

不確実性に対応する属性推薦を用いた人物画像検索

宜 鯤† 吉田 登†† 董テイテイ†† 劉 健全††

† 京都大学大学院経済研究科 〒 606-8501 京都市左京区吉田本町

†† NEC ビジュアルインテリジェンス研究所 〒 211-8666 神奈川県川崎市中原区下沼部 1753

E-mail: †yi.kun.52n@st.kyoto-u.ac.jp, ††{n-yoshida14,t-dong,jqliu}@nec.com

あらまし 属性に基づいた人物画像検索は、画像から検出した人物の属性を用いて、指定した属性条件を満たしている人物画像を検索する。多くの既存研究は属性の識別やギャラリー画像のランキングに重心を置き、どの属性を指定すべきかについてほとんど言及していない。一方、ユーザが年齢、性別、服装の色などの数多くある属性を用いる場合、素早くターゲットを探し出すためには、ギャラリー画像が持つ属性の構造や特徴を捉え、適切な属性をユーザに推薦することも重要である。また、ユーザが属性を指定する際に、誤認識や誤記憶により、ターゲットの年齢や服装の色を間違えることがあり、ユーザ回答をそのまま受け入れると、ターゲットが排除される恐れがある。そこで、本研究は、ユーザ回答の不確実性を考慮に入れた属性推薦を用いた人物画像検索システムを提案する。既存手法との比較実験により、提案手法の優位性を示す。

キーワード 人物検索, 画像, 属性推薦, 不確実性

1 はじめに

近年、画像または映像から対象人物を検索することはインテリジェンスなビデオ監視、人とコンピューターの相互作用などを含む数多くの分野での幅広い応用が考えられ、大きく注目されている [18], [19], [20].

外観属性に基づく人物画像検索は、対象人物の画像が入手できない場合に特に有用である [13], [14], [15], [8]. 本手法を用いた場合、年齢、性別、服装の色、鞆等の所持品の種類などの属性から、対象人物が持つ特徴的な属性を検索クエリに入力することで、見つけることができる。逆に、ユーザーが対象人物の非特徴的な属性、例えば映像中のほとんどの人物が有する属性を検索クエリに入力した場合、該当する人物の数が多くなり、対象人物の発見に余計な時間がかかる恐れがある。従って、素早く対象人物を探し出すためには、検索クエリに入力すべき適切な属性をユーザに推薦することが重要である。そのためには、ギャラリー画像群が持つ属性の構造や分布を捉え、属性を推薦する順番を決めるポリシーが必要である。

数多くのデータの集合の中から、ユーザの好みに合うものを素早く見つけるために入力すべき項目を推薦する技術はプロダクトサーチの分野で研究されている [6], [2], [4]. しかし、本手法を人物画像検索における属性推薦に適用した研究例は我々の知る限りではまだ存在しない。それは、以下の2つの問題点が原因であると考えられる。

一つ目は、ディテクターの推定によって付与される属性が不確実性を持つことである。プロダクトサーチの場合、プロダクトの提供者がラベルを付けることが一般的である。それに対し、人物画像検索ではラベル付けされていない膨大なギャラリー画像から検索を行うが、手作業で属性ラベルを付けることは現実的ではないため、ディテクターによる属性推定を用いてラベル付

けを行う必要がある。ディテクターで推定した属性情報は不確実性を含んでおり、不確実性は推定した属性に対する確信度スコアとして出力される。多くの従来技術では確信度の閾値処理でラベルを付与しており、これによって検索で不都合が生じる場合がある。例えばある人物に対して男性、女性それぞれの確信度スコアが 0.1, 0.4 であり、これを閾値 0.5 で処理した場合、この人物は性別に関するラベルを有さず、どちらを条件に検索を行っても結果から排除されてしまう。従って、推定された属性の不確実性を適切に捉えることが課題になる。

二つ目は、ユーザが推薦に基づいて属性を回答する際に、必ずしも正しい属性を答えるわけではないという点である。ユーザは記憶から対象人物の特徴を思い出して回答するため、対象人物の一部の特徴を忘れていたり、間違えて記憶していたなどの理由により回答が誤る可能性がある。それに加え、対象人物を目撃した際の照明条件などの環境条件や視点、隠れの有無などにより、ユーザが対象人物の属性をそもそも誤って認識することも、ユーザ回答に不確実性が生じる要因として挙げられる。このようにユーザが指定した属性条件に不確実性が含まれると、ユーザが指定した検索条件と、ユーザが探したい人物の属性情報とにギャップが生じ、望む結果を得ることが難しくなる。結果として、ユーザは検索結果から対象人物を見つけてことができなくなり、対象人物を見つけるために何回も条件を変更しながら検索を試さなければならなくなる。

さらに、同じ人物画像に対し、ユーザによる認識とディテクターによる認識がずれている可能性がある。一例として、ユーザから「対象人物の上着の色は青である」という属性が入力されたとき、図 1a, 図 1b で示されているように、ユーザとディテクターにとって「青」の認識がずれていた場合、対象人物は結果から排除される恐れがある。

そこで、ユーザ入力の不確実性とディテクター推定結果の不



(a) ユーザにとっての青色の (b) デテクターにとっての
上着. 青色の上着.

図 1: ユーザとデテクターの認識ずれ.

確実性に対応する属性推薦手法が必要である。私たちはデテクターが推定する属性の確信度を活用するために、確信度スコアをソフトラベルとして用いた、ランキングに基づく属性推薦手法を提案した。また、ユーザ入力の不確実性及びユーザとデテクターの認識ずれに対応するために、スコアのスムージングのメカニズムとそれを用いたランキング手法を提案した。提案手法の評価について、ユーザに属性を推薦して回答してもらい、対象人物が見つかるか否かを判定するプロセスを繰り返し、 k 問までに対象人物が見つかる成功率、 k 問までの MRR (Mean Reciprocal Rank), 対象人物を探し出すまでに必要な質問数によって評価する方法を提案した。PA100K [1] データセットを使用した実験により、提案した手法が他の手法よりも優れていることを明らかにした。

2 関連研究

2.1 人物画像検索

人物画像検索について、多くの既存研究は属性の推定やクエリと画像の類似度に注目している [16], [1]. 一部の既存研究は問合せ画像または自然言語のクエリで検索する人物画像検索を研究している [9], [10], [11], [12], [17]. [9] はクエリ画像とギャラリー画像それぞれについて推定された属性ベクトルのノルムを考慮にいた人物画像検索手法を提案した。[10] はマルチスケールな人物認識及び検索手法を提案した。

属性を用いた人物画像検索についても、大きく注目されている。[13] は GAN を人物画像検索に応用する手法を提案した。[14] は属性と画像の階層マッチングモデルを提案し、人物画像検索への応用を示した。[15] は属性推定と人物認識を同時に行うための新しいロス関数を提案し、それに基づいて人物画像検索手法を提案した。[8] は問合せ属性の不確実性と画像から推定された属性の不確実性を考慮した人物画像検索手法を提案した。

表 1: 符号定義

符号	説明
G	すべての人物画像を含むギャラリー
Q	すべての属性の集合
Q_d	同じ種類の属性からなる Q の部分集合
g_i	ギャラリーに属する i 番目の人物画像
a_i	g_i の真の属性
\hat{a}_i	推定された g_i の属性の確信度スコア
\tilde{a}_i	ユーザが知っている g_i の属性
q_j	エージェントからの j 番目の属性に関する質問
π	各人物画像が対象人物になる確率
s_t	t 回の質問をした時点でユーザから各属性に関する質問の 回答
$r_{i,t}$	g_i が t 回の質問をした時点でのランキング
k	毎回表示される人物画像数
m	最大質問数

2.2 Learning-to-ask

Learning-to-ask は、数多くの候補が存在している集合の中から、ユーザの好みに合っているものを素早く見つけるために、ユーザが答えるべき項目を推薦することに関する研究である。プロダクトサーチの領域では、答えるべき属性を適切に選ぶことがすでに研究されてきた。[6] はインタラクティブなプロダクトサーチで最小数のクエリを用いてターゲットを検索するために、Sequential Bayesian Search (SBS) という手法を提案した。[7] は SBS を拡張し、ユーザからの回答に基づき、最小数のクエリを用いて関連ドキュメントを検索するアルゴリズムを提案した。[2] はユーザ回答の不確実性を考慮に入れ、不確実な回答に対応するプロダクトサーチアルゴリズムを提案した。[4] は、Learning to Ask のフレームワークを提示し、情報獲得とターゲット検索をバランスした質問選択ポリシーを提案した。[5] は強化学習手法を Learning to Ask のポリシー学習に応用し、質問エージェントが継続的な対話を通じて質問選択の最適なポリシーを学習できるシステムを提案した。

しかしながら、人物画像検索の領域では「Learning-to-ask」の応用についてほとんど研究されていない、本研究はその空白を補填することを目的の一つとする。

3 提案手法

3.1 問題定義

本論文では、数多くの人物画像が存在しているギャラリーからユーザが一枚の対象人物画像を検索することを想定する。エージェントは検索を補助するために、ユーザに答えるべき属性を推薦し、ユーザの回答に基づいて検索結果を提示する。

ギャラリー G には、複数枚の人物写真 ($g_i, 1 \leq i \leq |G|$) が存在している。人物画像 g_i は、デテクターにより各属性が推定され、 $|Q|$ 次元の確信度スコアのベクトル \hat{a}_i が付与される。ユーザは対象人物 g_i を知っており、ユーザが認識する属性情報を $|Q|$ 次元のベクトル \tilde{a}_i とする。一方、人物画像 g_i の真の属性が $|Q|$ 次元のベクトル a_i で表されるとする。ここで、 $a_i, \hat{a}_i, \tilde{a}_i$ は必

ずしも一致しない。ギャラリー G に属する $|G|$ 枚の人物画像の属性確信度スコアをまとめる $|G| \times |Q|$ の行列を G_{mat} とする。ユーザはギャラリー内の一枚の人物画像を順次対象人物として検索を行う。各人物画像が対象人物になる確率を π とする。

また、人物画像の属性について、人物の年齢、性別、服の色など、グループに分けることができる。例を挙げると、年齢に関する属性の部分集合 Q_1 は、「15歳以下」、「15~30歳」、「30~45歳」、「45~60歳」と「60歳以上」の五つの属性を含んでいる。符号の説明は表 1 にまとめる。

人物検索は以下の流れで行う。

- (1) エージェントは「対象人物の年齢は 15~30 歳ですか？」などの一つの属性について回答することを推薦し、ユーザは自分が認識する対象人物の情報に基づき、「YES」または「NO」を解答する。
- (2) エージェントはユーザが入力した回答を用い、ギャラリーにある人物画像をランキングし、上位の k 枚の人物画像をユーザに提示する。
- (3) 対象人物が提示されるまたは最大質問数 m に達すると、人物検索が終了する。そうでなければ、(1) に戻り、エージェントは次の属性を推薦する。

t 回のやりとりが終了した時点で、ユーザから得られた各属性に関する回答を $|Q|$ 次元のベクトル s_t は以下の式で表される。

$$s_t \in \{-1, 0, 1\}^{|Q|} \quad (1)$$

例を挙げると、1 番目の属性を聞いたことがない場合、 s_t の 1 位の値を 0 にする。1 番目の属性を聞き、ユーザから「YES」と答えると、 s_t の 1 位の値を 1 にする。1 番目の属性を聞き、ユーザから「NO」と答えると、 s_t の 1 位の値を -1 にする。

t 回のやりとりが終了した時点で、 s_t に基づき、エージェントはランキングの結果を表示し、次の推薦属性の選択をする。

3.2 問題点

既存手法をそのまま人物画像検索に応用する場合、以下の二つの問題がある。

- (1) ディテクターの認識で得られる確信度を使わず、属性の「YES/NO」だけを用いて選択される属性は必ずしも最適な属性ではない。
- (2) ユーザは対象人物の属性を誤認識、誤記憶する可能性がある。ディテクターの認識とずれている可能性がある。

ディテクターで推定された属性情報は不確実性を含んでおり、それは属性の確信度スコアで表示される。前述の通り、多くの従来手法では閾値処理によって属性の「YES/NO」だけを使っている。従って、確信度スコアが閾値付近にある真の属性を誤って認識する可能性がある。例を挙げると、対象人物の真の年齢が「15~30歳」であるが、光などの環境条件や隠れの有無など原因により「15~30歳」の確信度スコアが低く閾値を下回った

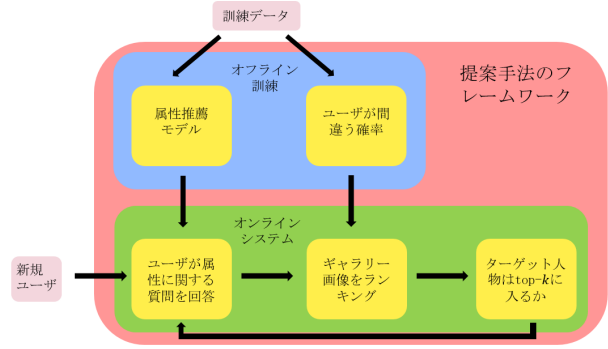


図 2: 提案手法のフレームワーク。

場合、ディテクターは「15~30歳ではない」と誤認識する。その状態でユーザがその人物を検索する際に、「15~30歳」の属性を指定すると、対象人物が結果から排除される。これに対し、閾値処理に基づく属性の「YES/NO」ラベルではなく、確信度スコアをソフトラベルとして用いることで、ユーザが「15~30歳」を指定したとしても対象人物は排除されず、他の指定された属性と合わせた統合スコアのランキングによって対象人物を見つけられる可能性がある。

また、ユーザは対象人物の属性について回答する際に、必ずしも正しく答えるわけではない。これは、ユーザが対象人物の一部の特徴を忘れていたり、そもそも間違っていて認識している可能性があるからである。例を挙げると、対象人物の真の年齢が「30~45歳」であるにも関わらず、ユーザがこの人物の年齢を 20 代後半くらいと誤認識した場合、「年齢は 15~30 歳ですか？」の質問に対し、「YES」と答えてしまう。すると、誤情報がランキングシステムに流され、対象人物は結果から排除される恐れがある。私たちはこれらの問題を解決するために、属性推薦モデルとスムージングメカニズムを提案した。提案手法のフレームワークは図 2 のように示されている。

3.3 属性推薦モデル

3.3.1 ランキングに基づく属性推薦

問題 1 を解決するために、私たちはランキングに基づく属性推薦手法を提案した。前述の通り、一枚の人物画像 g_i は、各属性の確信度スコアの $|Q|$ 次元ベクトル \hat{a}_i を持つ。ユーザの回答が s_t となる場合、確信度スコアを用いて各人物画像のランキングは以下ようになる。

$$ranking = \text{argsort } G_{mat} \cdot s_t^T \quad (2)$$

人物画像 g_i が t 回の属性推薦に回答した時点でのランキングを $r_{i,t}$ とする。

g_i が対象人物である場合、次の属性について回答を得ることで、 g_i のランキングがより上位に上がることが望ましい。言い換えると、次に推薦すべき属性は $r_{i,t} - r_{i,t+1}$ を最大化する属性である。すなわち、すべての人物画像が対象人物になる確率 π を考慮に入れた上で、最も望ましい属性を推薦するモデルを得ること

Algorithm 1 ランキングに基づく属性推薦手法

- 1: ギャラリー G , 属性集合 Q , すでに聞かれた属性問題 Q' , ユーザの回答 s_t をインプットする
- 2: 各人物画像現在のランキング $R = (r_{1,t}, r_{2,t}, \dots, r_{|G|,t})$ を計算する
- 3: **for** q in $Q \cap (Q')^C$ **do**
- 4: 属性 q を聞く場合のランキング $R' = (r'_{1,t}, r'_{2,t}, \dots, r'_{|G|,t})$ を計算する
- 5: 属性 q を聞くことで得られるリワード $(R - R') \cdot \pi$ を計算する
- 6: **end for**
- 7: リワードが最も高くなる属性を推薦する

Algorithm 2 強化学習による近似

- 1: 環境初期化
- 2: Replay Buffer を初期化
- 3: 二つの Dueling DQN ネットワーク Q^A, Q^B を初期化
- 4: **for** g in $|G_{train}|$ **do**
- 5: **for** t in $1 : m$ **do**
- 6: **if** $Unif(0, 1) < \epsilon$ **then**
- 7: ランダムな質問を選択する
- 8: **else**
- 9: 現在のポリシー Q^A と Q^B により属性を選択する
- 10: **end if**
- 11: ユーザに属性を提示し, 現時点の状態 s_t と次の状態 s_{t+1} を観測する
- 12: 選択された属性 q の回答を得ることで得られるリワード R を計算する
- 13: トランジション (s_t, q, s_{t+1}, R) を Replay Buffer に保存する
- 14: Replay Buffer からランダムに一部のトランジションを抽出する
- 15: 抽出されたトランジションを用い, Q^A または Q^B をアップデートする
- 16: **end for**
- 17: **end for**

が目的である. 本モデルを得るための疑似コードは Algorithm 1 で示されている.

3.3.2 強化学習による近似

残念ながら, 前述のアルゴリズムはランキングの計算を繰り返すため, 計算コストが高い. 特に, ギャラリーに数多くの人物画像が存在している場合は計算コストが膨大になる. そこで, 私たちは強化学習モデル Double Dueling DQN [21], [22] を用い, 上記のランキングに基づく属性推薦手法のポリシーを学習させ, 最適なポリシーの近似解を得た. 疑似コードは Algorithm 2 で示されている.

3.4 スムージングメカニズム

ユーザの回答の不確実性及びユーザとディテクターの認識ずれに対応するために, 私たちはスムージングメカニズムを提案した.

ユーザが対象人物の属性を間違っ て入力する場合, またユーザとディテクターの認識がずれている場合, 表 2 に示す問題が

表 2: ユーザが対象人物の属性を間違っ て入力する, またはユーザとディテクターの認識がずれている場合

	ユーザで「YES」	ユーザで「NO」
ディテクターで「YES」	問題なし	対象人物を排除
ディテクターで「NO」	対象人物を排除	問題なし

生じる.

そこで, 対象人物を排除しないために, 「年齢は 15~30 歳ですか?」の質問に対し, ユーザが「YES」と答えても, 年齢グループにある他の属性, 「15 歳以下」, 「30~45 歳」, 「45~60 歳」と「60 歳以上」にもある程度確信度スコアを与える. 言い換えると, ランキングの計算は,

$$ranking = \text{argsort } G_{mat} \cdot s_t^T$$

ではなく, s_t の確信度スコアを再配分した s'_t に入れ替え,

$$ranking = \text{argsort } G_{mat} \cdot (s'_t)^T$$

で計算される. 確信度スコアの再配分については, 学習データを用いてあらかじめ決められる. 例を挙げると, 学習データでユーザが「15~30 歳」と思う人物が 100 人おり, その 100 人に対するディテクターの認識では「15 歳以下」が 12 人, 「15~30 歳」が 60 人, 「30~45 歳」が 20 人, 「45~60 歳」が 8 人, 「60 歳以上」が 0 人であったとする. この場合, ユーザが「年齢は 15~30 歳ですか?」の質問に対し, 「YES」と答える場合, ランキングを計算する時は $s_t = (0, 1, 0, 0, 0)$ ではなく, $s'_t = (0.12, 0.6, 0.2, 0.08, 0)$ のように確信度を再配分したものをを用いる.

4 実 験

以下の 4 つ観点で我々の手法 (CSBAS-SRM) を評価した. **Q1.** 20 の質問以内でターゲットとなる人物を見つけることができるか **Q2.** ターゲットとなる人物を見つけるために必要な質問数が少なくなるか **Q3.** 20 の質問以内にターゲットとなる人物を見つけることができなかった場合でも, ターゲットとなる人物のランキングがより高くなるか **Q4.** 提案されたモデルの公平性が高いか (ターゲット画像毎に必要な質問数のばらつきが小さいか) **Q5.** 属性推薦手法 (CSBAS) とスコア再配分手法 (SRM) はそれぞれが機能しているか **Q6.** ユーザーに表示される画像の数 k が, 上記の指標に影響するか

4.1 データセットと実験設定

4.1.1 データセット

私たちは PA100K [1] からランダムに 1000 枚の写真を選出し, より詳細な属性認識を行うため, 手動作業で属性ラベルを付けた.

4.1.2 実験設定

前述のように, ユーザのご記憶やご回答, ディテクターの誤認識などが原因でユーザの回答とディテクターの認識にずれが生じる場合がある. そこで我々は, 以下 4 つの状況を想定して提案手法の評価を行った.

表 3: 成功率の評価結果. シンボル (†) は既存手法に対して顕著な効果があることを示す.

	Case 1	Case 2	Case 3	Case 4
CSBAS-SRM	76.0%†	34.3%†	31.2%†	18.5%†
SBS	63.7%	28.9%	26.1%	15.7%
QSBPS	64.8%	28.4%	26.8%	15.1%
SBSTAR	64.7%	27.6%	26.0%	14.9%

- **Case 1.** ユーザとディテクターの認識にずれがない場合.
- **Case 2.** ユーザが 30% の確率で誤回答をする場合.
- **Case 3.** ディテクターが誤認識をする場合. これは、ユーザは人手で正解付けされた属性に基づいて回答するが、検索時のスコア計算はディテクターの推定結果に基づいて行うこととした.
- **Case 4.** Case3 のようにディテクターの誤認識が含まれ、かつユーザが 30% の確率で誤回答をする場合.

4.1.3 シミュレーション設定

ランダムに 200 枚の画像を抽出し、それぞれをターゲットとした検索を実施した. すべてにおいてユーザは前述 3.1 の通りエージェントとのやり取りに基づいて検索条件を設定していくものとし、提示される画像の枚数 k は 1 とした.

4.1.4 評価指標

エージェントとの 20 回のやり取りがないでターゲット画像が見つかる成功率、ターゲット画像を見つけるまでのやり取り数と、20 回のやり取り後の MRR を評価した. さらに、検索の公平性を評価するため、Gini 係数を算出して評価した.

4.1.5 Baselines

既存手法との差を評価するため、SBS [6], QSBPS [2] and SBSTAR [7] の 3 つの手法との比較を行った. 評価実験において、それぞれの手法で設定するパラメータ等は、個々の論文で言及されているパラメータを用いた.

4.2 Q1: 20 回のやり取りでの成功率

まず、20 回のエージェントとユーザのやり取りによってターゲット人物が見つかる成功率を評価した. 表 3 に結果が示されており、提案手法はすべての case において既存手法を大きく上回る成功率を実証した. Case1 のようにユーザとディテクターの認識にずれがない状況において既存手法を上回っていることから、ソフトラベルを用いたランキングに基づく属性推薦手法が有効だと考えた. また、Case2 4 のようにユーザの誤回答やディテクターの誤認識がある場合、すべての手法で成功率は低下する傾向にあったものの、我々の提案手法が最も優れる成功率を示し、認識ずれに対応するためのスコア再配分手法が有効であったことを示唆している. ソフトラベルを用いた属性推薦手法 (CSBAS) とスコア再配分手法 (SRM) の効果切り分けについては 4.6 にて言及する.

表 4: MRR 評価結果. シンボル (†) は既存手法に対して顕著な効果があることを示す.

	Case 1	Case 2	Case 3	Case 4
CSBAS-SRM	0.826†	0.443†	0.429†	0.270†
SBS	0.764	0.411	0.383	0.245
QSBPS	0.725	0.386	0.358	0.185
SBSTAR	0.744	0.377	0.364	0.193

4.3 Q2: ターゲットが見つかるまでのやり取り数

ターゲットが見つかるまでのやり取り数を評価するために、各回数のやり取り時点までにおける累積成功率を評価した.

図 4 に示した通り、提案手法は Case1 のほとんどのやり取り数において既存手法よりも優れた累積成功率を有しており、提案手法が適切な属性を推薦することで、より少ない属性入力でターゲット人物をいち早く発見できていることを実証した. 一方 case2 4 のようにユーザやディテクターの認識にずれがある場合では、やり取り数が少ない時点において既存手法の方が優れるが、やり取り数が大きくなるにつれて提案手法の方が優れる結果となった. これは、スコアの再配分手法 (SRM) が認識ずれを補う効果があり、やり取り数が増えて検索に用いる属性数が増える程効果を発揮する傾向があることを示している.

4.4 Q3: Mean Reciprocal Rank

20 回のやり取りを終えてもターゲット画像が見つからない場合があるが、この時でもターゲット画像がより上位に位置していればユーザはより少ない結果確認でターゲットを見つけることができる. そこで、20 回のやり取りでターゲットが見つからなかった場合の MRR を評価した. 表 4 に図示した通り、提案手法はすべての case において最も良い MRR を示し、既存手法よりも優れていることを実証した.

4.5 Q4: 公平性

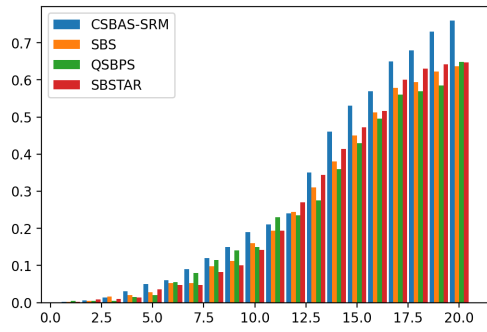
前述 4.3 の通り、case2 から 4 のようにユーザとディテクターの認識にずれがある場合、少ないやり取り数の時点では既存手法の方が成功率が高い場合も見られた. これは少ないやり取りではたまたま認識ずれを含む回答が得られず、ターゲットが早く見つかる場合があることが影響している可能性がある. そこで我々は、MRR の分散 (Gini 係数) を評価した. Gini 係数が低く公平性が高い (=MRR の分散が小さい) ほど、たまたまに影響されない良い検索システムであると言える.

一例として MRR における Gini 係数は以下の式で算出した. 表 5 に各 case における各手法のやり取り数と MRR の Gini 係数を示す.

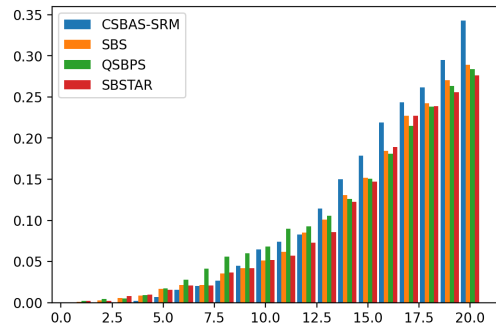
$$\text{Gini-MRR} = \frac{\sum_{i=1}^n \sum_{j=1}^n |MRR_i - MRR_j|}{2n \sum_{i=1}^n MRR_i}. \quad (3)$$

4.6 Q5: Ablation Study

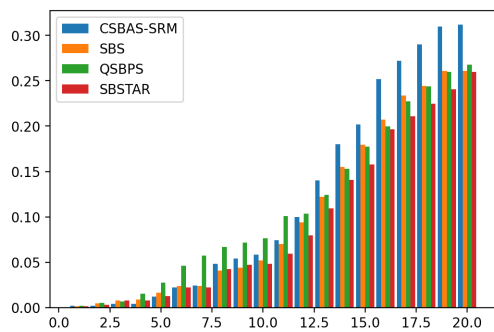
ディテクターの不確実性によってギャラリー画像のラベルが不適切になる問題に対処するためのソフトラベルを用いた属性推薦手法 (CSBAS) と、ユーザ回答とディテクターの認識のず



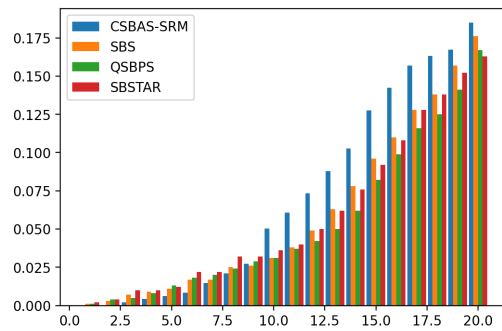
(a) Case 1: ユーザとディテクターに認識ずれがない場合.



(b) Case 2: ユーザが一定確率で誤回答する場合.



(c) Case 3: ディテクターが誤認識をする場合.



(d) Case 4: ディテクターの誤認識とユーザの誤回答の双方が含まれる場合.

図 3: t 回目のやり取り後時点における累積成功率.

表 5: 公平性評価結果. シンボル (\dagger) は既存手法に対して顕著な効果があることを示す.

		CSBAS-SRM	SBS	QSBPS	SBSTAR
Case 1	Gini-MRR	0.197\dagger	0.217	0.251	0.223
	Gini-#やり取り数	0.309	0.309	0.294	0.31
Case 2	Gini-MRR	0.557	0.547	0.569	0.542
	Gini-#やり取り数	0.189\dagger	0.262	0.225	0.262
Case 3	Gini-MRR	0.519	0.529	0.576	0.522
	Gini-#やり取り数	0.239	0.241	0.254	0.239
Case 4	Gini-MRR	0.657	0.667	0.73	0.664
	Gini-#やり取り数	0.122	0.234	0.124	0.232

れに対応するためのスコア再配分手法 (SRM) それぞれの効果を評価するため、Ablation study を行った. 前章同様成功率と MRR を評価した. 表 6 に示す通り、case 2-4 では SRM によって成功率も MRR も改善する傾向が見られ、期待通りスコア再配分によってずれの影響が弱められていることが実証できた.

表 6: 属性推薦とスコア再配分の切り分け評価.

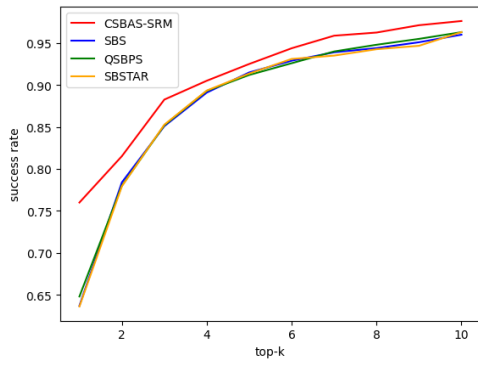
		Case 1	Case 2	Case 3	Case 4
Success rate	CSBAS-SRM	76.0%	34.3%	31.2%	18.5%
	w/o SRM	76.0%	32.1%	29.3%	15.2%
MRR	CSBAS-SRM	0.826	0.443	0.429	0.270
	w/o SRM	0.826	0.427	0.412	0.238

4.7 Q6: 画像表示数の影響

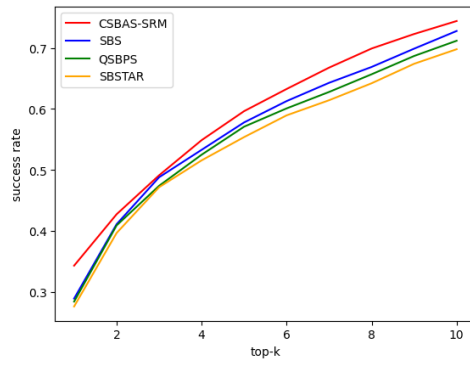
最後に、ユーザに各やり取り後に提示する画像数 k が検索性能に与える影響を評価した. 実際のユースケースではユーザは検索結果上位を目的に応じた数だけ閲覧してターゲットか否かの判定を行うと想定される. そこで、 k を 1 から 10 まで変化させた場合の成功率を評価した. 図 4 に示す通り、すべての手法で画像表示数が増える程成功率が向上した. また、すべての場合において提案手法が最も優れた成功率を示した.

5 まとめと今後の課題

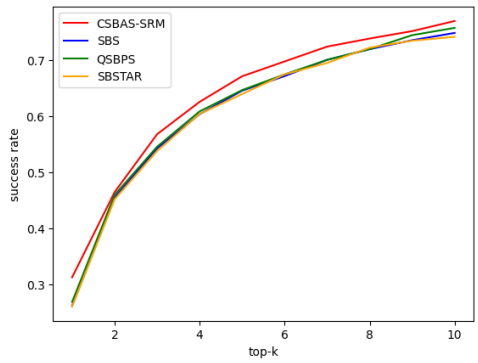
本研究では、ユーザが入力する検索クエリの属性条件と、ディテクターが推定する属性の両方が不確実性を持つ条件の下で数



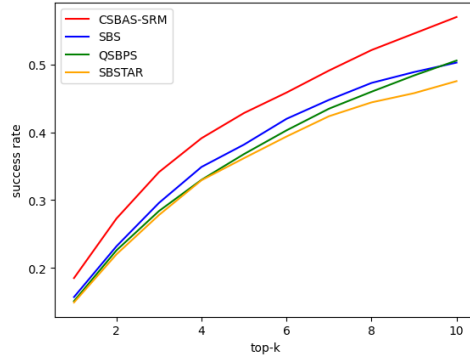
(a) Case 1: ユーザとディテクターに認識ずれがない場合.



(b) Case 2: ユーザが一定確率で誤回答する場合.



(c) Case 3: ディテクターが誤認識をする場合.



(d) Case 4: ディテクターの誤認識とユーザの誤回答の両方が含まれる場合.

図 4: 画像表示数 k と成功率の関係.

多くの人物画像から対象人物を早く検索するために、属性のソフトラベルを用いたランキングに基づく属性推薦手法と、スコア再配分手法を提案した。ユーザとディテクターの両方が正しい場合、ユーザが一定の確率で間違える場合、ディテクターとユーザの認識がずれている場合、及びディテクターとユーザの認識はずれている上でユーザが一定の確率で間違える場合、四つの状況をシミュレートして実験を行い、本手法の有用性を実証した。今後は、スムージングメカニズムを拡張し、ユーザが間違える場合及び認識がずれている場合だけでなく、両方も正しい場合でもスコアの再分配により属性推薦を改善していく。また、大規模なギャラリーに応用する場合の効率性を改善していく。

文 献

- [1] X. Liu, H. Zhao, M. Tian, L. Sheng, J. Shao, S. Yi, J. Yan, and X. Wang, "Hydraplus-net: Attentive deep features for pedestrian analysis," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 350–359.
- [2] J. Zou and E. Kanoulas, "Learning to ask: Question-based sequential bayesian product search," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 369–378.
- [3] J. Zou, Y. Chen, and E. Kanoulas, "Towards question-based recommender systems," in *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, 2020, pp. 881–890.
- [4] Y. Chen, B. Chen, X. Duan, J.-G. Lou, Y. Wang, W. Zhu, and Y. Cao, "Learning-to-ask: Knowledge acquisition via 20 questions," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2018, pp. 1216–1225.
- [5] H. Hu, X. Wu, B. Luo, C. Tao, C. Xu, W. Wu, and Z. Chen, "Playing 20 question game with policy-based reinforcement learning," *arXiv preprint arXiv:1808.07645*, 2018.
- [6] Z. Wen, B. Kveton, B. Eriksson, and S. Bhamidipati, "Sequential bayesian search," in *International Conference on Machine Learning*. PMLR, 2013, pp. 226–234.
- [7] J. Zou, D. Li, and E. Kanoulas, "Technology assisted reviews: Finding the last few relevant documents by asking yes/no questions to reviewers," in *The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2018, pp. 949–952.
- [8] T. Dong and J. Liu, "Person search by uncertain attributes," in *Proceedings of the 2022 International Conference on Multimedia Retrieval*, 2022, pp. 215–218.

- [9] D. Chen, S. Zhang, J. Yang, and B. Schiele, “Norm-aware embedding for efficient person search,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 12 615–12 624.
- [10] X. Lan, X. Zhu, and S. Gong, “Person search by multi-scale matching,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 536–552.
- [11] B. Munjal, S. Amin, F. Tombari, and F. Galasso, “Query-guided end-to-end person search,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 811–820.
- [12] Y. Yan, J. Li, J. Qin, S. Bai, S. Liao, L. Liu, F. Zhu, and L. Shao, “Anchor-free person search,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7690–7699.
- [13] Y.-T. Cao, J. Wang, and D. Tao, “Symbiotic adversarial learning for attribute-based person search,” in *European Conference on Computer Vision*. Springer, 2020, pp. 230–247.
- [14] Q. Dong, S. Gong, and X. Zhu, “Person search by text attribute query as zero-shot learning,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3652–3661.
- [15] B. Jeong, J. Park, and S. Kwak, “Asmr: Learning attribute-based person search with adaptive semantic margin regularizer,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 016–12 025.
- [16] D. Li, X. Chen, and K. Huang, “Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios,” in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*. IEEE, 2015, pp. 111–115.
- [17] S. Li, T. Xiao, H. Li, B. Zhou, D. Yue, and X. Wang, “Person search with natural language description,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1970–1979.
- [18] Z. Wang, R. Hu, C. Liang, Y. Yu, J. Jiang, M. Ye, J. Chen, and Q. Leng, “Zero-shot person re-identification via cross-view consistency,” *IEEE Transactions on Multimedia*, vol. 18, no. 2, pp. 260–272, 2015.
- [19] Y. Wang, J. Shen, S. Petridis, and M. Pantic, “A real-time and unsupervised face re-identification system for human-robot interaction,” *Pattern Recognition Letters*, vol. 128, pp. 559–568, 2019.
- [20] H. Wang, S. Gong, X. Zhu, and T. Xiang, “Human-in-the-loop person re-identification,” in *European conference on computer vision*. Springer, 2016, pp. 405–422.
- [21] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, “Dueling network architectures for deep reinforcement learning,” in *International conference on machine learning*. PMLR, 2016, pp. 1995–2003.
- [22] H. Van Hasselt, A. Guez, and D. Silver, “Deep reinforcement learning with double q-learning,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 30, no. 1, 2016.