

# 洋服の偏りと視覚的非類似性を考慮した 最適カプセルワードローブ構築の拡張

田中 勇太<sup>1,a)</sup> 尾崎 知伸<sup>1,b)</sup>

**概要:** 着回しを考えた最低限の洋服集合をカプセルワードローブ (CW) と呼ぶ。与えられた洋服群からの CW の構築は、組み合わせ最適化問題として定式化でき、これまでに劣モジュラ最適化を通じて、トピックモデルに基づく互換性と多様性に関して (近似) 最適な値を持つ CW を構築する手法が提案されている。本研究では、既存手法における評価関数を拡張し、互換性・多様性に加え、視覚的非類似性と洋服の偏りをも考慮した CW の構築を目指す。

## Optimal capsule wardrobe construction by considering visual dissimilarity and balance of use

TANAKA YUTA<sup>1,a)</sup> OZAKI TOMONOBU<sup>1,b)</sup>

**Abstract:** A fashion capsule wardrobe (CW) is a minimum set of clothes that can be worn in different ways. The construction of CWs can be formulated as a combinatorial optimization problem in general, and a sophisticated method has been proposed to construct optimal CWs on compatibility and diversity based on topic modeling and submodular optimization. In this study, in order to improve the quality of CWs, we extend the evaluation function in the existing method by incorporating additional criteria considering visual dissimilarity and balance of use. The effectiveness of the proposed criteria is assessed quantitatively using RichWear Dataset.

### 1. はじめに

着回しを考えた必要最小限の洋服集合をカプセルワードローブ (Capsule Wardrobes, CW) と呼ぶ。CW の構築は、日々の洋服選択時間の短縮や、洋服購入費用の抑制につながるだけでなく、アパレル業界におけるブランド戦略や個人の自己表現のための有用な手段としても認識され、近年多くの注目を集めている。しかし、有用な CW の構築にはファッションに関する一定レベルの知識やセンスが必要とされ、ファッションに精通していない利用者にとって、CW を構築することは必ずしも容易ではない。これらのことを背景に、ファッション初心者の支援を目的とした CW 自動構築アルゴリズムがいくつか提案されている [1, 3, 5, 6].

本研究では、Hsiao らによって提案された CW 自動構築手法 [1] に着目する。詳細は後述するが、この手法は (1) CW から生成される各コーディネート (洋服の組み合わせとしての) 質を表す互換性 (Compatibility) と (2) CW から生成可能なコーディネートの多彩さを表す多様性 (versatility) の 2 つの観点から CW を評価するものであり、トピックモデルを用いた各観点の評価と劣モジュラ最適化を通じ、教師情報なしに最適な CW を構築する。最適 CW の評価基準である互換性・多様性はともに CW が満たすべき重要な側面を捉えているが、両者のみを CW の評価基準とした場合、例えば見た目が酷似する洋服が多数含まれるなど、他の観点に関して必ずしも良質な CW が構築されるとは限らない。そこで本研究では、CW 評価の新たな観点として、視覚的非類似性および利用洋服の偏りの 2 点を対象とし、それぞれに対する新たな評価基準を提案する。またこれらを通じ、より多様な側面に対する良質な CW の自動

<sup>1</sup> 日本大学文理学部情報科学科  
〒156-8550 東京都世田谷区桜上水 3-25-40

a) nu.tanaka.yuta@gmail.com

b) ozaki.tomonobu@nihon-u.ac.jp

構築の実現を目指す。

本論文の構成は以下のとおりである。2章で、関連研究について言及する。3章で本研究の基礎となるCW構築アルゴリズム [1] を説明し、4章で洋服の偏りと視覚的非類似性を考慮した新たな評価基準を提案する。5章で評価実験を行い、最後に6章でまとめと今後の課題を示す。

## 2. 関連研究

これまでに、CWの自動構築に関するいくつかの研究が行われている [1,3,5,6]。本研究で対象とする [1] は、もっとも初期の研究の一つである。この手法では、画像処理技術を用いてコーディネート画像から各ファッションアイテム（洋服や靴など）を抽出するとともに、抽出画像に対する属性抽出を行うことで、コーディネートを文書（単語集合）化する。これにより、確率モデル（トピックモデル）を用いたファッションアイテム・コーディネートの扱いを実現している。

一方 [5] では、トップスとボトムスの集合を対象とし、グラフ構造を利用したCW構築を提案している。具体的には、各アイテムを頂点とし、妥当なコーディネートを構成できるトップス-ボトムス間に辺を付与することで、洋服集合を二部グラフに変換するとともに、その最適部分グラフ（に含まれるアイテム）をCWとして抽出する。その際、コスト（予算）を考慮することで、CW構築を予算制約付き多目的最適化問題として定式化している。なお、最適部分グラフの特定はNP完全な問題となるため、実際のCW抽出には、グラフ理論に基づく貪欲なヒューリスティックアルゴリズムを採用している。

加えて [6] では、CWの個人化について議論を行っている。アイテムの購入記録を用いてユーザの好みをモデル化するとともに、体型などを考慮した評価モデルを構築している。また [3] では、洋服の組み合わせの良し悪しを大域的特徴（例：色、形）だけでなく、局所的特徴（例：袖、ロゴ）の両方から推定することで、より高精度コーディネートの互換性推定を行ったうえで、CW構築を行った。

## 3. 互換性と多様性に基づく最適CWの構築

本章では文献 [1] に従い、互換性と多様性を基準とする最適CWの構築アルゴリズムを導入する。

トップスやボトムス、シューズ等の種別のことを層（レイヤ）と呼び、層  $i$  に関するアイテムの全体集合を  $A_i = \{s_i^0, \dots, s_i^{|A_i|-1}\}$  と表記する。層の総数を  $m$  とするとき、各層から一枚ずつアイテムを選択することで構成されるサイズ  $m$  の集合  $o = \{s_1^{c_1}, \dots, s_m^{c_m}\} (\forall i s_i^{c_i} \in A_i)$  をコーディネートと呼ぶ。このとき定義より、洋服の全体集合  $A = \bigcup_i^m A_i$  から、要素数が  $\prod_i |A_i|$  であるコーディネート集合  $\mathcal{Y} = A_1 \times \dots \times A_m$  が構築可能であることが

分かる。

文献 [1] では  $A = \bigcup_i^m A_i$  を入力とし、各層から丁度  $T$  個のアイテムを選択することでCWを構築する。すなわち、各層  $i$  において選択されたアイテムの集合を  $A_{iT} = \{s_i^{j_1}, \dots, s_i^{j_T}\} \subseteq A_i$ 、そこから構築されるコーディネートの全体集合を  $\mathbf{y} = A_{1T} \times \dots \times A_{mT} \subseteq \mathcal{Y}$  としたとき、互換性に関する評価基準  $C(\mathbf{y})$  と多様性に関する評価基準  $V(\mathbf{y})$  の和を最大化するコーディネート集合

$$\mathbf{y}^* = \operatorname{argmax}_{\mathbf{y} \subseteq \mathcal{Y}} (C(\mathbf{y}) + V(\mathbf{y}))$$

を生成するカプセルワードローブ  $A^* = \bigcup_i^m A_{iT}^*$  を特定する。

互換性  $C(\mathbf{y})$  と多様性  $V(\mathbf{y})$  は、コーディネートを文書、コーディネートを構成する各洋服が持つ属性を単語、ファッションスタイルをトピックと見做し、確率モデルの一つであるトピックモデル（Correlated Topic Models [2]）を用いて算出する。具体的には、互換性

$$C(\mathbf{y}) = \sum_{o_j \in \mathbf{y}} c(o_j) \quad \text{where } c(o_j) = p(o_j | \boldsymbol{\mu}, \boldsymbol{\sigma}, \beta)$$

は、 $\boldsymbol{\mu}$ ,  $\boldsymbol{\sigma}$ ,  $\beta$  をパラメータとするトピックモデルにおけるコーディネート  $o_j$  の尤度  $p(o_j | \boldsymbol{\mu}, \boldsymbol{\sigma}, \beta)$  の和として算出される。これは、よく見られるコーディネートは互換性・妥当性が高く、逆にあまり見られないコーディネートは質が低いという仮定によるものである。また多様性は、トピックモデルにおけるモデル数を  $K$  とし、

$$V(\mathbf{y}) = \sum_{i=1}^K \left( 1 - \prod_{o_j \in \mathbf{y}} (1 - p(z_i | o_j)) \right)$$

と算出する。ここで  $p(z_i \bmod o_j)$  はコーディネート  $o_j$  におけるスタイル  $z_i$  の確率である。式より、各スタイル  $z_i$  に対して大きな確率を持つコーディネート  $o_j$  が存在する場合に、高い多様性が与えられる。詳細は論文 [1] を参照されたい。

カプセルワードローブ  $A^*$  の選択は、互換性  $C(\mathbf{y})$  と多様性  $V(\mathbf{y})$  が、(1) コーディネート集合に関して劣モジュラ性を満たすこと、また (2) 他層の洋服集合  $A_{jT}$  ( $i \neq j$ ) を固定した場合に層  $i$  の洋服集合に関して劣モジュラ性を満たすことに着目し、層ごとの劣モジュラ最適化を繰り返すことで達成される。すなわち、EMアルゴリズムのように、「注目レイヤを変えながら、そのレイヤに関する最適化を行う」という操作を、前回のループでのスコアとの差が一定の閾値を下回るまで繰り返すことで、(劣モジュラ最適化という意味での) 近似解を求める。その疑似アルゴリズムを図 1 に示す。

## 4. 視覚的非類似性と利用洋服の偏りに基づくCWの評価

本研究では、より多様な側面に対して高品質なCWを構

### Algorithm 1

```

1:  $A_{iT} := \emptyset, \forall i$  # 選択するアイテムの初期化
2:  $\delta_{obj} := \epsilon + 1$  #  $\epsilon$  は収束のための許容度
3:  $pre_{obj} := 0$ 
4: while  $\delta_{obj} \geq \epsilon$  do
5:   for each layer  $i = 0, 1, \dots, (m-1)$  do
6:      $A_{iT} := \emptyset$ 
7:     for each time step  $t = 1, 2, \dots, T$  do
8:        $g = select\_item(A_{iT})$ 
9:       # 最大ゲインアイテムの選択
10:       $A_{iT} := A_{iT} \cup \{g\}$ 
11:     end for
12:   end for
13:    $cur_{obj} = calc\_closet\_score(A_{iT})$  # 現在のスコアの計算
14:    $\delta_{obj} = cur_{obj} - pre_{obj}$ 
15:    $pre_{obj} = cur_{obj}$ 
16: end while

```

図 1 互換性と多様性の最適化に基づく CW 構築アルゴリズム [1]

築するため、二つの新たな評価関数を導入する。一つ目の評価関数は、洋服に対する視覚的非類似性であり、見た目がよく似たコーディネートが多数生成されることを防ぐことを目的とする。また二つ目の評価関数は、使用洋服の偏りに関するものであり、洗濯などで特定の洋服が利用できなくなった場合に生成されるコーディネートの質が極端に落ちることを防ぐことを目的とする。以下それぞれについて説明する。

#### 4.1 視覚的非類似性

各層の洋服集合  $A_i$  から  $T$  枚ずつ洋服を選択することで得られるコーディネートの集合  $\mathbf{y} = A_{1T} \times \dots \times A_{mT} \subseteq \mathcal{Y}$  に対し、視覚的非類似性を以下のように定義する。

$$S(\mathbf{y}) = \sum_i Cov(A_{iT}) \quad \text{where}$$

$$Cov(A_{iT}) = \left| \left\{ s_i^j \in A_i \mid \sigma_S \leq \min_{s_i^x \in A_{iT}} dist(s_i^j, s_i^x) \right\} \right|$$

ここで  $dist(s_i^j, s_i^x)$  は 2 つのアイテム  $s_i^j, s_i^x$  間の視覚的非類似度を、 $\sigma_S \geq 0$  は利用者による閾値をそれぞれ表す。

$S(\mathbf{y})$  は、各層において選択された洋服集合  $A_{iT}$  が、当該する層の洋服集合  $A_i$  をどの程度（視覚的に）被覆しているかを評価する。言い換えれば、集合被覆の観点から間接的に非類似性を評価するものであり、 $A_{iT}$  が視覚的に多様なアイテムを含む場合、 $A_{iT}$  中の各要素（アイテム）が被覆する範囲がそれぞれ異なり、結果として  $Cov(A_{iT})$  および  $S(\mathbf{y})$  は大きな値をとる。

#### 4.2 使用洋服の偏り

使用洋服の偏りを以下のように定義する。

$$B(\mathbf{y}) = |\{s \in o_j \mid o_j \in \mathbf{y}, c(o_j) \geq \sigma_B\}|$$

表 1 最適化したクローゼットのスコア

	$w_c \cdot C$	$w_v \cdot V$	$w_s \cdot S$	$w_b \cdot B$
$C_{all}$	8.709	7.524	11.812	8.666
$C_{SB}$	11.414	7.152	12.319	10.666
$C_{CV}$	8.848	7.689	10.280	6.666
$C_{random}$	0.046	6.769	8.673	0.000

ここで  $c(o_j)$  はコーディネート  $o_j$  の互換性を、 $\sigma_B$  は利用者による閾値を表す。式が示す通り、 $B(\mathbf{y})$  は一定以上の互換性を持つコーディネート  $o_j$  の構築に利用される洋服の数を表す。適切な閾値  $\sigma_B$  を設定することで、（総和としては互換性に寄与するが）多数の凡庸なコーディネートしか生成できない洋服の排除が期待できる。

先述した通り、本研究では、より多様な側面に対して高品質な CW の構築を目的としている。このため、提案した 2 つの指標を独立に利用するのではなく、互換性と多様性も含めた形での最適化を行う。具体的には、重み  $w_x (x \in \{c, v, s, b\}, w_x \geq 0)$  を用いた各基準の重み付き和

$$w_c \cdot C(\mathbf{y}) + w_v \cdot V(\mathbf{y}) + w_s \cdot S(\mathbf{y}) + w_b \cdot B(\mathbf{y})$$

を最大化する CW を構築する。また各側面の重みを適切に調整することで、利用者の要求を反映した柔軟な CW 構築の実現が期待できる。

## 5. 評価実験

提案する評価基準を用いた CW 構築を評価するため、RichWear データセット [4] を対象とした評価実験を行った。

### 5.1 提案指標の効果

提案指標を評価する、すなわち

- (1) 提案指標が最適化に寄与していること、および
- (2) 提案指標を用いることでより高質な CW が構築されること

を確認するため、利用する指標の組み合わせを変えながら、以下の条件で CW を構築した。

- トップス・ボトムス・シューズの 3 層を対象とする ( $m = 3$ )
- RichWear データセットから、各層に対するアイテムをそれぞれ 400 準備する ( $|A_i| = 400$ )
- 各層から 4 つのアイテムを選択する ( $T = 4$ )

実験結果を表 1 にまとめる。表中において  $C_{all}$  は、すべての指標 ( $C(\mathbf{y}), V(\mathbf{y}), S(\mathbf{y}), B(\mathbf{y})$ ) を用いた場合を表す。また  $C_{SB}$  は指標  $S(\mathbf{y})$  と  $B(\mathbf{y})$  のみを、 $C_{CV}$  は指標  $C(\mathbf{y})$  と  $V(\mathbf{y})$  のみを用いた場合を表す。なお  $C_{CV}$  は既存手法（ベースライン）[1] に相当する。一方  $C_{all}$  は最適化を行わず、ランダムにアイテムを選択することで CW を構築した場合を表す。

実験結果より、4 つの指標を利用する  $C_{all}$  においても、

**表 2** 各クローゼットのロバスト性

指標	$C_{all}$	$C_{SB}$	$C_{CV}$
ロバスト性	36	48	24

各指標の最小値である  $w_v \cdot V = 7.524$  が合計スコア 36.711 の 20%程度を占めており、適切なパラメタ設定の基であればすべての指標がそれぞれ最適化に寄与していることが確認できる。

また  $C_{all}$  と  $C_{random}$  を比較すると、各評価値の差はそれぞれ 8.663, 0.754, 3.138, 8.666 であり、結果として合計スコアの差は 21.223 となっている。一方、 $C_{all}$  とベースラインである  $C_{CV}$  を比較すると、 $w_c \cdot C$  と  $w_v \cdot V$  に関してはそれぞれ差が  $-0.139$  と  $-0.165$  と若干の低下が見られるが、 $w_s \cdot S$  と  $w_b \cdot B$  に関してはそれぞれ差が 1.532 と 2 であり、全体としても大きく質が向上していることが分かる。このことより、互換性・多様性をほとんど犠牲にすることなく、視覚的非類似性と偏りを導入することに成功していることが確認できる。

## 5.2 CW のロバスト性

提案指標が多様な側面に対して高質な CW の構築に寄与することを確認するため、新たな観点として頑健性（ロバスト性）を下式で定義し、前節で構築された CW を評価した。

$$\text{score}(\mathbf{y}) = \min_{s \in o_j \in \mathbf{y}} (r(s))$$

$$\text{where } r(s) = |\{o_j \in \mathbf{y} \mid c(o_j) \geq \sigma_B \wedge s \notin o_j\}|$$

式  $\text{score}(\mathbf{y})$  では、CW に含まれる各アイテム  $s$  に対し、 $s$  を使わなくても構築することのできる互換性の高いコーディネート数  $r(s)$  を求めている。また、 $r(s)$  の最小値は、一つのアイテムが何らかの理由で利用できなくなった場合にも構築可能な良質なコーディネート数を表しており、アイテムの欠損という側面から CW の頑健性（ロバスト性）を評価している。

実験結果を表 2 に示す。結果から、ロバスト性スコアが昇順に  $C_{CV} < C_{all} < C_{SB}$  であることが確認できる。すなわち、偏りの指標を用いない最適化である  $C_{CV}$  に比べ、偏りの指標を用いる  $C_{all}$ ,  $C_{SB}$  の方がロバスト性が高く、さらに両者の間では、互換性基準を含まない  $C_{SB}$  が最も高い、ということになる。加えて、このスコアが低い手法ほど  $w_b \cdot B$  の値も大きくなっていることから、洋服の偏りの指標がロバスト性のある CW 構築に寄与していることが読み取れる。

## 5.3 指標間の相関

各指標がそれぞれ独立した観点から CW を評価していることを確認するため、ランダム選択を用いて 140,000 件の CW 群を構築し、指標間の相関計数を算出した算出され

**表 3** 各指標間の相関係数

	$w_v \cdot V$	$w_s \cdot S$	$w_b \cdot B$
$w_c \cdot C$	0.28	0.05	0.59
$w_v \cdot V$		0.05	0.09
$w_s \cdot S$			0.03

た相関係数を表 3 に示す。

結果より、 $w_s \cdot S$  は他の指標との相関が非常に小さく、独立した指標であることが確認できる。一方、 $w_b \cdot B$  は、 $w_c \cdot C$  との相関が 0.59 と高く、必ずしも独立しているとは言えない結果となった。これは、 $B(\mathbf{y})$  の計算に  $C(\mathbf{y})$  を用いているためであると考えられる。

## 6. まとめ

本研究では、既存の CW 構築アルゴリズム [1] に対して、視覚的非類似性と使用洋服の偏りに関する新たな指標を導入した。また実験を通じ、これらの指標を用いることで、多様な観点からの高質な CW 構築が実現されることを確認した。

今後の課題としては、互換性との強い相関が確認された偏りに関する指標の改善が挙げられる。また、評価指標が人の認識に沿った評価指標になっているかなどの定性的な評価をして、実際の社会への有効性を示す必要があると考えている。加えて、CW の実利用を前提とした新たな評価指標の考案と、大規模データを対象としたインタラクティブ利用を可能にする高速アルゴリズムの開発も大きな課題である。

## 参考文献

- [1] W.-L. Hsiao and K. Grauman : Creating Capsule Wardrobes from Fashion Images, *Proc. of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.
- [2] J. D. Lafferty and D. M. Blei : Correlated topic models, *Advances in Neural Information Processing Systems* 18, 2006.
- [3] H. Chen, Y. Lin, F. Wang, and H. Yang : Tops, Bottoms, and Shoes: Building Capsule Wardrobes via Cross-Attention Tensor Network, *Proc. of the 15th ACM Conference on Recommender Systems*, pp.453–462, 2021.
- [4] F.-H. Huang, H.-M. Lu, and Y.-W. Hsu : From Street Photos to Fashion Trends: Leveraging User-Provided Noisy Labels for Fashion Understanding, *IEEE Access*, 9:49189–49205, 2021.
- [5] S. Patil, D. Banerjee, and S. Sural : A Graph Theoretic Approach for Multi-Objective Budget Constrained Capsule Wardrobe Recommendation, *ACM Transactions on Information Systems*, 40(1):1–33, 2022.
- [6] X. Dong, X. Song, F. Feng, P. Jing, X.-S. Xu, and L. Nie : Personalized Capsule Wardrobe Creation with Garment and User Modeling, *Proc. of the 27th ACM International Conference on Multimedia*, pp.302–310, 2019.
- [7] A. Bardey, M. Booth, G. Heger, and J. Larsson : Finding yourself in your wardrobe: An exploratory study of lived experiences with a capsule wardrobe, *International Journal of Market Research*, 64(1):113–131, 2021.