キーワードに基づく組み合わせの良さを考慮したファッションコーディネート検索

有川　魁人† 　加藤　誠†† 　吉川　正俊†

† 京都大学大学院情報学研究科　〒 606–8501 京都府京都市左京区吉田本町
†† 筑波大学大学院図書館情報メディア研究科　〒 305–8550 茨城県つくば市春日 1-2
E-mail: †arikawa@db.soc.i.kyoto-u.ac.jp, ††mpkato@acm.org, †††yoshikawa@i.kyoto-u.ac.jp

あらまし　本研究では, キーワードをクエリとした組み合わせを考慮したファッションコーディネート検索の手法を提案する. 既存のコーディネート検索手法ではコーディネートの組み合わせを考慮し, かつクエリからコーディネートを検索することは困難である. 提案手法では弱教師あり学習に基づきファッション投稿型 SNS で代表的な wear の投稿情報を用いて組み合わせの良さを推定している. また検索時にはクエリとコーディネートをそれぞれ分散表現に変換し, 分散表現を前向きニューラルネットワークを用いて適合度を予測する. その適合度をもとにクエリについてコーディネートの順位付けを行う.

キーワード　ファッション検索，深層学習, 情報検索

1 はじめに

ファッションコーディネートとは日々の生活の中で必ず行う活動であり, 自分の嗜好や天気, 季節, 場所など様々な状況に合わせてコーディネートの組み合わせを考える必要がある. 近年ではファストファッションの台頭によりコーディネートは更に多様化した. ファストファッションとは「最新の流行を採り入れながら低価格に抑えた衣料品を, 短いサイクルで世界的に大量生産・販売するファッションブランドやその業態.」であり, 2018 年の世界アパレル専門店売上ランキングでは Top3 が 1 位から順に ZARA, H&M, UNIQLO とファストファッションブランドが独占している. このようにファストファッションの台頭に加えて, 人々の服の購買方法も店頭で購入するだけではなく, EC サイトを利用したネット通販が増加しており, 人々は安価で手軽に様々な服を購入することが可能となり, 簡単に多種多様なコーディネートを組み合わせることが可能となった. 一方でコーディネートの組み合わせは無数に考えられるため, ファッションに不慣れな人に対してコーディネートの組み合わせを考えることは負担がかかると考えられる. 2017 年にマイナビ社が行なった調査によれば, 10 代から 30 代の女性 110 名を対象にしたアンケート調査 [1] で約 77%の女性がコーディネートを組み合わせることが難しいと回答した. その理由としては「トップスとボトムスの相性がわからない」,「気温や天気に合わせた服装がわからない」等が挙げられ, この調査結果から人々の大半はコーディネートの組み合わせに対して苦手意識を持っていることが推測される.

このような問題を解決する手段の 1 つとしてコーディネートの検索が考えられ, コーディネートを検索する手法は大きく 3 つに分けられる.

1 点目はキーワードをクエリとしたコーディネート検索である. Instagram や Wear などの SNS ではブーリアン検索が用いられており, キーワードに対して一致したタグを持つコーディネートを検索する手法である. しかし上述の手法ではキーワードマッチングによる検索手法であるため, 複数のキーワードで構成されるクエリの場合はクエリの条件を満たすコーディネートが得られない可能性がある. また Zoghbi ら [2] はキーワードをクエリとしてクエリの意味に適合したコーディネートの全身画像を検索する手法を提案している. しかし上述の手法ではコーディネートの全身画像を検索する手法であるため, トップスとボトムスの相性を考慮した検索を行うことが難しい

2 点目はコーディネート画像をクエリとしてクエリ画像に類似したコーディネートを検索する手法である. Laenen ら [3], [4] はテキストとコーディネート画像をクエリとしてクエリ画像に類似したコーディネートを検索する手法を提案している. しかし上述の手法は類似画像検索であり, 特定のキーワードからコーディネートを検索, コーディネートの組み合わせの良さを考慮したコーディネートの検索を行うことが難しい.

3 点目はファッション画像をクエリとしてクエリ画像に適合したファッション画像を検索してコーディネートを生成する手法である. Song ら [5] はトップス画像とボトムス画像, そしてそれぞれの画像に対して服の形状や色を説明するテキストを学習データとして組み合わせの良さを推定し, トップスまたはボトムス画像をクエリとしてそれぞれに適合したトップスまたはボトムス画像を検索する手法を提案している. 同種の研究として, Han ら [6] はトップス, ボトムス, 鞄などの任意のファッション画像をクエリとして, クエリに対して組み合わせの良いファッション画像を推定しコーディネートを生成している. しかし上述の手法はクエリが画像であり, あるファッション画像に対して適合したファッション画像を検索してコーディネートを生成する手法であるため, キーワードからコーディネートの検索を

行うことがは難しい.

そこで本論文ではキーワードをクエリとして, クエリの意味に適合し組み合わせの良さを考慮したコーディネートの検索を行う手法を提案する. また本論文ではコーディネートをトップスとボトムスの組み合わせと定義し, あるクエリが与えられた時に検索結果上位のコーディネートはクエリに適合し, トップスとボトムスの組み合わせが良いと仮定する. 具体的な検索方法は, クエリは word2vec 用いて分散表現に変換し, コーディネートは ResNet を用いて分散表現に変換する. これらのベクトルを横に結合して前向きニューラルネットワークを介することで クエリと画像の適合度を推定しキーワードの意味とコーディネートの特徴を捉えた検索の手法を提案する. またコーディネートの組み合わせの良さは SNS 上のデータを用いて弱教師あり学習に基づき推定する.

本論文の貢献はキーワードに基づく組み合わせを考慮したコーディネート検索の有用性を明らかにしたことである. 具体的にはコーディネート投稿型 SNS で代表的な wear のタグ情報から弱教師あり学習に基づきキーワードから組み合わせの良いコーディネートを検索する手法を提案した. また評価実験では検索モデルの代表的な評価指標を用いてキーワードに対して組み合わせを考慮しない検索モデル, クエリ尤度検索モデルと比較を行い, 組み合わせを考慮しない検索モデルの検索性能を上回ったことを明らかにした.

本論文の構成は次の通りである. 第 2 節では関連研究としてファッションに関連する研究, 弱教師あり学習を用いた検索方法について説明する. 第 3 節ではキーワードをクエリとして組み合わせの良いコーディネートを検索する手法を提案し, 第 4 節では評価実験を通して提案手法の有用性を確認する. 第 5 節では本論文の結論とともに今後の課題について説明する.

2 関 連 研 究

ファッションに関連する研究では複数のカテゴリに大別される. 本論文では, コーディネート画像検索, コーディネートの組み合わせ推定の関連研究について説明する.

2.1 コーディネート画像検索

Zoghbi ら [2] はクエリとコーディネート画像をペアとして共通空間に埋め込むことでクエリからコーディネート画像を検索する手法を提案している. 具体的には, クエリは bag-of-words でベクトルに変換し, 画像は CNN を用いてベクトルに変換している. これらのベクトルを Bilingual LDA(BiLDA) を用いて複数のトピックとして表現し共通空間に埋め込んでいる. 検索時にクエリのトピックとコーディネート画像のトピックの適合度を推定し順位付けを行う. 本研究ではクエリとコーディネート画像を分散表現に変換し, クエリとコーディネートの適合度を推定するという点で Zoghbi らの手法と類似しているが, 本研究では前向きニューラルネットワークを用いて適合度の推定を行う. また Zoghbi らの手法ではコーディネートの組み合わせの良さについては考慮されておらず, 本研究ではコーディネートの組み合わせの良さを考慮した検索が可能である.

2.2 コーディネートの組み合わせ推定

Song ら [7] はトップス画像, またはボトムス画像を入力として入力に適合したトップス画像またはボトムス画像を検索する手法を提案している. [7] ではトップス画像とボトムス画像, そしてそれぞれの画像に対して服の形状や色を説明するテキストを Autoencoder へ入力し, Autoencoder の隠れ層から特徴ベクトル $\mathbf{i}^t, \mathbf{i}^b, \mathbf{c}^t, \mathbf{c}^b$ を抽出する. また抽出した特徴ベクトルの適合度 m を計算し, その計算結果をもとに Baysian Personalized Ranking (BPR) [8] を用いてランキングを生成する. 適合度 m は以下の式で計算される.

$$m = (1-\beta)(\mathbf{i}^t)^T\mathbf{i}^b + \beta(\mathbf{c}^t)^T\mathbf{c}^b \tag{1}$$

β はハイパーパラメータである. またトップス画像に対して組み合わせの良いボトムス画像を検索するタスクがあると仮定すると, BPR はトップスがあるボトムスをどの程度好むかという適合度 m を予測するモデルであり, トップスに対して組み合わせの良いボトムスとランダムサンプリングしたボトムスそれぞれの m を計算する. また計算された m の差を sigmoid 関数 σ を介してどちらのボトムスが好まれるかを予測する. トップス t に対してボトムス b_i がボトムス b_j より好まれる確率は以下の式で表される.

$$p(b_i > b_j; t) = \sigma(m_{tb_i} - m_{tb_j}) \tag{2}$$

また, 組み合わせの良さはファッションの専門家による知識を教師データとして推定する.　本研究ではコーディネートの組み合わせの良さを推定するという点で Song らの手法と類似しているが, 本研究では組み合わせの良さはファッション投稿型 SNS の投稿情報を教師データとして推定する. また Song らの手法では画像をクエリとしているため, 得られるコーディネートの数は限られてしまう. 本研究ではキーワードをクエリとして用いたコーディネート検索のため, 様々な種類のコーディネートを検索することが可能である.

3 組み合わせを考慮したコーディネート検索

3.1 概　　要

本節では, 第 1 節で述べた通り, キーワードから組み合わせの良いコーディネートを検索する手法の提案を行う.

提案手法の概要図を図 1 に示す. 本研究では入力をクエリ $q \in Q$ とし, Q はすべてのクエリの集合である. また出力をコーディネート $c \in C$ のリストとし, C はすべてのコーディネートの集合である. この時クエリ q は単語 $w_i \in W$ から構成され, W はすべての単語の集合である. またコーディネートをトップスとボトムスの組み合わせと定義し, トップスを $t \in T$, ボトムスを $b \in B$ とし, T, B はそれぞれすべてのトップス, ボトムスの集合である. また $c = (t, b)$ とする. またクエリとトップス, ボトムスをそれぞれ word2vec と ResNet を用いて分散表現に変換する.

この時, クエリの分散表現を $\mathbf{q} \in \mathbb{R}^n$, トップスとボトムスの分

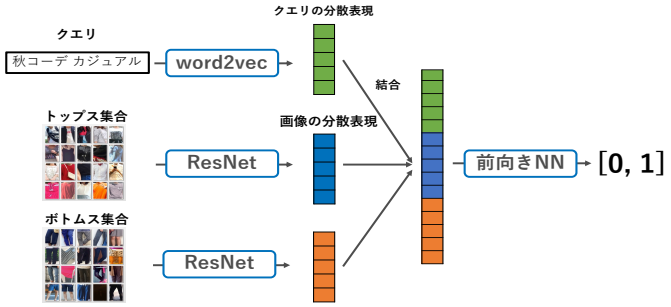

図 1　提案手法の概要図

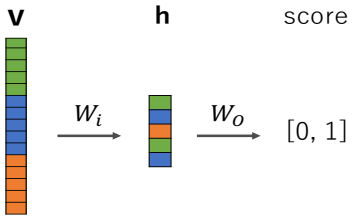

図 2　適合度計算の概要図

散表現をそれぞれ $\mathbf{t} \in \mathbb{R}^m, \mathbf{b} \in \mathbb{R}^m$ とする. またクエリと画像の分散表現を横に連結して前向きニューラルネットワークを介して適合度を計算する. また図 2 よりクエリと画像の分散表現を横に連結したベクトルを $\mathbf{v} \in \mathbb{R}^{n+2\times m}$ とすると適合度 (score) は以下の式で計算される:

$$\mathbf{h} = \mathbf{W_i}^T\mathbf{h} + \mathbf{b} \tag{3}$$

$$\text{score} = \text{sigmoid}(\mathbf{W_o}^T\mathbf{h} + \mathbf{c}) \tag{4}$$

$\mathbf{h}$ は前向き NN の隠れ層の写像ベクトル, ${W_o}^T$ は前向き NN のパラメータ, $\mathbf{b}, \mathbf{c}$ はバイアス項, sigmoid はシグモイド関数を表す. また適合度の大きさの順にコーディネートを出力する.

以降の小節では 3.2 節にクエリ, 画像のベクトル表現方法, 3.3 節に学習方法について説明する.

3.2　クエリ, 画像のベクトル表現

提案手法でははじめにクエリ, 画像それぞれを分散表現に変換している. クエリに対しては単語の意味を分散表現として表現するために word2vec [9] を用いて, 低次元の分散表現を生成している. また, クエリは 2 語の単語から成り立つ為, それぞれの単語をベクトルに変換した後, それらのベクトルで Max-Pooling をとり単一のクエリベクトルに変換している. また Max-Pooling でクエリベクトルを生成する手法は複数の単語ベクトルを単一のクエリベクトルで表現する手法の 1 つである. [10]

画像については画像の色, 形状などの特徴をベクトルとして抽出するために, 画像分類などの研究で顕著な効果を発揮している ResNet [11] を用いて, 低次元の特徴ベクトルを生成している.

3.3　学 習 方 法

本論文では, 学習データ集合を

$$T = \{(\mathbf{q}_1, \mathbf{x}_{11}, \mathbf{x}_{12}, y_1), ..., (\mathbf{q}_N, \mathbf{x}_{N1}, \mathbf{x}_{N2}, y_N)\}\ (\mathbf{x} = (\mathbf{t}, \mathbf{b})) \tag{5}$$

と定義する. 学習データはそれぞれクエリ $\mathbf{q}_i$, 2 つのコーディネート候補 $(t_{i1}, b_{i1}), (t_{i2}, b_{i2})$, どちらのコーディネートがクエリに適合しているか示すラベル $y_i \in \{-1, 1\}$ に対応している. また図 3 が示す様に, Pairwise でモデルを学習させており, 目的関数に Pairwise 学習でランキング付けをする為に広く使用されてきた Hinge Loss [12] を設定する. [12] より Pairwise でランキングを生成する場合 Hinge Loss は以下の式で表される.

$$\mathcal{L} = \sum_{i=1}^{N} \max\{0, \epsilon - y_i f(\mathbf{x}_{i1} - \mathbf{x}_{i2})\} \tag{6}$$

式 (6) の y_i は $f(\mathbf{x}_{i1} - \mathbf{x}_{i2})$ がどのクラスに属するか対応するラベルである. ϵ はクラスのマージンに相当し, ϵ が 0 に近づく程 $f(\mathbf{x}_{i1} - \mathbf{x}_{i2}) > 0$ の場合学習が行われない. また Hinge Loss を用いて学習データの i 番目で目的関数を計算する場合以下の式で表される.

$$\mathcal{L} = \max\{0, \epsilon - y_i[\text{score}(\mathbf{q}_i \| \mathbf{x}_{i1}) - \text{score}(\mathbf{q}_i \| \mathbf{x}_{i2})]\} \quad (\mathbf{x} = (\mathbf{t}, \mathbf{b})) \tag{7}$$

$\|$ は連結を示す. ϵ はハイパーパラメータである. また上述の学習方法は弱教師あり学習 [13], [14] に基づいており, 本研究では SNS の投稿で用いられるタグをクエリとして用いている. 本研究では 2 つのコーディネート候補のサンプリングには, クエリ, コーディネート集合が与えられた時, クエリ $\mathbf{q}_i$ に対してクエリ尤度が高いコーディネートを 1 組, コーディネート集合全体からコーディネートを 1 組それぞれランダムサンプリングして, 学習候補を生成する. クエリ尤度は SNS の投稿で用いられるタグを利用して計算を行う. またクエリ尤度モデルは多項分布に基づいており, 語は独立に生起すると仮定する. この時 SNS の投稿 D とクエリ Q とするとクエリ尤度 $P(Q|\theta_D)$ は以下の式で定義される.

$$P(Q|\theta_\text{D}) = \prod_{i=0}^{|Q|} P(w_i|\theta_D) = \prod_{w_i \in Q} \frac{\text{c}(w_i, \text{D})}{|\text{D}|} \tag{8}$$

この時 $w_i \in Q$ はクエリ Q 内の単語である. また $\text{c}(w_i, \text{D})$ は投稿 D において単語 w_i が出現する頻度, $|\text{D}|$ は投稿 D のタグ数を示す. 提案手法では単語 w_i が存在しない投稿のクエリ尤度が零になる問題を防ぐために線形補間スムージング [15] でクエリ尤度を計算している. スムージングを施したクエリ尤度モデルは以下の式で表される.

$$P(Q|\theta_\text{D}) = \prod_{w_i \in Q} \{\lambda \frac{\text{c}(w_i, \text{D})}{|\text{D}|} + (1 - \lambda) P(w_i|\theta_C)\} \tag{9}$$

$$P(w_i|\theta_C) = \frac{\sum_{\text{D}\in C} \text{c}(w_i, \text{D})}{\sum_{\text{D}\in C} |\text{D}|} \tag{10}$$

λ はハイパーパラメータであり, C は投稿 D の集合である. またその時のラベル y_i は以下の式で表される.

$$y_i = \text{sign}(\sum_{\{t,b\}\in x} P(\mathbf{q}_i|\text{D}_{\mathbf{x_{i1}}}) - \sum_{\{t,b\}\in x} P(\mathbf{q}_i|\text{D}_{\mathbf{x_{i2}}})) \tag{11}$$

$$\text{sign}(x) = \begin{cases} 1\ (x > 0) \\ -1\ (\text{otherwise}) \end{cases} \tag{12}$$

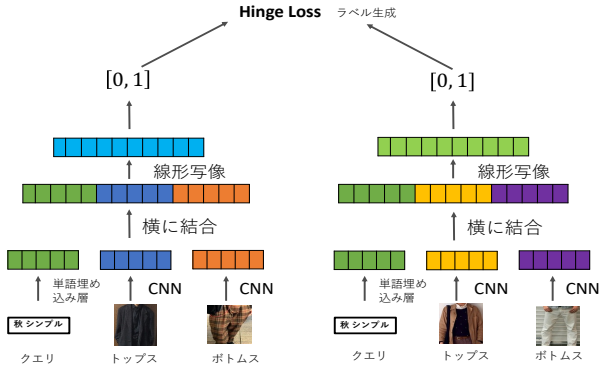

図 3　学 習 方 法

4　実 験 方 法

本節では, 4.1 節に実験で使用するデータセットについて説明する, 4.2 節では実験設定について説明する.

4.1　概　　要

本節で提案した手法が有効であるか組み合わせを考慮しないコーディネート検索をベースライン手法に設定し Precision@k (k=1, 3, 5, 10, 20), MAP@k で提案手法の有効性を測る.

また以降の小節で 4.2 節ではデータセットについて説明を行い, 4.3 節ではクエリの生成方法について説明を行い, 4.4 節ではトップス画像とボトムス画像の抽出方法について説明を行い, 4.5 節では実験設定についての説明を行い, 4.6 節にはベースライン手法と比較した結果を説明する.

4.2　データセット

本研究ではコーディネート投稿型 SNS「wear[1]」の投稿情報を利用してクエリ, コーディネートの生成を行なっている. 2020 年 2 月現在, wear ではユーザ数約 480 万人, 投稿数約 850 万件と多くのユーザに利用されており, wear ではユーザが自身で撮影したコーディネート画像を投稿しており, 本研究ではコーディネート画像からトップス, ボトムスのみを抽出し利用している.

またユーザは自身のコーディネートに対して任意の個数タグをつけることができる. 本研究ではこのタグを利用してクエリを生成している. また本研究では 2013 年 3 月 19 日から 2018 年 5 月 6 日の 1,806,147 件の投稿を利用している.

4.3　クエリの生成方法

本研究ではユーザが付与したタグからクエリを生成するにあたり, 最低 1,000 件以上の投稿に付けられているタグを使用している. 図 4 より最大約 20 万件の投稿にタグが付けられており, 「シンプル, カジュアル」などコーディネートの印象, 「デニム, ニット」など服の形状, 「休日スタイル, 春のコーデ, 夏コーデ」など状況や季節と言った様々な種類のタグが利用されていることがわかる. 本研究ではクエリの生成に, 「夏 派手」の様に 2 つのタグで 1 クエリを生成している. また関連性の薄いタグから生成されるクエリは適合したコーディネートを検索することが難しいと考えられ, 検索性能を低下させる恐れがある. そのためクエリの選定方法として Pointwise Mutual Information (PMI) に従いクエリを選定した. 具体的には, PMI が正かつクエリの 2 つのタグを含む投稿が最低 100 件あるクエリを選定している.

1 : https://wear.jp/

表 1　データセットの構成

#pair	*#query*	*#pair/#query*
1,488,000	1,488	1,000

4.4　トップス, ボトムス画像の抽出方法

本研究ではコーディネートの組み合わせの良さを推定するため, wear のコーディネート画像から適切にトップス, ボトムス画像を抽出しなければならない. そこでトップス, ボトムス画像の抽出方法として Openpose[2]を使用した. Openpose とは Pose Estimation と呼ばれる画像から人の骨格を検出する代表的な手法の 1 つであり, Deep Learning を応用して人の関節や動きを正確に検出することが可能である. 本研究では 18 COCO keypoints を使用しており [16], 体のパーツ (keypoints) を用いて骨格を描画している. トップスとボトムスの分割には腰の位置にある keypoints より上部をトップス, 下部をボトムスと定義し分割した. また服の形状に無関係な背景は取り除いている. また身体の一部分のみを写しているコーディネートについては適切にトップスとボトムスを抽出することができないためデータセットには含めていない.

4.5　実 験 設 定

本研究で使用するデータセットは (クエリ, クエリに適合したコーディネート, クエリに不適合なコーディネート, ラベル) を 1 ペアのデータとする. またクエリに適合したコーディネートについてはクエリ尤度が高くかつ wear に実在するコーディネートとする. またクエリに不適合なコーディネートはクエリ尤度が低くかつ wear に実在しないコーディネートとする. またデータセットの構成を表 1 に示す. また本研究では, 学習データと検証データの割合を 8:2 とする.

また本研究で提案している検索モデルおよび学習は Pytorch[3] で実装している. またモデルのネットワークのパラメータは Adam [17] で最適化している. 実験では学習率を 5×10^{-5} としバッチサイズは 128 に設定した. またクエリは日本語コーパス学習済みの word2vec を使用して単語埋め込みベクトルを次元数 300 で初期化している. トップス, ボトムス画像はそれぞれ学習済みの ResNet を使用して画像の特徴ベクトルを次元数 512 で初期化している.

4.6　検索性能に関する評価

本小節では 4.1 節で説明した通り提案手法の有効性を明らかにするために, ベースライン手法を設定して検索性能に関して比較を行う. 以下にベースライン手法の詳細について説明する.

2 : https://github.com/CMU-Perceptual-Computing-Lab/openpose
3 : https://pytorch.org/

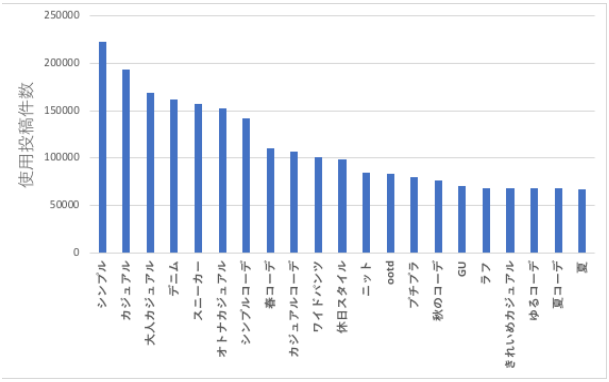

図 4　タグの分布

表 2　Precision と MAP の評価結果

Method	Precision@1	Precision@3	Precision@5	Precision@10	Precision@20
提案手法	0.9167	0.8773	0.8521	0.8111	0.7424
WCC	0.5799	0.5243	0.5014	0.4469	0.3925

Method	MAP@1	MAP@3	MAP@5	MAP@10	MAP@20
提案手法	0.9167	0.8459	0.8025	0.7314	0.6355
WCC	0.5799	0.4375	0.3841	0.3024	0.2306

組み合わせを考慮しないコーディネート検索手法 (Without Considering Combinations, WCC): トップスとボトムスを連結し 1 つのコーディネート画像として扱う. 提案手法と同様にクエリとコーディネート画像はそれぞれ word2vec と ResNet を用いて分散表現に変換する. またそれぞれのベクトルを横に結合し前向きニューラルネットワークを介することでクエリとコーディネート画像の適合度を推定する. ハイパーパラメータは提案手法と同じ数値に設定している. また 4.1 節で説明した通り代表的な検索指標を用いて提案手法とベースライン手法の検索性能の比較を行う. また, 検索指標に関しては Precision@k, MAP@k を用いる. また本研究では $k = \{1, 3, 5, 10, 20\}$ とする. また本研究では検証データ内の各クエリに対応するペアに使用した (トップス, ボトムス) の組み合わせを各クエリに対するコーディネート集合と定義し, 検索指標の算出を行なった. 表 2 に評価指標の評価結果を示す.

表 2 は提案手法およびベースライン手法に対して Precision@k, MAP@k を計算した結果である. 表 2 より Precision@k, MAP@k の順位は提案手法 > WCC であることがわかる. また提案手法が WCC より検索性能が上回った理由としては, クエリに対してコーディネートを学習させるよりもコーディネートをトップスとボトムスに分け学習させた方がトップス, ボトムスそれぞれの表現力が高まり, クエリの意味と合致したため検索性能を向上させたと推測する.

5　ま　と　め

本研究では, キーワードに基づく組み合わせの良さを考慮したファッションコーディネート検索の手法について提案した. 提案手法ではクエリ, トップス, ボトムスを word2vec, ResNet を介して分散表現に変換し, それらのベクトルを連結し前向きニューラルネットワークに介することでクエリとコーディネートの適合度を予測した. またトップスとボトムスの組み合わせの良さは弱教師あり学習に基いている.

実験では代表的な検索指標を用いてトップスとボトムスの組み合わせの良さを考慮しない検索モデルと比較を行なった結果, 我々の提案手法が全ての検索性能で上回り, 提案手法の有効性を明らかにした.

今後の課題として, 検索モデルの改良, 大規模なコーディネート集合における効率的な検索手法の提案について検討する予定である.

文　献

[1] お客様にコーディネートの悩みを聞いてみた。心に響くアドバイスとは！？ ｜アパレルクローゼット times. https://baito.mynavi.jp/apparel/times/nayami/article-305.html.

[2] Susana Zoghbi, Geert Heyman, Juan Carlos Gomez, and Marie-Francine Moens. Cross-modal fashion search. In *International Conference on Multimedia Modeling*, pp. 367–373. Springer, 2016.

[3] Katrien Laenen, Susana Zoghbi, and Marie-Francine Moens. Web search of fashion items with multimodal querying. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 342–350. ACM, 2018.

[4] Katrien Laenen, Susana Zoghbi, and Marie-Francine Moens. Cross-modal search for fashion attributes. In *Proceedings of the KDD 2017 Workshop on Machine Learning Meets Fashion*, Vol. 2017, pp. 1–10. ACM, 2017.

[5] Xuemeng Song, Fuli Feng, Xianjing Han, Xin Yang, Wei Liu, and Liqiang Nie. Neural compatibility modeling with attentive knowledge distillation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 5–14. ACM, 2018.

[6] Xintong Han, Zuxuan Wu, Yu-Gang Jiang, and Larry S Davis. Learning fashion compatibility with bidirectional lstms. In *Proceedings of the 25th ACM international conference on Multimedia*, pp. 1078–1086. ACM, 2017.

[7] Xuemeng Song, Fuli Feng, Jinhuan Liu, Zekun Li, Liqiang Nie, and Jun Ma. Neurostylist: Neural compatibility modeling for clothing matching. In *Proceedings of the 25th ACM international conference on Multimedia*, pp. 753–761, 2017.

[8] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*, 2012.

[9] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pp. 3111–3119, 2013.

[10] Dinghan Shen, Guoyin Wang, Wenlin Wang, Martin Renqiang Min, Qinliang Su, Yizhe Zhang, Chunyuan Li, Ricardo Henao, and Lawrence Carin. Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms. *arXiv preprint arXiv:1805.09843*, 2018.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[12] Hang Li. Learning to rank for information retrieval and natural language processing. *Synthesis Lectures on Human Language Technologies*, Vol. 4, No. 1, pp. 1–113, 2011.

[13] Mostafa Dehghani, Hamed Zamani, Aliaksei Severyn, Jaap Kamps, and W Bruce Croft. Neural ranking models with weak supervision. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 65–74. ACM, 2017.

[14] Hamed Zamani, Mostafa Dehghani, W Bruce Croft, Erik Learned-

Miller, and Jaap Kamps. From neural re-ranking to neural ranking: Learning a sparse representation for inverted indexing. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 497–506. ACM, 2018.

[15] Chengxiang Zhai and John Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval. In *ACM SIGIR Forum*, Vol. 51, pp. 268–276. ACM, 2017.

[16] openpose/output.md at master · cmu-perceptual-computing-lab/openpose · github. `https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/output.md`. (Accessed on 02/03/2020).

[17] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.