

グループフェアネスを考慮したウェブ検索タスク

酒井 哲也^{1,a)}

概要: 本稿では、NTCIR-17 に提案しているウェブ検索タスク (FairWeb-1) の構想について述べる。このタスクでは、検索エンジンユーザの観点から見た適合性のみならず、検索されるエンティティ (人やコンテンツ) の観点から見たグループフェアネスを考慮する。本稿の目的は (1) 早い段階からタスク仕様についてフィードバックをいただくこと、および (2) 関連研究コミュニティにタスクの意義をご理解いただき、参加をご検討いただくことである。FairWeb-1 の主な新規性は、グループフェアネスを評価する際に順序的グループを適切に扱っている点と、名義的・順序的グループを同時に扱う交差的グループフェアネスを扱っている点である。このタスクを通して、グループフェアネスの観点からのウェブ検索アルゴリズムの課題を明らかにすることを目指している。

A Web Search Task That Takes Group Fairness into Account

TETSUYA SAKAI^{1,a)}

Abstract: We describe our plan for FairWeb-1, a new web search task we are proposing for NTCIR-17. This task considers not only relevance from the viewpoint of search engine users but also group fairness from the viewpoint of entities (i.e., people or contents) that are being sought. This paper aims at (1) collecting feedback on the task design early on; and (2) letting the importance of the task be known to relevant research communities, and to encourage task participations. The task is novel in that it properly handles ordinal groups when evaluating group fairness, and in that it accommodates the evaluation of intersectional group fairness given both ordinal and nominal groups. Through this task, we aim to clarify open research questions for web search algorithms regarding group fairness.

1. はじめに

情報アクセスの評価型国際会議 NTCIR (NII Testbeds and Community for Information access Research)^{*1} は 1999 年以来約 1 年半毎に開催されており [8], [10], 2023 年 12 月には第 17 回 (NTCIR-17) を迎える。本稿では、NTCIR-17 に提案しているウェブ検索タスク (FairWeb-1) の構想について述べる。このタスクでは、検索エンジンユーザの観点から見た適合性 (relevance) のみならず、検索されるエンティティ (人やコンテンツ) の観点から見たグループフェアネス (group fairness) を考慮する。例えば、国際会議の運営委員を選出するために「情報検索の研究者の情報を収集したい」という情報要求をユーザがもって

る状況を考える。このとき、検索エンジンが h-index の非常に高い著名な研究者の情報ばかり出力した場合、h-index が現時点では高くない研究者にチャンスが与えられないことになりうる。これを防ぐアプローチのひとつとして、例えば h-index が低い研究者のグループ、中くらいの研究者のグループ、高い研究者のグループというグループ分けを考え、各グループの検索エンジンユーザへの露出 (exposure) ができるべく公平になるように検索アルゴリズムを工夫することが考えられる。これがグループフェアネスの考え方である。また、上記のような特定の観点に基づきエンティティ (entity; この例では「研究者」) をグループ化したときのグループの集合を属性集合 (attribute set) と呼ぶ。人間というエンティティを扱う場合、グループフェアネスの観点から重要な属性集合の例としてジェンダーや国籍が挙げられる。

本稿の目的は (1) 早い段階から FairWeb-1 のタスク仕様

¹ 早稲田大学 (Waseda University)

^{a)} tetsuyasakai@acm.org

^{*1} <http://research.nii.ac.jp/ntcir/index-ja.html>

についてフィードバックをいただくこと、および (2) 関連研究コミュニティにタスクの意義をご理解いただき、参加をご検討いただくことである。FairWeb-1 の主な新規性は、グループフェアネスを評価する際に名義的グループ (例えば「女性」) に加えて順序的グループ (例えば「h-index が 50 を超える研究者」) も適切に扱っている点と、両者を同時に扱う交差的グループフェアネス (intersectional group fairness) を扱っている点である。^{*2} このタスクを通して、現状の検索アルゴリズムにおけるグループフェアネスの観点からの課題を明らかにすることを目指している。

2. 提案するタスクの概要

2.1 システムの入出力

本タスクにおける参加システムへの入力には以下の 3 つである。

- 検索トピック。一定の条件を満たすエンティティに関する情報要求を記述したもの。
- 上記検索トピックに付随するひとつ以上の属性集合。
- 上記各属性集合に対するターゲット分布。これは、属性集合に属するグループ群に対して定義される確率質量関数 (Probability Mass Function: 以下 PMF) 。

例えば、検索トピック「情報検索の研究者」(エンティティ: 研究者) とともにジェンダーに関する属性集合を与え、各ジェンダーグループの露出が均等になるようにターゲット分布を設定する。この例において、情報検索の研究をしている個々の研究者は、後述のアノテーションにより適合エンティティ (relevant entity) と認定される可能性がある。

参加システムの出力は各検索トピックに対するウェブページのランクつきリスト (SERP: Search Engine Result Page) である。この SERP には、以下の 2 つの性質を満たすことが求められる。

適合性 検索トピックに適合した情報を (なるべく高い順位に) 多く含むこと。

グループフェアネス SERP が達成した各属性集合上の分布 (実績分布) が、ターゲット分布になるべく近いこと。例えば、前述の検索トピックに対して、男性研究者の情報ばかりを含んだ SERP は適合性の観点では高得点となりうるが、ジェンダーに関するグループフェアネスの観点からは低得点となる。

システム評価においては、従来のアドホック検索タスク [4] と同様に n 件の検索トピックを用い、後述する評価指標をサイズ n のトピック集合に関して平均する。本タスクは、検索トピックに関する適合性に加えて各属性集合に関するグループフェアネスを扱っているため、アドホック

検索タスクを一般化したものとなっている。なお n は予備実験をもとに統計的検出力の観点から決定する [3], [10]。

2.2 タスクのワークフロー

本タスクは以下の手順で実施される予定である。

- (1) 各参加チームは、オーガナイザから与えられたサイズ n のトピック集合に含まれる各トピック、および付随する属性集合とターゲット分布に対してシステムにより SERP を作成し、これらの結果をまとめたラン (run) というファイル [4], [10] を期日までに提出する。
- (2) オーガナイザは、スクリプト NTCIRPOOL^{*3} を用いて各トピックに対し深さ k のプールファイル [4], [10] を作成する。これは各システムの SERP の上位 k 件のウェブページの和集合により構成される。
- (3) アノテータは、プールファイル中の各ウェブページを精査し、高々 3 件の適合エンティティをアノテータする。ここで、各エンティティが検索トピックに適合するか否かは、そのエンティティに関する根拠 (rationale) をウェブ検索により探して判断する。
- (4) オーガナイザは、アノテータによる適合エンティティのアノテーションをもとに、プールファイル中の各ウェブページに対し適合性およびグループ所属性 (group membership) を導出する。その上で、各ランを適合性およびグループフェアネスの観点から評価する。

上記手順の (1) に関連して、3 章で本タスクで扱う検索トピックと属性集合について説明する。手順 (2) は通常のアドホック検索 [4], [10] と要領が同じなので説明を省略する。手順 (3) の適合エンティティのアノテーションについては 4 章で説明する。手順 (4) のページ適合性およびグループ所属性の導出について 5 章で、ランの評価について 6 章で説明する。

3. エンティティ、トピック、属性集合

3.1 エンティティタイプおよびトピックタイプ

2.1 節で述べたように、本タスクにおける検索トピックは適合エンティティが満たすべき条件を記したものである。エンティティとしては以下の 5 種類を考える。

R エンティティ 研究者 (researcher)。

M エンティティ 映画 (movie)。

T エンティティ Twitter アカウント。

Y エンティティ YouTube 映像。

B エンティティ 本 (book)。

各トピックは上記エンティティのいずれか 1 つを扱うものとし、そのようなトピックをそれぞれ R トピック, M トピックなどと呼ぶ。各トピックタイプの具体例を以下に示す。

^{*2} ウェブ検索ではなく Wikipedia 検索を扱う TREC 2022 Fair Ranking Track (https://fair-trec.github.io/docs/Fair_Ranking_2022_Participant_Instructions.pdf) でも交差的グループフェアネスが扱われる予定だが、少なくとも現時点では順序的グループの性質が考慮されていない。

^{*3} <https://research.nii.ac.jp/ntcir/tools/ntcirpool-en.html>

R *information retrieval researchers*
M *Daniel Craig 007 movies*
T *Twitter accounts that provide info on COVID*
Y *Coldplay covers on YouTube*
B *Sci-fi novels published in 1999*

ここで、下線部はトピックタイプの手がかりとなる部分を
示し、イタリック体の部分は適合エンティティの満たすべ
き条件に相当する。

本タスクでは、上記各トピックタイプについて、以下の
各節で説明する属性集合を考慮した上でグループフェアネ
スの観点からシステム評価を行う。

3.2 R トピックの扱う属性集合

HIINDEX 各研究者の Google scholar^{*4} h-index の値を
 x とするとき、以下の 4 つの順序的グループを考える:
グループ 1 ($x < 10$), グループ 2 ($10 \leq x < 30$), グ
ループ 3 ($30 \leq x < 50$), グループ 4 ($x \geq 50$).

GENDER 研究者の公式バイオグラフィー中に he もし
くは she という指示代名詞が使われているか否かをも
とに、以下の 3 つの名義的グループを考える: グルー
プ 1 (he), グループ 2 (she), グループ 3 (other). す
なわち other は単にバイオグラフィー中に he も she も
出現しないという意味である。

3.3 M トピックの扱う属性集合

REVIEWS 各映画の IMDb ページ^{*5} における review 数
を x とするとき、以下の 4 つの順序的グループを考える:
グループ 1 ($x < 100$), グループ 2 ($100 \leq x < 10K$),
グループ 3 ($10K \leq x < 1M$), グループ 4 ($x \geq 1M$).

ORIGIN 各映画の IMDb ページにおける “country of
origin” に記載された情報に基づき、各映画を 8 程度の
地理的領域 (geographic region) のうち 1 つ以上にマッ
ピングする。地理的領域へのマッピングは、米 CIA が
管理する The World Factbook^{*6} に記載された情報をも
とに行う予定である。ORIGIN は、FairWeb-1 が扱
う属性集合のうち唯一、エンティティに関する soft な
グループ所属性 (soft group membership) [7] を扱う
ものである。すなわち、例えば日米共同制作の映画の
場合のように、あるエンティティが複数の地理的領域
に同時に所属することを許容する。

3.4 T トピックの扱う属性集合

FOLLOWERS 各 Twitter アカウントの follower 数を x
とするとき、REVIEWS と同様に 4 つの順序的グルー
プを考える。

3.5 Y トピックの扱う属性集合

SUBSCS 各 YouTube 映像の投稿者の subscriber 数を x
とするとき、REVIEWS と同様に 4 つの順序的グルー
プを考える。

3.6 B トピックの扱う属性集合

RATINGS 各書籍の Amazon.com ページにおける
customer global rating 数を x とするとき、REVIEWS
と同様に 4 つの順序的グループを考える。

4. 適合エンティティのアノテーション

通常のアドホック検索タスクでは、適合性判定者がプー
ルされた各文書に対して適合性判定を行う。これに対し
て、本タスクにおけるアノテータの仕事は、プールされた
各文書から適合エンティティを検出し、それに付随する情
報を抽出することである。このために文書の適合性判定
ツール [4] に似たブラウザベースの判定者向けツールを用
意する。なお、各ウェブページから検出する適合エンティ
ティは高々 3 つとする。これには以下の 2 つの意味があ
る。(1) 多くの適合エンティティを含むページは個々のエ
ンティティに関する情報が比較的少ないものと仮定し、
このようなページを検索結果に含めることを積極的に推
奨しない。(2) ひとつのページに対するアノテーションの
労力が多大になることを回避する。なお、アノテーション
の信頼性向上のため、各ページのアノテーションは 2 人の
アノテータに独立に行ってもらう予定である。

アノテーションに含まれる情報はトピックタイプにより
異なる。表 1 に各適合エンティティに対しアノテータする
情報の一覧を示す。なお、順序的グループの所属性に用い
る生の統計量 (例えば h-index の値) は時間とともに変化
するが、アノテーションはあくまでその時点でウェブ上で
確認できる情報をもとに行う。

5. ページ適合性およびグループ所属性の導出

4 章で示したアノテーション作業が完了したのち、オー
ガナイザはその結果をもとに以下のように各ページの適合
性およびグループ所属性を導出する。

5.1 ページ適合性の導出

ページ d から抽出された適合エンティティの集合を $E(d)$
とする。(この集合のサイズは、2 名のアノテータが全く異
なる適合エンティティを 3 件ずつ抽出した場合に最大値 6
をとる。) 適合エンティティ e の適合レベル $r(e)$ を、2 名
の判定者が共に適合エンティティと認定した場合に 2、片
方の判定者のみが適合エンティティと認定した場合に 1 と
定義する。このとき、 d の適合レベル $g(d)$ を以下のように
定義する。

*4 <https://scholar.google.com/>

*5 <https://www.imdb.com/>

*6 <https://www.cia.gov/the-world-factbook/>

表 1 各適合エンティティに対しアノテートする情報.

タイプ	判定インタフェースにより記録する情報
R	研究者名 公式バイオグラフィーの URL he/she/other (上記ページより判定) Google Scholar URL (見つからない場合は NA) h-index の値 (上記ページより取得)
M	映画名 IMDb の URL Review 数 (上記ページより取得) 国・地域の一覧 (上記ページの country of origin 情報より取得)
T	Twitter アカウントの URL Follower 数 (上記ページプロフィールより取得)
Y	YouTube 映像の URL subscriber 数 (上記ページのアカウント情報より取得)
B	amazon.com 上の本の URL global rating 数 (上記ページより取得)

$$g(d) = \begin{cases} 0 & (E(d) = \emptyset); \\ \max_{e \in E(d)} r(e) & (\text{otherwise}). \end{cases} \quad (1)$$

すなわち、ページ中の適合エンティティの適合レベルの最大値をページ全体の適合レベルとする。従ってページ適合レベル $g(d)$ の最大値 G もまた 2 である。

5.2 ページのグループ所属性の導出

まず、属性集合 C が順序的グループ $C_1, \dots, C_{|C|}$ により構成される場合を考える。適合エンティティ $e \in E(d)$ に関する生の統計量 (例えば HINDEX を考えた場合の特定の研究者の h-index の値) をグループに写像するフラグ $F(e, C_i)$ を考える。すなわち、 e がグループ C_i に属するなら $F(e, C_i) = 1$ 、属さないなら $F(e, C_i) = 0$ とする。例えば 3.2 節で定義した HINDEX グループによれば、h-index が 5 の研究者はグループ 1 に属するので $F(e, C_1) = 1, F(e, C_2) = F(e, C_3) = F(e, C_4) = 0$ である。ひとつのエンティティが複数のグループに属することはないので、 F は適合エンティティに対する hard なグループ所属性関数である [7]。

上記のグループ所属性関数に基づき、ページ d がグループ C_i に所属する確率を以下で定義する。

$$P(d, C_i) = \begin{cases} 1/|C| & (E(d) = \emptyset); \\ \frac{|\sum_{e \in E(d)} F(e, C_i)|}{|\sum_i \sum_{e \in E(d)} F(e, C_i)|} & (\text{otherwise}). \end{cases} \quad (2)$$

次に、属性集合が名義的グループにより構成される場合であるが、このうち GENDER については各適合エンティティに対し hard なグループ所属性関数が定義される (すなわち he, she, other は排他的である) ので、式 2 が同様に適用できる。一方、ORIGIN については一般にひとつの映画に対し複数の地理的領域が対応するため、 F の代わり

に soft なグループ所属性を扱う $G(e, C_i)$ を用いる。すなわち、適合エンティティ $e \in E(d)$ に対応する地理的領域の集合を $ORIGIN(e) (\subseteq C)$ とし、 $|ORIGIN(e)| = m (\geq 1)$ であるとき、

$$G(e, C_i) = \begin{cases} 1/m & (C_i \in ORIGIN(e)); \\ 0 & (\text{otherwise}). \end{cases} \quad (3)$$

このとき、ページ d がグループ C_i に所属する確率を以下で定義する。

$$P(d, C_i) = \begin{cases} 1/|C| & (E(d) = \emptyset); \\ \frac{|\sum_{e \in E(d)} G(e, C_i)|}{|\sum_i \sum_{e \in E(d)} G(e, C_i)|} & (\text{otherwise}). \end{cases} \quad (4)$$

式 3 より、 $G(e, C_i)$ が $F(e, C_i)$ を一般化したものであることは明らかである。逆にいうと、soft なグループ所属性関数 $G(e, C_i)$ において $m = 1$ としたものが hard なグループ所属性関数 $F(e, C_i)$ である。また、エンティティのグループ所属性が hard であるか soft であるか否かにかかわらず、ページのグループ所属性は一般に soft である。例えば、あるページから h-index が 5 の研究者と h-index が 50 の研究者が検出された場合、このページのグループ所属確率はグループ 1 および 4 の両方について正となる。

6. 評価指標

6.1 定義

本タスクではシステムの評価に GFR (Group Fairness and Relevance) フレームワーク [7] を用いる。GFR は、SERP L を以下のように評価する。

$$GFR(L) =$$

$$\sum_{k=1}^{|L|} Decay_{L@k} \left(w_0 Utility_{L@k} + \sum_{m=1}^M w_m DistrSim_{L@k}^m \right). \quad (5)$$

式 5 において、 $Decay_{L@k}$ は Expected Reciprocal Rank (ERR) [1], [10] のユーザモデルに基づくユーザが第 k 位の文書 $d_{L@k}$ を閲覧した上で SERP L を放棄する確率である。まず、ユーザがページ $d_{L@k}$ に満足する確率をページ適合レベル $g(d_{L@k}) \in \{0, 1, 2\}$ により以下のように求める。

$$p_{L@k}^{\text{sat}} = \frac{2^{g(d_{L@k})} - 1}{2^G}. \quad (6)$$

5.1 節で示したように、本タスクでは $G = 2$ である。これに基づき、 $Decay_{L@k}$ は以下のように計算される。

$$Decay_{L@k} = \begin{cases} p_{L@1}^{\text{sat}} & k = 1 \\ p_{L@k}^{\text{sat}} \prod_{j=1}^{k-1} (1 - p_{L@j}^{\text{sat}}) & \text{otherwise}. \end{cases} \quad (7)$$

式 5 における $Utility_{L@k}$ は第 k 位で L を放棄したユーザにとっての L の有用性を示す指標で、本タスクでは以下

の2種類を考慮する.

$$Utility_{L@k}^{ERR} = 1/k, \quad (8)$$

$$Utility_{L@k}^{iRBU} = \phi^k. \quad (9)$$

式7と式8の積がERR, 式7と式9の積がIntentwise Rank-Biased Utility (iRBU) と呼ばれる適合性のみに基づく評価指標である. Sakai ら [9] による SERP 優劣判断におけるユーザと評価指標の整合性の実験結果に基づき, iRBU については $\phi = 0.99$ とする. $Utility_{L@k}^{ERR}$ が順位 k とともに $1, 0.5, 0.33, 0.25, \dots$ と急激に減少するのに対し, $Utility_{L@k}^{iRBU}$ ($\phi = 0.99$) は $0.99, 0.98, 0.97, 0.96, \dots$ と非常に緩やかに減少することに注意してほしい. これは iRBU の立場からの SERP 上位 k 件の有用性は k (すなわちユーザが閲覧する文書数) が大きくなってあまり目減りしないことを意味する.

式5における $DistrSim_{L@k}^m$ は, 第 m の属性集合に関し, SERP L の上位 k 件の文書による実績分布 $D_{L@k}^m$ のターゲット分布 D_*^m との類似度を測る関数である. まず, 上位 k 件の文書による実績分布 $D_{L@k}^m$ は, グループ C_i の確率 $P_{L@k}^m(C_i)$ を以下のように計算したものである.

$$P_{L@k}^m(C_i) = \frac{\sum_{j=1}^k P(d_{L@j}, C_i)}{k}. \quad (10)$$

その上で, $DistrSim_{L@k}^m$ は2つのPMFのダイバージェンスを用いて以下のように計算される.

$$DistrSim_{L@k}^m(D_{L@k}^m \parallel D_*^m) = 1 - Divergence_{L@k}^m(D_{L@k}^m \parallel D_*^m) \quad (11)$$

ここで, ダイバージェンス関数は $[0, 1]$ の範囲をとる3種類を用意している. まず, 名義的なグループを扱う場合は Jensen-Shannon Divergence (JSD) [2], [6] を用いる. 一方, 順序的なグループを扱う場合はこれに適した Normalised Match Distance (NMD) もしくは Root Normalised Order-aware Divergence (RNOD) を用いる [5], [6].

以下, 本タスクに依存しない形でダイバージェンス関数を定式化するために $P_{L@k}^m(C_i)$ を P_i と略記し, これに対応するターゲット分布における確率を P_i^* と略記する. さらに, PMF $D_{L@k}^m, D_*^m$ をそれぞれ D, D^* と略記する.

まず, 名義的なグループ向けのダイバージェンスである JSD は以下のように定義される.

$$JSD(D, D^*) = \frac{KLD(D \parallel D') + KLD(D^* \parallel D')}{2}, \quad (12)$$

ここで D' はグループ C_i の確率が $P_i' = (P_i + P_i^*)/2$ により定義されるPMFで,

$$KLD(D \parallel D') = \sum_{i \text{ s.t. } P_i > 0} P_i \log_2 \frac{P_i}{P_i'}, \quad (13)$$

$$KLD(D^* \parallel D') = \sum_{i \text{ s.t. } P_i^* > 0} P_i^* \log_2 \frac{P_i^*}{P_i'} \quad (14)$$

である.

順序的グループに適したNMDはPMF上のEarth Mover's Distanceであり, $CP_i = \sum_{j=1}^i P_j, CP_i^* = \sum_{j=1}^i P_j^*$ とするとき以下のように計算できる.

$$NMD(D, D^*) = \frac{\sum_{i=1}^{|C|} |CP_i - CP_i^*|}{|C| - 1}. \quad (15)$$

一方, RNODは以下のように定義される. まず, Distance Weighted sum of squares (DW) を各グループ C_i について以下のように計算する.

$$DW_i = \sum_{j=1}^{|C|} |i - j| (P_j - P_j^*)^2. \quad (16)$$

次に, $C^* = \{C_i \in C \mid P_i^* > 0\}$ (すなわちターゲット分布上の確率が正であるグループの集合) を考え, 2つのPMFをもとに以下のOrder-aware Divergence (OD) を計算する.

$$OD(D \parallel D^*) = \frac{\sum_{i \text{ s.t. } C_i \in C^*} DW_i}{|C^*|}. \quad (17)$$

RNODは正規化したODの平方根である.

$$RNOD(D \parallel D^*) = \sqrt{\frac{OD(D \parallel D^*)}{|C| - 1}}. \quad (18)$$

順序的グループを扱う場合はこれに適したダイバージェンスを用いたほうがよい理由は以下の例を考えれば明らかであろう. 今, h-indexが低いグループを強調したターゲット分布を考え, h-indexが中くらいのグループを強調した実績分布 X と, h-indexが高いグループを強調した実績分布 Y があるものとする. この場合, JSDなどの指標では「 Y よりは X のほうが好ましい」という評価ができず, 「いずれも外れ」という評価しかできない [5], [6].

式5は重み w_0, w_1, \dots, w_M を含むが, 本タスクではこれらの値を全て等しいものとする. すなわち, GFRはひとつの適合性指標と $M \in \{1, 2\}$ 個のグループフェアネス指標の平均により求める. 3章で示したように, RトピックおよびMトピックについては $M = 2$ であり, 順序的グループを含む属性集合と名義的グループを含む属性集合の双方を考慮した交差的グループフェアネスを扱うことになる.

6.2 計算例

本タスクにおける評価方法を図1の例題により説明する. ここでは簡単のため, SERPのサイズを $|L| = 3$ としており, RトピックのHINDEX属性集合のみを考える. この例では, 第1位の文書が適合エンティティとして研究者X, Y, Z (h-indexがそれぞれ5, 6, 20, 適合レベルがそれぞれ2,1,2) を含んでおり, 第2位の文書には適合エンティティ

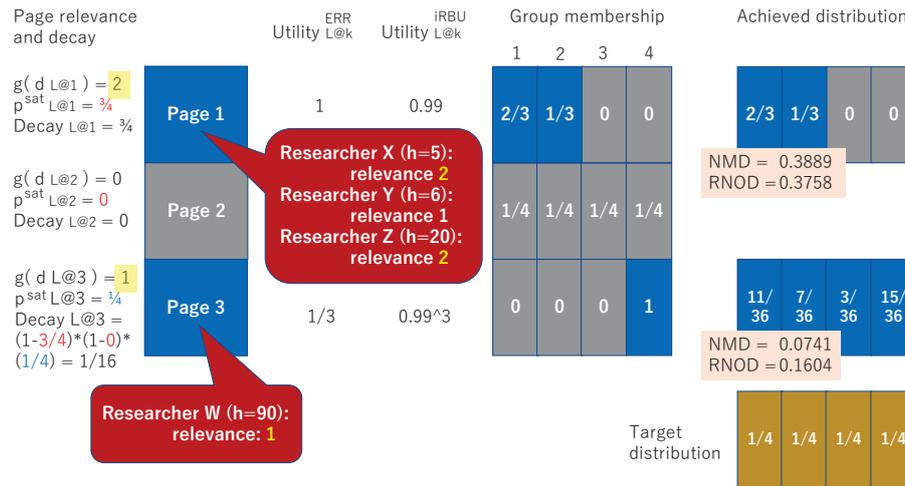


図 1 GFR の計算方法を示すための例題。

がなく、第 3 位の文書が適合エンティティとして研究者 Z (h-index が 90, 適合レベルが 1) を含んでいる。

まず、適合エンティティの適合レベルと式 1 より、上記 3 文書の適合レベルはそれぞれ 2,0,1 となる。従って、各文書の満足確率は式 6 より $3/4, 0, 1/4$, 第 $k (= 1, 2, 3)$ 位における $Decay_{L@k}$ は式 7 より $3/4, 0, 1/16$ となる。また、式 8 より $Utility_{L@1}^{ERR} = 1$, $Utility_{L@3}^{ERR} = 1/3$, 式 9 より $\phi = 0.99$ のとき $Utility_{L@1}^{iRBU} = 0.99$, $Utility_{L@3}^{iRBU} = 0.99^3 = 0.9801$ となる。式 5 の Decay と Utility の積の和の部分 (すなわち ERR と iRBU) はそれぞれ 0.7708, 0.8031 となる。

次に、図 1 の右側を用いてグループフェアネスを評価する。まず、第 1 位において研究者 X および Y がグループ 1 に所属し、Z がグループ 2 に所属することから、式 2 よりグループ所属確率ベクトル $(2/3, 1/3, 0, 0)$ が得られる。これはそのまま第 1 位における実績分布となる。一方、第 2 位については適合エンティティが存在しないので一様分布、第 3 位についてはグループ 4 に所属する研究者 W のみを含むことから $(0, 0, 0, 1)$ となる。 $k = 3$ における実績分布は式 10 より $(11/36, 7/36, 3/36, 15/36)$ である。以下、紙面の都合上 RNOD の議論を割愛する。ターゲット分布 (一様分布) の累積確率ベクトルは $(1/4, 2/4, 3/4, 1)$ であり、 $k = 1, 3$ における実績分布の累積確率ベクトルは $(2/3, 1, 1, 1)$ と $(11/36, 18/36, 21/36, 1)$ であるから、NMD は式 15 より 0.3889, 0.0741 である。式 5 における Decay と DistrSim の積の和は $(3/4) * (1 - 0.3889) + (1/16) * (1 - 0.0741) = 0.5162$ となる。

7. まとめ

本稿では、NTCIR-17 に向けて設計したグループフェアネスを考慮したウェブ検索タスク (FairWeb-1) の構想について紹介した。「WebDB 夏のワークショップ」のオーディエンスより有益なフィードバックをいただくこと、またタ

スク参加のご検討をいただくことを期待している。

謝辞 本タスクの設計に有用なフィードバックを下された荒井ひろみ氏 (RIKEN AIP), Nuo Chen 氏 (早稲田大学), Michael Ekstrand 氏 (Boise State University) に感謝します。

参考文献

- [1] Chapelle, O., Metzler, D., Zhang, Y. and Grinspan, P.: Expected Reciprocal Rank for Graded Relevance, *Proceedings of ACM CIKM 2009*, pp. 621–630 (2009).
- [2] Lin, J.: Divergence Measures Based on the Shannon Entropy, *IEEE Transactions on Information Theory*, Vol. 37, No. 1, pp. 145–151 (1991).
- [3] Sakai, T.: *Laboratory Experiments in Information Retrieval: Sample Sizes, Effect Sizes, and Statistical Power*, Springer (2018).
- [4] Sakai, T.: How to Run an Evaluation Task: with a Primary Focus on Ad Hoc Information Retrieval, *Information Retrieval Evaluation in a Changing World: Lessons Learned from 20 Years of CLEF* (Ferro, N. and Peters, C., eds.), Springer, pp. 71–102 (2019).
- [5] Sakai, T.: A Closer Look at Evaluation Measures for Ordinal Quantification, *Proceedings of the 1st International Workshop on Learning to Quantify* (2021).
- [6] Sakai, T.: Evaluating Evaluation Measures for Ordinal Classification and Ordinal Quantification, *Proceedings of ACL-IJCNLP 2021* (2021).
- [7] Sakai, T., Kim, J. Y. and Kang, I.: A Versatile Framework for Evaluating Ranked Lists in terms of Group Fairness and Relevance, (online), available from <https://arxiv.org/abs/2204.00280> (2022).
- [8] Sakai, T., Oard, D. W. and Kando, N.(eds.): *Evaluating Information Retrieval and Access Tasks: NTCIR's Legacy of Research Impact*, Springer (2020).
- [9] Sakai, T. and Zeng, Z.: Retrieval Evaluation Measures that Agree with Users' SERP Preferences: Traditional, Preference-based, and Diversity Measures, *ACM TOIS*, Vol. 39, No. 2 (2021).
- [10] 酒井哲也: 情報アクセス評価方法論: 検索エンジンの進歩のために, コロナ社 (2015).