

多様な指標を考慮したコーディネート系列の生成

Generation of fashion coordination sequences considering various factors

菊池 愛海¹ 尾崎 知伸^{1*}
Aimi Kikuchi¹ Tomonobu Ozaki¹

¹ 日本大学 文理学部

¹ College of Humanities and Sciences, Nihon University

Abstract: This study addresses the automatic generation of fashion coordination sequences. Specifically, by using techniques on word embedding, topic modeling, and submodular optimization, we first propose a framework for generating fashion capsule wardrobes (CWs) that reflect the user's preferences in addition to the quality of the outfits and the diversity of styles. We then perform the selection of a multi-day fashion coordination sequences from the CWs under some realistic constraints, considering continuity and the schedule of each day. The effectiveness of the proposed framework is assessed through evaluation experiments on real datasets.

1 はじめに

近年、洋服の多様化が進み、個人の嗜好や時間・場所・場面を考慮した洋服の着合わせ、すなわちコーディネートを楽しむことは身近なものとなっている。その際、理想的なコーディネートに必要な洋服をすべて購入、準備できるとは限らず、既に所持している洋服から工夫してコーディネートを組み立てることが一般的である。しかし、場面に応じた適切な洋服を選択するとともに、自身の好きなスタイルに寄せたコーディネートを行うためには、一定の知識や経験が必要とされ、必ずしも容易ではない。更に日常生活を考えると、直近の予定を考慮して複数日分のコーディネートを用意する必要もあり、問題を難しくしている。

これらの問題を解決するための一つの方向性として、カプセルワードローブ (Capsule Wardrobe, CW) が挙げられる。CW とは、着合せ (組み合わせ) の良い多様なコーディネート多数を多数組み合わせることのできる小数の洋服集合を意味し、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。またファッションに関する知識やセンスが不足している利用者の支援を目的に、いくつかの CW 自動構築アルゴリズムが提案されている [1, 2, 3, 4, 5]。これらの CW 自動構築アルゴリズムを利用することで、コーディネート準備の簡略化が見込まれるが、実際には自身の好みを CW へと

反映すること、および日々の予定を考慮しながら (洋服集合である) CW から具体的なコーディネート系列を考えることは必ずしも容易ではない。

これらのことを背景に、本研究では、トップス・ボトムス・シューズの3アイテムを対象とし、CW に着目したコーディネート系列の自動生成に取り組む。具体的には、単語埋め込み、トピックモデル、最適化の各技術を利用し、着合わせの良さとスタイルの多様性に加え、利用者の好みを反映した CW を生成する。さらに CW から、いくつかの現実的な制約のもと、連続性と各日の予定を考慮した複数日分のコーディネート系列を生成する技術を提案する。

本論文の構成は以下のとおりである。2章で先行研究について言及する。3章で提案手法の詳細を説明した後、4章で評価実験を行う。最後に5章でまとめを行い、今後の課題を述べる。

2 関連研究

これまでに、ファッションコーディネートに関して様々な研究が行われている。

例えば文献 [6] では、ファッションアイテムやカテゴリ、スタイルをそれぞれ実数ベクトル化することで、コーディネートスタイルに適したファッションアイテムを検索する手法を提案している。また文献 [7] では、ファッション画像とテキストを同一空間に埋め込むことで、テキストを用いたファッションアイテムの検索を実現している。

*連絡先: 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
E-mail: ozaki.tomonobu@nihon-u.ac.jp

一方、CWの自動構築に関しても、いくつかの研究が報告されている。もっとも初期の研究の一つである[1]では、画像処理技術を用いたファッションアイテムに対する属性抽出を通じ、コーディネートを書き(属性集合)化した上で、トピックモデル[8]を用いて算出される交換性と多様性を基準にCWを構築する。これに対し文献[2]では、交換性と多様性の他に視覚的類似性等をCW構築基準に加えることを、文献[3]では、大域的特徴(色、形など)に加えて局所的特徴(袖、ロゴなど)の双方を利用して交換性を推定することをそれぞれ提案している。加えて[5]では、購入記録を用いた嗜好のモデル化と体形等を考慮した評価モデルを構築することで、CWの個人化を行っている。また[4]では、トップスとボトムスを対象とした二部グラフに対する最適部分グラフ抽出問題として、CW構築を実現している。

詳細は後述するが、本研究では文献[6, 7]等と同様、ファッションアイテムの実ベクトル化を行う。またCW構築アルゴリズムとして[1]による手法を採用する。これらを組み合わせることで、利用者の嗜好を反映したCW構築を実現する。

3 コーディネート系列の生成

3.1 問題の定式化

本研究では、トップス・ボトムス・シューズの3アイテムを対象とし、CWに着目したコーディネート系列の自動生成に取り組む。以下では、対象とする問題の定式化を行う。

トップスの集合を T 、ボトムスの集合を B 、シューズの集合を S と表記し、各集合に含まれるファッションアイテム $o \in T \cup B \cup S$ を、そのアイテムに付与されたタグや説明から得られる単語の集合 $D_o = \{d_1, \dots, d_{|D_o|}\}$ と同一視する。加えて、単語の全体集合を $W = \bigcup_{o \in T \cup B \cup S} D_o$ と表記する。一方、コーディネート $c = \langle t, b, s \rangle$ を、各要素からそれぞれ一つのアイテム $t \in T, b \in B, s \in S$ を選択することで構成されるタプルで表現する。またトップス、ボトムス、シューズの各集合 $T' \subseteq T, B' \subseteq B, S' \subseteq S$ に対し、構築可能なコーディネートの全体集合を

$$C(T', B', S') = \{\langle t, b, s \rangle \mid t \in T', b \in B', s \in S'\}$$

と表記する。以上の準備のもと、以下に示す手順を経て、長さ N のコーディネート系列を抽出する。

1. 利用者によって与えられるスタイルを表すキーワード $w \in W$ に適合するトップス・ボトムス・シューズの集合 $T_w \subseteq T, B_w \subseteq B, S_w \subseteq S$ を特定する。

2. 文献[1]による手法を用い、適合ファッションアイテムの集合 T_w, B_w, S_w からカプセルワードローブ $CW_w = \langle T_w^{cw}, B_w^{cw}, S_w^{cw} \rangle$ ($T_w^{cw} \subseteq T_w, B_w^{cw} \subseteq B_w, S_w^{cw} \subseteq S_w$) を抽出する。

3. i 番目のコーディネート c_i に対する嗜好を表すキーワード $w_i \in W$ を与え、いくつかの現実的な制約のもと、後述する評価関数を最大化する長さ N のコーディネート系列

$$CS_w^{\{w_1, \dots, w_N\}} = [c_1, \dots, c_N] \\ (c_i \in C(T_w^{cw}, B_w^{cw}, S_w^{cw}))$$

を生成する。

以下、それぞれの手順について詳細を説明する。

3.2 適合ファッションアイテムの抽出

スタイルを表すキーワード $w \in W$ に適合するファッションアイテムを特定するために、単語埋め込み技術を利用する。

まず、各アイテム o に対する BOW 表現をトランザクションとするデータベース $\{D_o \mid o \in T \cup B \cup S\}$ を対象に単語埋め込み技術 word2vec[9, 10] を適用することで、各単語 $w \in W$ のベクトル表現 \vec{w} を獲得する。次いで、得られたベクトル表現を用い、構成する単語の平均ベクトルとしてアイテム o のベクトル表現 $\vec{o} = \frac{1}{|D_o|} \sum_{d \in D_o} \vec{d}$ を獲得する。

以上の準備の下、閾値 σ を利用者によって与えられる自然数とし、キーワード w に対するベクトル間のコサイン類似度 $\text{sim}(w, o) = \cos(\vec{w}, \vec{o})$ の各集合 $X \in \{T, B, S\}$ における上位 σ 件を、 w に適合するアイテムの集合 $X_w \subseteq X$ とする。

3.3 カプセルワードローブの構築

適合ファッションアイテム集合 X_w ($X \in \{T, B, S\}$) からのカプセルワードローブ $CW_w = \langle T_w^{cw}, B_w^{cw}, S_w^{cw} \rangle$ の生成には、文献[1]の手法を用いる。具体的には、構築するCWに含まれるアイテムの数 θ を指定し、後述する交換性に関する評価値 $C(\cdot)$ と多様性に関する評価値 $V(\cdot)$ に関し、その和が最大となる CW^* 、すなわち

$$CW^* = \arg \max_{\substack{CW = \langle T', B', S' \rangle, \\ T' \subseteq T_w, |T'| = \theta, \\ B' \subseteq B_w, |B'| = \theta, \\ S' \subseteq S_w, |S'| = \theta}} (C(CW) + V(CW))$$

を CW_w として獲得する。

CWに関する評価は、コーパスとして与えられるコーディネート集合 $C = \{\langle t_1, b_1, s_1 \rangle, \dots, \langle t_{|C|}, b_{|C|}, s_{|C|} \rangle\}$

を対象に、ファッションアイテムの種別を接頭辞とする単語から構成される文書の集合

$$\bigcup_{c=(t,b,s) \in C} bow_c \text{ where } bow_c = \bigcup_{x \in \{t,b,s\}} \{x+d \mid d \in D_x\}$$

(記号 + は単語の連結を行う演算子) を獲得し、トピックモデル [8] を適用すること得られる確率モデル P を用いる。

文書 (単語集合) bow の P における尤度 $c(bow)$ を用い、 $CW = \langle T', B', S' \rangle$ の互換性、すなわち着合せの良さに関する評価値は、生成可能なコーディネートの尤度の和、すなわち

$$C(CW) = \sum_{c \in C(T', B', S')} c(bow_c)$$

と定義される。一方、 P におけるトピック集合 G と文書 bow におけるトピック $g \in G$ の確率を $p(g \mid bow)$ を用い、 CW の多様性に関する評価値は、コーディネートによるトピックのカバー、すなわち

$$V(CW) = \sum_{g \in G} \left(1 - \prod_{c \in C(T', B', S')} (1 - p(g \mid bow_c)) \right)$$

と定義される。

3.4 最適コーディネート系列の特定

カプセルワードローブ CW_w からのコーディネート系列 $CS_w^{\{w_1, \dots, w_N\}} = [c_1, \dots, c_N]$ の生成には、以下の 4 評価関数の総和を用いる。

$$\begin{aligned} C(CS_w^{\{w_1, \dots, w_N\}}) &= \sum_{i=1}^N c(bow_{c_i}) \\ V(CS_w^{\{w_1, \dots, w_N\}}) &= \frac{1}{|G|} \sum_{g \in G} \left(1 - \prod_{i=1}^N (1 - p(g \mid bow_{c_i})) \right) \\ D(CS_w^{\{w_1, \dots, w_N\}}) &= \sum_{i=1}^{N-1} dist(\vec{c}_i, \vec{c}_{i+1}) \\ M(CS_w^{\{w_1, \dots, w_N\}}) &= \sum_{i=1}^N dist(\vec{c}_i, \vec{w}_i)^{-1} \end{aligned}$$

評価関数 $C(\cdot)$ は、コーパスから構築された確率モデル P における尤度を用いて、系列に含まれるコーディネートの着合わせの良さ (互換性) を評価する。また $V(\cdot)$ は、コーディネート集合の多様性を評価する。

一方、残り 2 つの評価関数 $D(\cdot)$ と $M(\cdot)$ は、適合ファッションアイテムの抽出に利用した単語やアイテムに対するベクトル空間を利用する。コーディネート $c = \langle t, b, s \rangle$ に対するベクトル表現を、構成要素のベクトルの平均値、すなわち $\vec{c} = \frac{1}{3} (\vec{t} + \vec{b} + \vec{s})$ とする。ま

た、2 つのベクトル \vec{v}_1, \vec{v}_2 間の距離を $dist(\vec{v}_1, \vec{v}_2)$ と表記する。これらを用い、連続性に関する評価関数 $D(\cdot)$ は、連続するコーディネート c_i と c_{i+1} の違いが大きい系列に対し高い評価を与える。一方、キーワードに対する適合性に関する評価関数 $M(\cdot)$ は、指定されたキーワード w_i とコーディネート c_i の距離が近い系列を高評価する。

4 評価実験

提案したファッションコーディネート系列生成手法の有効性を評価するために、IQON3000[11] データセットを用いて評価実験を行った。実験では、765 のコーディネートを対象とし、トプス、ボトムス、シューズ以外のアイテムを削除するとともに、各アイテム o に対する説明文を形態素解析することで単語集合 D_o を獲得した。また、スタイルを表すキーワードを $w \in \{\text{カジュアル, かわいい}\}$ 、適合ファッションアイテム集合に関するパラメタを $\sigma = 10$ 、 CW に含めるアイテム数を $\theta = 5$ 、生成する系列長を $N = 4$ とした。加えて、各日に対する嗜好を $[w_1, w_2, w_3, w_4] = [\text{かわいい, 冬, 軽やか, フォーマル}]$ とするとともに、以下の制約の下でコーディネート系列の生成を行った。

制約 1 : すべての日でトプスが異なる

制約 2 : 3 種以上のボトムスを含む

制約 3 : 尤度が一定値未満のコーディネートを含まない

4.1 定量評価

実験では比較のため、合計評価値を最大化する系列に加え、各評価値 ($C(\cdot), V(\cdot), D(\cdot), M(\cdot)$) を最大化する系列およびいくつかの評価値の和を最大化する系列を特定した。実験結果を表 1 に示す。

表 1 より、 $w = \text{カジュアル}$ の場合、妥当性 (C) だけに着目してコーディネート系列を生成しても、他の基準での評価が低いものが生成されてしまうことが伺える。一方、多様性 (V) と連続性 (D) は関連が強いこともあり、多様性 (V) を最大化するコーディネート系列は、他の基準でも比較的高い評価が得られることが示された。また多様性を最大化するコーディネート系列と、全評価基準の総和を最大化するコーディネート系列とを比較した場合、特に適合性 (M) の評価に大きな差が認められ、複数の基準を同時に考慮することで、各評価においてバランスの良いコーディネート系列が生成されることが示唆された。

表 1: 実験結果：得られたコーデ系列の評価値

最大化基準	C	V	D	M	合計
<i>w</i> = カジュアル					
C	0.232	0.128	0.209	0.025	0.594
V	0.210	0.303	0.211	0.061	0.785
D	0.214	0.223	0.225	0.051	0.713
M	0.134	0.172	0.165	0.224	0.695
C + V	0.212	0.284	0.171	0.041	0.708
D + M	0.157	0.212	0.219	0.114	0.702
C + V + D	0.195	0.201	0.211	0.036	0.643
C + V + M	0.201	0.221	0.192	0.074	0.688
C + V + D + M	0.200	0.219	0.216	0.181	0.816
<i>w</i> = かわいい					
C	0.301	0.122	0.196	0.032	0.551
V	0.137	0.210	0.201	0.018	0.566
D	0.122	0.208	0.241	0.019	0.590
M	0.115	0.180	0.185	0.194	0.674
C + V	0.218	0.182	0.163	0.022	0.585
D + M	0.152	0.185	0.237	0.153	0.727
C + V + D	0.222	0.194	0.221	0.041	0.678
C + V + M	0.190	0.198	0.189	0.123	0.700
C + V + D + M	0.265	0.209	0.229	0.112	0.815

w = かわいいの場合、適合性 (*M*) に着目した最大化が総合評価の高いコーディネート系列の生成に寄与することが伺える。また、全評価基準の総和を最大化するコーディネート系列における各基準での評価値は、単一基準を最大化する系列の値には及ばないものの、それぞれ十分に高い値を示しており、欠点の少ないコーディネート系列が生成されていると考えられる。

4.2 定性評価

定性評価として、服装に気を遣う 20~30 代の女性 20 名に対してアンケート調査を行った。アンケートでは、本研究の方法で生成したコーディネート系列に対し、以下の 3 項目の質問を提示した。

質問 1：コーディネートはまともですか？(1：まともではない ~ 4：まともである)

質問 2：毎日異なる印象を受けますか？(1：受けない ~ 4：受ける)

質問 3：1 日目~4 日目でそれぞれどのような印象を受けますか。以下のスタイルを当てはめてください(かわいい, 冬, 軽やか, フォーマル)

質問 1 は、コーディネートの妥当性に関する評価に対応するものであり、コーディネートが妥当かどうか

表 2: アンケート結果：質問 1・質問 2

<i>w</i>	1	2	3	4
質問 1				
カジュアル	14.3%	28.6%	38.1%	19.0%
かわいい	0.0%	14.3%	28.6%	57.1%
質問 2				
カジュアル	4.8%	0.0%	19.0%	76.2%
かわいい	19.0%	23.8%	38.1%	19.0%

表 3: アンケート結果：質問 3

選択肢	1 日目	2 日目	3 日目	4 日目
<i>w</i> = カジュアル				
かわいい	45%	0%	30%	0%
冬	5%	0%	5%	60%
軽やか	5%	65%	5%	0%
フォーマル	20%	0%	30%	25%
未回答・その他	25%	35%	30%	15%
<i>w</i> = かわいい				
かわいい	45%	10%	5%	0%
冬	15%	45%	0%	15%
軽やか	0%	20%	5%	55%
フォーマル	5%	0%	70%	0%
未回答・その他	35%	25%	20%	30%

を評価する。また質問 2 は、多様性に関する評価に対応するものであり、毎日なるべく異なるコーディネートが生成されているかを評価するものである。一方質問 3 は、キーワードを用いたコーディネートに対する評価・確認に対応する。正答率が高いほど視覚的にも良いと言える。回答者に提示するコーディネート系列には、各コーデを構成する画像のみを示し、使用したキーワードの説明はすべて省略している。アンケートの集計結果を表 2 に示す。

表 2 より、まず *w* = カジュアルの場合、質問 2 よりコーディネートの多様性では高い結果を得られた。しかし、質問 1 の妥当性においては評価が分散し、妥当性があるとは断言できない結果となった。表 3 よりキーワード通りの印象を受けているの評価に関しては、1 日目の「かわいい」のキーワードに関しては被験者も「かわいい」という印象を受けたことが分かる。しかし、その後の 2 日目から 4 日目に関してはキーワード通りの印象を受けていないことが分かり、表 2 の結果からも照らし合わせると、コーディネートとしては妥当だがキーワード通りの印象は受けていないという結果になった。

次に *w* = かわいいの場合、質問 1 よりコーディネートの妥当性で一定の高い結果を得られた。質問 2 より、*w* = カジュアル に比べ、コーディネートの多様性で

は評価が分散し低い結果となった。表3よりキーワード通りの印象を受けているかの評価に関しては、1日目と2日目のキーワードに関しては被験者もキーワード通りの印象を受けたことが分かり、 $w =$ カジュアルよりも良い結果となった。しかし、その後の3日目と4日目に関してはキーワード通りの印象を受けていないことが分かる。表2の結果からも照らし合わせると、ある程度キーワード通りの印象を持つ妥当なコーディネート系列が生成されているという結果となった。

5 まとめと今後の課題

本研究では、カプセルワードローブに着目したコーディネート系列の生成手法を提案した。より具体的には、トップス・ボトムス・シューズからなるコーディネートを対象に、指定されたスタイルを反映したCWの構築手法、および互換性と多様性に加え、連続性や嗜好との適合性を考慮したコーディネート系列の評価関数を提案した。また実データを対象とした定量・定性評価により、その有効性を確認した。

今後の課題として、より多様なスタイル・嗜好を用いて生成される長大なコーディネート系列を対象とした提案手法の評価が挙げられる。また、より詳細かつ柔軟にスタイルや嗜好を反映する仕組みの開発や、流行や視覚的側面、コストなどを考慮した評価関数の提案が挙げられる。

謝辞 本研究の一部は、JSPS 科研費 JP22K12173 の助成を受けたものである。

参考文献

- [1] W.-L. Hsiao and K. Grauman : Creating Capsule Wardrobes from Fashion Images, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7161–7170, 2018.
- [2] Y. Tanaka and T. Ozaki : An extension of optimal fashion capsule wardrobe construction by considering visual dissimilarity and number of good coordinates, *Proc. of the 2022 Tenth International Symposium on Computing and Networking Workshops*, pp.224–228, 2022.
- [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] 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.
- [5] 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.
- [6] 知見 優一, 北山 大輔 : ファッション SNS におけるアイテム・スタイル特徴に基づく検索システムとその評価, 第12回データ工学と情報マネジメントに関するフォーラム, B2-2, 2020.
- [7] A. Rubio, LongLong Yu, E. Simo-Serra, and F. Moreno-Noguer : Multi-modal joint embedding for fashion product retrieval, *Proc. of the 2017 IEEE International Conference on Image Processing*, pp.400–404, 2017.
- [8] G. Xun, Y. Li, W. X. Zhao, J. Gao, and A. Zhang : A Correlated Topic Model Using Word Embeddings, *Proc. of the 26th International Joint Conference on Artificial Intelligence*, pp.4207–4213, 2017.
- [9] T. Mikolov, K. Chen, G. Corrado and J. Dean : Efficient Estimation of Word Representations in Vector Space, arXiv preprint, arXiv:1301.3781, 2013.
- [10] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean : Distributed Representations of Words and Phrases and their Compositionality, *Advances in Neural Information Processing Systems* 26, pp.3111–3119, 2013.
- [11] X. Song, X. Han, Y. Li, J. Chen, X.-S. Xu, and L. Nie : GP-BPR: Personalized Compatibility Modeling for Clothing Matching, *Proc. of the 27th ACM International Conference on Multimedia*, pp.320–328, 2019.