**個人主義的公平**であるとする．

4. 個人主義的公平に関する観察

本章では，具体的な選好順序の例を挙げ，個人主義的な公平マッチング例，及び全体最適や安定なマッチングではあるが個人主義的公平ではないマッチング例を示す．

4.1 マッチング例

ここで，マッチングについての例を示す．$n = 4$ とし，男性集合を $\{A, B, C, D\}$，女性集合を $\{a, b, c, d\}$ とする．それぞれの選好順序の例を表 1 に示す．ここで，各男性の各女性に対する順位を並べたもの，及びその逆を表にする．例えば，A 行について，男性 A の選好順序において a の順位は 3 である．同様に，b, c, d の順位はそれぞれ 1, 2, 4 である．ある女性に対する全ての男性の順位を総和したものが，その女性の優秀度となる．例えば，a 列について，男性 A の選好順序において a の順位は 3 である．同様に，B, C, D における a の順位はそれぞれ 3, 3, 4 である．このとき，a の優秀度は $3 + 3 + 3 + 4 = 13$ となる．以上について表 2 に示す．

この選好順序における全てのマッチングについて，不満足度にあたるコスト関数 C の値，全体最適であるか，個人主義的公平であるか，安定であるかについて表 3 に全列挙する．

表 1 例として用いる選好順序

男性の選好順序					女性の選好順序				
A:	b	c	a	d	a:	B	A	C	D
B:	d	c	a	b	b:	A	C	D	B
C:	d	b	a	c	c:	A	B	D	C
D:	d	c	b	a	d:	D	C	B	A

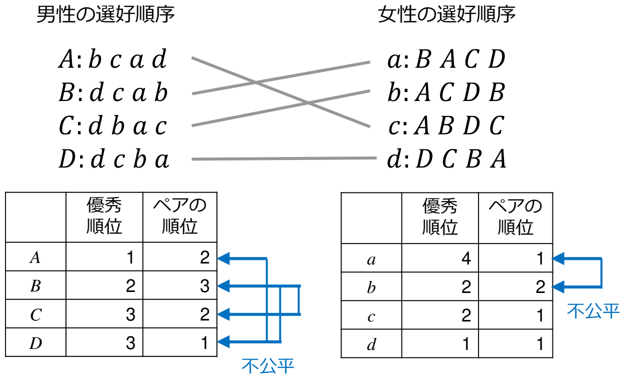

図 1 全体最適なマッチング

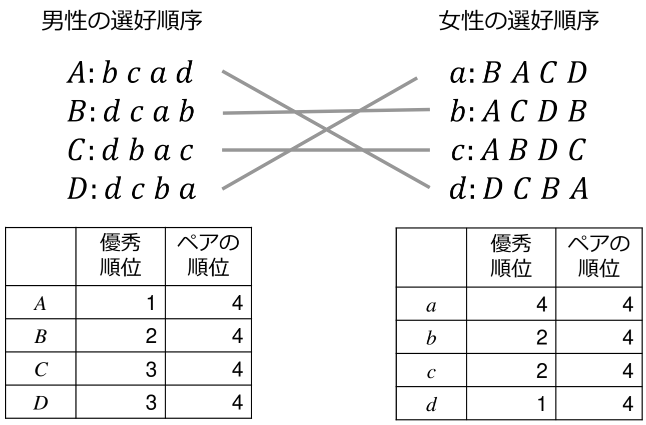

図 2 個人主義的公平なマッチング

上記の例において，全体最適なマッチング $\mathcal{M}_T$ は，$\{(A, c), (B, a), (C, b), (D, d)\}$ である．優秀順位及び $\mathcal{M}_T$ における各人のマッチング相手の順位を図 1 に示す．このとき，$C(\mathcal{M}_T) = 13$ である．しかしながら，このマッチング $\mathcal{M}_T$ は個人主義的公平ではない．なぜなら，例えば A の優秀順位は 1 であり D の優秀順位 3 よりも小さいにも関わらず，A のマッチング相手の順位は 2 であり D のマッチング相手の順位 1 よりも大きい．

次に個人主義的公平なマッチングの例を示す．上記の例において，個人主義的公平なマッチング $\mathcal{M}_F$ は，$\{(A, d), (B, b), (C, c), (D, a)\}$ のみである．優秀順位及び $\mathcal{M}_F$ における各人のマッチング相手の順位を図 2 に示す．このとき，$C(\mathcal{M}_F) = 32$ である (各人にとって最悪の組み合わせである)．

比較のため，安定なマッチングの例を示す．

表 2　優秀度と優秀順位

女性の優秀度と優秀順位					男性の優秀度と優秀順位				
	a	b	c	d		A	B	C	D
A	3	1	2	4	a	2	1	3	4
B	3	4	2	1	b	1	4	2	3
C	3	2	4	1	c	1	2	4	3
D	4	3	2	1	d	4	3	2	1
優秀度	13	10	10	7	優秀度	8	10	11	11
優秀順位	4	2	2	1	優秀順位	1	2	3	3

表 3　本例におけるマッチングの全列挙

マッチング $\mathcal{M}$	$C(\mathcal{M})$	全体最適	公平性	安定性
$(A,a),(B,b),(C,c),(D,d)$	23			
$(A,a),(B,b),(C,d),(D,c)$	21			
$(A,a),(B,c),(C,b),(D,d)$	15			
$(A,a),(B,c),(C,d),(D,b)$	18			
$(A,a),(B,d),(C,b),(D,c)$	18			
$(A,a),(B,d),(C,c),(D,b)$	23			
$(A,b),(B,a),(C,c),(D,d)$	16			
$(A,b),(B,a),(C,d),(D,c)$	14			
$(A,b),(B,c),(C,a),(D,d)$	14			✓
$(A,b),(B,c),(C,d),(D,a)$	17			
$(A,b),(B,d),(C,a),(D,c)$	17			
$(A,b),(B,d),(C,c),(D,a)$	22			
$(A,c),(B,a),(C,b),(D,d)$	13	✓		
$(A,c),(B,a),(C,d),(D,b)$	16			
$(A,c),(B,b),(C,a),(D,d)$	19			
$(A,c),(B,b),(C,d),(D,a)$	22			
$(A,c),(B,d),(C,a),(D,b)$	19			
$(A,c),(B,d),(C,b),(D,a)$	19			
$(A,d),(B,a),(C,b),(D,c)$	21			
$(A,d),(B,a),(C,c),(D,b)$	26			
$(A,d),(B,b),(C,a),(D,c)$	27			
$(A,d),(B,b),(C,c),(D,a)$	32		✓	
$(A,d),(B,c),(C,a),(D,b)$	24			
$(A,d),(B,c),(C,b),(D,a)$	24			

上記の例において，安定なマッチング $\mathcal{M}_S$ は，$\{(A,b),(B,c),(C,a),(D,d)\}$ のみである．優秀順位及び $\mathcal{M}_S$ における各人のマッチング相手の順位を図 3 に示す．このとき，$C(M) = 14$ である．

4.2　考察

実は任意の選好順序の条件において、必ずしも個人主義的公平なマッチングが存在するとは限らない．例えば表 4 で示す選好順序の条件においては、個人主義的公平なマッチングは存在しない．また，前節で示したように個人主義的な公平性を満足するものの，全体最適性を大きく損なうマッチングしか存在しない場合もある．そこで，下位 k 番目以内の人が自分よりも好み順位が上の人とマッチングするのを許容する，またはある 2 名がマッチングした相手の好み順位の差が ℓ 以内であれば差がないとみなし許容する，

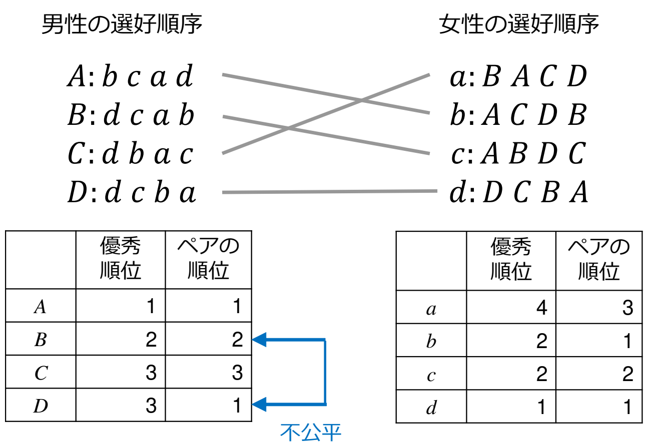

図 3　安定なマッチング

といった条件を緩和した定義が有用であると考えられる．

定義 5. 全ての $m \in M$ 及び $w \in W$ について，

表 4　個人主義的公平なマッチングが存在しない選好順序の例

男性の選好順序					女性の選好順序				
A:	d	b	a	c	a:	B	D	A	C
B:	d	b	c	a	b:	C	B	A	D
C:	c	d	a	b	c:	D	B	A	C
D:	c	a	d	b	d:	D	C	B	A

(1) 全ての $m' \in M_{<O(m)+k}$ について

$$P_m(\mathcal{M}(m)) < P_{m'}(\mathcal{M}(m')) + \ell$$

(2) 全ての $w' \in W_{<O(w)+k}$ について

$$P_w(\mathcal{M}(w)) < P_{w'}(\mathcal{M}(w')) + \ell$$

が成り立つとき，このマッチング $\mathcal{M}$ を (k, ℓ)-**個人主義的公平**であるとする．

例えば，前節で示した全体最適なマッチングは $(0, 1)$-個人主義的公平であり，安定なマッチングは $(1, 0)$-個人主義的公平である．これにより，個人主義的公平度を定量的に評価することもできる．

5. 方針違反検知手法

前章では，全体最適であるマッチングや安定なマッチングが，必ずしも個人主義的公平ではないことを示した．これらはいずれの方針が優れているか一概に決定することはできない．予めユーザに方針を提示し、ユーザに判断を委ねることを可能にするべきであると考える．しかしながら，実際に実施されているマッチング処理が，提示された方針と一致していることをいかにしてユーザが確信できるか，という点が課題となる．本章では、この課題についての解決手法を提案する．

5.1 概要

以下が本論文における方針違反検知手法の要件である．

- マッチングサーバで実施したマッチング処理が提示された方針と一致しない場合，ユーザはそれを検知できる
- マッチングサーバ以外にユーザの選好順序を知られることがない

提案手法においては，検知を行うための検知サーバを置く．検知サーバは semi-honest とし，ユーザの選好順序に興味はあるが、決められた処理以外の処理を実施することはないものとする．悪意のあるマッチングサーバの目的は提示された方針とは異なるマッチング処理を実施し，ユーザにそれを検知されないことである．ユーザには悪意のあるものは含まれず，虚偽の入力を与えるなど決められた処理以外の処理を実施することはないものとする．

提案手法では，完全準同型暗号を用いる．ユーザは，自身の選好順序をマッチングサーバに送付する．マッチングサーバは受け取った選好順序から，マッチング結果を各

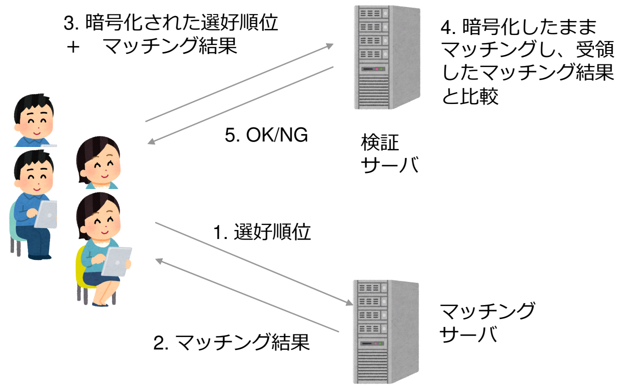

図 4　方針違反を検知する提案手法の概要

ユーザに送付する．ユーザは完全準同型暗号により，暗号化した選好順序とマッチング結果を検知サーバにも送付する．検知サーバは暗号化した選好順序に対し、提示されたマッチング処理を実施し、その結果をユーザから受け取った暗号化されたマッチング結果と照合し，結果をユーザに送付する．提案手法の概要を図 4 に示す．

5.2 完全準同型暗号

完全準同型暗号は暗号化したまま任意の演算が可能となる暗号である．以下に完全準同型暗号の定義を示す [9].

定義 6. 完全準同型暗号 HE とは，正当性，簡潔性，安全性を満たすアルゴリズム (Setup, Keygen, Enc, Dec, Eval) の 5 つ組である．

- **セットアップ** params $\leftarrow$ Setup(1^λ): 公開パラメータ param を出力する．
- **鍵生成** (pk, sk) $\leftarrow$ Keygen(param): 公開の暗号鍵 pk と秘密の復号鍵 sk を出力する．
- **暗号化** $c \leftarrow$ Enc(pk,μ): 公開鍵 pk を用いて，メッセージ $\mu \in \{0, 1\}$ を暗号化し，暗号文 c を出力する．
- **復号** $\mu \leftarrow$ Dec(sk,c): 秘密鍵 sk を用いて，暗号文 c を復号しメッセージ μ を出力する．
- **準同型演算** $\hat{c} \leftarrow$ Eval$(\mathcal{C}, (c_1, \ldots c_\ell)$, pk$)$: 公開鍵 pk を用いて，$c_1, \ldots c_\ell$ に対して，任意の回路 $\mathcal{C} : \{0, 1\}^\ell \to \{0, 1\}$ を適用し，暗号文 $\hat{c}$ を出力する．

正当性とは，暗号アルゴリズム及び準同型演算アルゴリズムによって得られた暗号文が正しく復号できる性質である．簡潔性とは，準同型演算を行った後の暗号文の大きさが演算した回路の大きさに依らない性質である．安全性とは，任意の攻撃者に対し，攻撃成功率が無視できるほど小さい性質である．正当性，簡潔性，安全性の厳密な定義についてはここでは言及しないが，これらを満足する手法は存在する．

任意の回路に対して準同型演算可能であれば，任意のアルゴリズムに対して準同型演算可能な完全準同型暗号を構成できることは自明である．ゆえに，以下のような完全準同型暗号を改めて定義する．以下，完全準同型暗号という

ときには，下記の定義を適用することとする．

定義 7. 完全準同型暗号 HE とは，正当性，簡潔性，安全性を満たすアルゴリズム $(\mathsf{Setup}, \mathsf{Keygen}, \mathsf{Enc}, \mathsf{Dec}, \mathsf{Eval})$ の 5 つ組である．

- **セットアップ** $\mathsf{params} \leftarrow \mathsf{Setup}(1^{\lambda})$: 公開パラメータ param を出力する．
- **鍵生成** $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{Keygen}(\mathsf{param})$: 公開の暗号鍵 pk と秘密の復号鍵 sk を出力する．
- **暗号化** $c \leftarrow \mathsf{Enc}(\mathsf{pk}, \mu)$: 公開鍵 pk を用いて，任意のメッセージ μ を暗号化し，暗号文 c を出力する．
- **復号** $\mu \leftarrow \mathsf{Dec}(\mathsf{sk}, c)$: 秘密鍵 sk を用いて，暗号文 c を復号しメッセージ μ を出力する．
- **準同型演算** $\hat{c} \leftarrow \mathsf{Eval}(\mathcal{A}, (c_1, \dots c_\ell), \mathsf{pk})$: 公開鍵 pk を用いて，$c_1, \dots c_\ell$ に対して，任意のアルゴリズム $\mathcal{A}$ を適用し，暗号文 $\hat{c}$ を出力する．

5.3 提案手法

ここでは一般的な検知手法を示す．n 人のユーザ $\{u_1, \dots u_n\}$ がいるものとする．各ユーザはそれぞれの入力 $\{m_1, \dots m_n\}$ を持つ．あらかじめ実行サーバが実行するアルゴリズム $\mathcal{A}$ は与えられているものとする．$\mathcal{A}$ は決定的アルゴリズムとする．

(1) ユーザの代表者が $\mathsf{Setup}(1^{\lambda})$ を実行し，param を取得し，これを公開する．

(2) ユーザの代表者が $\mathsf{Keygen}(\mathsf{param})$ を実行し，$(\mathsf{pk}, \mathsf{sk})$ を取得する．pk は公開し，sk はユーザ全員に配布する．

(3) 各ユーザ u_i は m_i を安全な通信路を用いて実行サーバに送付する．

(4) 実行サーバは $r' \leftarrow \mathcal{A}'(m_1, \dots, m_n)$ を実行し，r' を全てのユーザに送付する．

(5) 各ユーザ u_i は $c_i \leftarrow \mathsf{Enc}(\mathsf{pk}, m_i \parallel r')$ を安全な通信路を用いて検知サーバに送付する．ここで，$\parallel$ は文字列の連結を表す．

(6) 検知サーバは $x \leftarrow \mathsf{Eval}(\mathcal{A}^V, (c_1, \dots, c_n), \mathsf{pk})$ を実行し，x を全てのユーザに送付する．ここでアルゴリズム $\mathcal{A}^V$ は以下の通りである．
   (a) m_i と r' を分割する．
   (b) $r = \mathcal{A}(m_1, \dots, m_n)$ を実行する．
   (c) $r = r'$ であれば 1 を，$r \neq r'$ であれば 0 を出力する．

(7) 各ユーザ u_i は $\mathsf{Dec}(x, \mathsf{sk})$ を実行し，1 であれば実行サーバがアルゴリズム $\mathcal{A}$ を実行したことを確信できる．

マッチングにおいては，各ユーザの選好順序，マッチングアルゴリズム，マッチング結果 (を適切に符号化したもの) がそれぞれ $m, \mathcal{A}, r$ とみなすと，上記の手法により方針違反を検知することができる．

5.4 考察

sk を全てのユーザで共有しているため，もし c_i が漏洩した場合には参加しているユーザに m_i が漏洩するリスクがある．マルチパーティコンピュテーションならばその問題を解決できる可能性があるが、検知を行う際に全ユーザが参加する必要がある．

一般に完全準同型暗号の処理は非常に重い．しかしながら，今回の提案手法では，マッチングサーバの処理は独立に実行できるため，本来のサービス提供処理への影響はない．不信を抱いたユーザのみ検知サーバに問い合わせる必要がある．全てのユーザが検知を必要とすることは現実的には考えにくいため，本提案手法は実用性と透明性が両立できる手法であると考えられる．他の実現手段として Verifiable Computation を応用することで，検知サーバが不要になるメリットが生まれる可能性があるが，本提案のように本来のサービス提供処理と分離することができず，代償として全てのユーザの処理時間が大きく増大する．

本稿での提案手法では，マッチングが一致することを確認する手段を採用している．一方で，マッチングがある条件を満たしているかどうかを判定する効率の良い手段が確立していれば，条件を満たしているかどうかのみ検証する手法に拡張することは容易である．本稿では，あるマッチングが個人主義的公平であるかどうかを効率よくアルゴリズムを示しているわけではないが，今後そのようなアルゴリズムが発見されれば，本手法はより実用性の高い解決策となりうる．

提案手法は，乱択アルゴリズムに適用することはできない．また, 実際のサービスでは，ユーザの選好順序をユーザ自身が入力するのではなく，マッチングサーバが所有するデータから推定したものを使用するケースが考えられるが，現状ではそのようなケースには対応できない．

6. 既存研究

二部グラフにおけるマッチングの最適化問題は，労働者を仕事に割り当てる問題の数学モデルなどとして，古くからよく研究されている．安定結婚問題は，このマッチングに安定性という概念を採用した問題であり，Gale と Shapley により提案された [3]．Gale と Shapley らはまた，安定なマッチングを $\mathcal{O}(n^2)$ で見つけるアルゴリズムを提案した．安定結婚問題についてはこれまでに様々な拡張問題が考えられてきた．拡張問題の一分野として，安定結婚問題における公平性の議論がある．Gale-Shapley アルゴリズムによって見つかる安定なマッチングは，男性（または女性）最適なマッチングであり，また全体としてのコストの最適化などは考慮されていなかった．最小コスト安定マッチング，最小後悔安定マッチング [7]，男女平等安定マッチング [8] などの安定結婚問題を拡張した最適化問題が研究されてきた．これらの問題は，全体最適的な意味合いでの

公平性を意味している．本稿では，必ずしも安定とは限らない男女マッチングにおける個人主義的な意味合いでの公平性について新たに定義しており，このような観点から議論している既存研究は著者の知る限りない．

近年機械学習分野においては，公平性の議論が活発に行われている．この分野においては，大きく分けてグループ公平性（Group Fairness）と個人公平性 (Individual Fairness) が議論されている．グループ公平性の定義の一つである Demographic parity は，性別などのセンシティブ属性で条件づけられた場合の予測値の分布が一致することをもって公平とみなす [4]．個人公平性の定義の一つである Lipschitz property [5] は，似たような特徴を持つ二者に対する予測値の分布が一定範囲に収まることが保証される場合に公平とみなす．強化学習の典型モデルである多腕バンディット問題における個人公平性に関して，Fair Bandit がある [6]．Fair Bandit における公平性は，利得の高い腕を選択する確率が利得の低い腕を選択する確率を下回ることがない制約を満たすことである．本稿における個人主義的公平性の定義は，Fair Bandit における公平性の概念の影響を受けているが，対象が機械学習ではなく選好順位を持つ男女のマッチング問題に関する点が異なる．

7. おわりに

本稿では，安定結婚問題における男女マッチングの問題設定をベースとして，これまでに考えられてきた全体最適的な意味合いでの公平性や男女間の公平性とは異なる観点で，個人主義的公平性を新たに定義した．次に，全体最適なマッチングや安定なマッチングが、個人主義的な不公平を生じる事例を示した．さらに，例えば個人主義的に公平なマッチングを実施していると提示しておきながら，実際には全体最適なマッチングや安定なマッチングなど，異なる方針によってマッチングを実施されることを検知する手法について提案した．

参考文献

[1] Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction*, 30(1):127–158, 2020.

[2] AI ネットワーク社会推進会議. 報告書 2021〜「安心・安全で信頼性のある AI の社会実装」の推進〜 第 3 章「安心・安全で信頼性のある AI の社会実装」の推進の取組.

[3] D. Gale and L.S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69:9–15, 1962.

[4] Moritz Hardt, Eric Price, and Nathan Srebro. Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*, pages 3323–3331, 2016.

[5] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. *ITCS 2012 - Innovations in Theoretical Computer Science Conference*, pages 214–226, 2012.

[6] Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. Fairness in Learning: Classic and contextual bandits. *Advances in Neural Information Processing Systems*, pages 325–333, 2016.

[7] R.W. Irving, P. Leather, and D. Gusfield. An efficient algorithm for the “optimal” stable marriage. *Journal of the ACM*, 34(3):532–543, 1987.

[8] A. Kato. Complexity of the sex-equal stable marriage problem. *Japan Journal of Industrial and Applied Mathematics*, 10:1–19, 1993.

[9] Zvika Brakerski and Renen Perlman. Lattice-based fully dynamic multi-key FHE with short ciphertexts. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9814(468):190–213, 2016.