

# オンラインショッピングサイトにおけるコーディネート類似度判定手法とその印象評価

1村上大志 2黒澤義明 2目良和也 2竹澤寿幸

1広島市立大学情報科学部 2広島市立大学大学院 情報科学研究科

## 1 はじめに

通信販売業界において、2012年の1年間での世帯購入率は8割、個人利用率は6割を超え、非常に巨大な市場になってきている[1]。2012年においてオンラインショッピングを利用して購入された品目は、「本・雑誌・コミック」、「食料品・飲料(アルコール類をのぞく)」に次いで「レディースファッション・靴」が多い[2]。また、オンラインショッピングに限らず、通信販売により2012年に購入された商品は「婦人衣料品」の占める割合が最も高い。

近年、ファッション関連のオンラインサイトでは、商品を単体で販売するのみでなく、複数の写真を組み合わせてコーディネートを作成し、販売することで、購買者を増やす傾向がある。コーディネートを作成する際に重要になる要素は、コーディネート紹介写真や文中で印象や雰囲気述べている単語である。

本研究では、あるコーディネートを選択すると類似したコーディネート提案するシステムを構築するための基礎情報の調査を行う。システムの概要を図1に示す。このシステムは、企業側からは売り上げ向上、ユーザ側からは普段は着用しないコーディネートに気づくという利点がある。

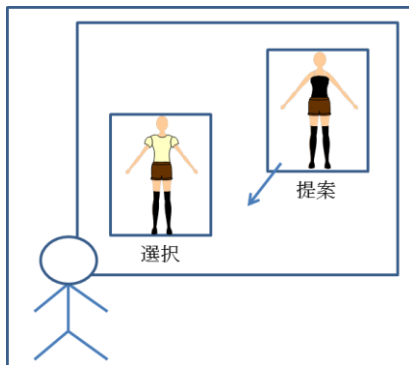


図1: システムの概要

## 2 関連研究

本章は、本研究に関連する研究について述べる。

### 2.1 ファッションイメージキーワードに基づいたコーディネートシステムの提案

神間ら[3]は、ファッションを記述する感性語として定義されているイメージキーワード()と「気づき」の関係について調べた結果、イメージキー

ワードがコーディネートの推薦に有用であることを示している。

本研究では、着用アイテムに属性辞書を用いて、自動でタグ付けを行い属性を付与し、その属性を基に類似度を求める。

### 2.2 ファッション雑誌を用いたファッション推薦システム

岩田ら[4]は、雑誌の写真画像から、コーディネートに関する情報を自動抽出し、抽出した上下の衣服領域の類似度に基づく手法とトピックモデルに基づく手法を用いて、コーディネートを推薦するシステムを提案している。

本研究では、コーディネートの画像情報ではなく、画像に付与されている着用アイテムの情報や紹介文中の特徴語を用いる。

### 2.3 服の着用経験に基づき服自身がコーディネート推薦するシステム

福田ら[5]は、ユーザがその日の気分を色に変換し、システムから提示されたカラーパレットから選択する。選択された色のイメージに最も近い服が推薦され、そこから得ることが出来る情報を蓄積し、その情報を用いて推薦を行う。

本研究では、色情報ではなく、画像に付与されている着用アイテムの情報や紹介文中の特徴語を用いる。

## 3 提案手法

本研究で対象とするデータは、オンラインショッピングサイト内にあるコーディネート紹介ページに付与されている着用アイテムの情報、およびコーディネート紹介文である。本研究で対象とするデータを図2に示す。

本研究における提案手法を以下に3つ示す。なお、アイテムタグ手法と全タグ手法に対し、属性辞書を用いて着用アイテム名に自動でタグ付けを行う。

- アイテムタグ手法: アイテムのタグのみを用いて類似度を求める手法
- 全タグ手法: 全てのタグを用いて類似度を求める手法
- 特徴語手法: コーディネート紹介文中の特徴語を用いる手法

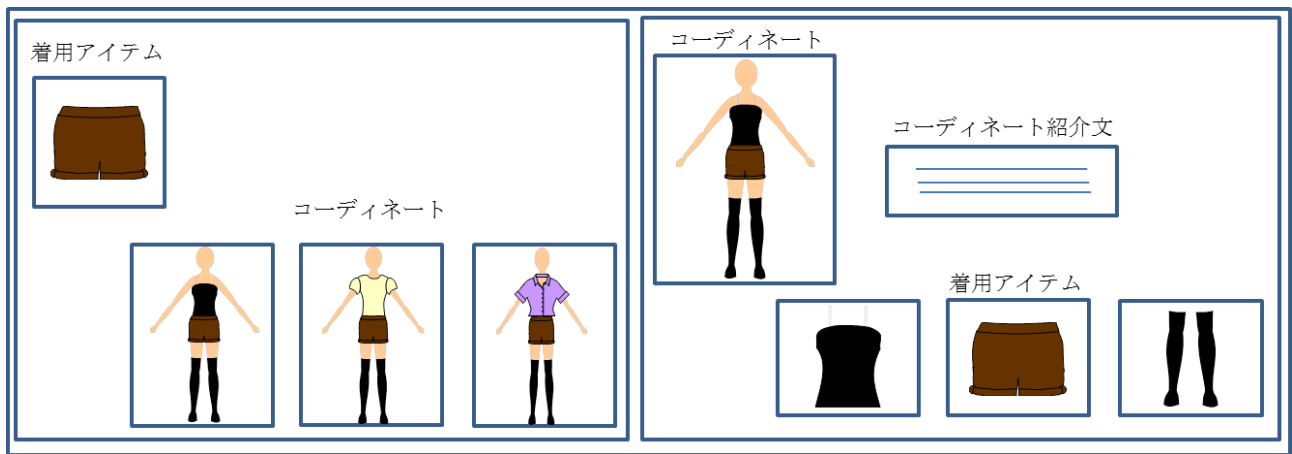


図 2：本研究で対象とするアパレルサイトの典型例

属性辞書はアパレル関連の書籍[6][7][8]およびファッション用語辞典 apparel-fashion wiki[9]を基に作成し、色、ディテール、アイテム、素材、柄、シルエット、技術、国名、人物名、サイズ、小ささ、トレンド、その他、不明、不必要の 15 種類である。タグ付けの流れを図 3 に示す。

### 3.1 アイテムタグ手法

アイテムタグ手法では、タグ付けされた着用アイテムのアイテム(最後の部分)のタグからコサイン類似度を求める。属性として収集したタグは 43 種類であった。また、各コーディネートの平均タグ数は、1.96 であった。収集したタグの一部を表 1 に示す。

表 1：収集したタグ

T シャツ	コンビネゾン	スタジャン
トップス	ブラウス	アウター
カーディガン	ニット	ガウン
プルオーバー	V チュニック	ショート
パンツ	シャツ	キュロット

アイテムタグ手法で得られたコーディネートの組み合わせは全 241 種類であった。しかし、トップスとパンツの組み合わせのコーディネートが 55 件、トップスとスカートの組み合わせのコーディネートが 51 件と 1 つのコーディネートに属している数が非常に多くなった。これは、アイ

テムタグ手法では、収集したタグの種類が 43 種類と少なかったためと考えられる。コーディネートの上位 10 件を表 2 に示す。

表 2：コーディネートの上位 10 件

コーディネート		件数(件)
ワンピース		85
トップス	パンツ	55
トップス	スカート	51
プルオーバー	スカート	42
ジャケット	ワンピース	35
プルオーバー	パンツ	31
カーディガン	ワンピース	31
T シャツ	パンツ	26
スカート		26
T シャツ	スカート	24

テムタグ手法では、類似度を基に提案できるコーディネートの量が多いのでユーザにとって有用ではない。よって、本研究では全タグ手法および特徴語手法を用いて、実験、考察を行う。

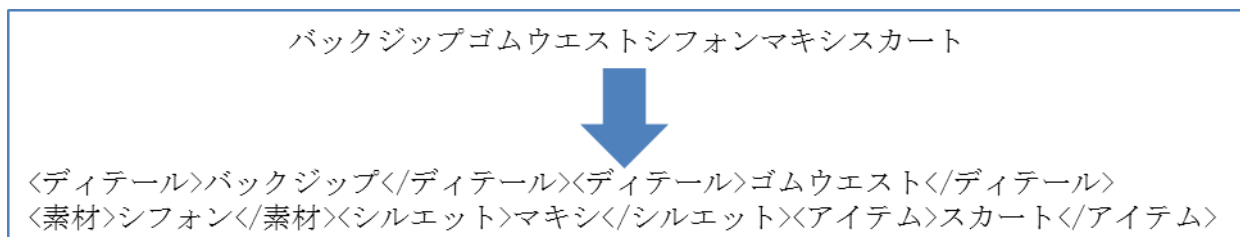


図 3：タグ付けの流れ

### 3.2 全タグ手法

全タグ手法では、タグ付けされた着用アイテムの全てのタグからコサイン類似度を求める。属性として収集したタグは 637 種類であった。収集したタグの一部を表 3 に、コーディネート の平均タグ数を表 4 に示す。表 3 の着用アイテム数とは、各コーディネートに付与されているアイテムの数である。着用アイテム数 2 の平均タグ数 9.34 は類似度が 0.5 の場合、4~5 個のタグが一致していることを示す。

表 3: 収集したタグ

ジャンクフード	マルチカラー	アウター
ピーナッツ	サテン	カフス
T シャツ	リブ	ビジュアール
後ろリボン	ノーカラー	ガウン

表 4: コーディネートの平均タグ数

着用アイテム数	平均タグ数
1	4.66
2	9.34
3	13.19
4	17.86
全コーディネート	9.01

### 3.3 特徴語手法

特徴語手法では、コーディネート紹介文中の特徴語を用いる。特徴語とは、コーディネート紹介文から形態素解析を行った解析結果の中から形容詞 179 種類を用いて、「Yahoo!知恵袋データ(第 2 版)」から各単語の出現率の比を求め、ファッションに関連する形容詞である。以下に出現率の比の式(1)を示す。

$$\text{出現率の比} = \frac{\text{「カテゴリ---ファッション」内の単語の出現率}}{\text{それ以外のカテゴリ内の単語の出現率}} \quad (1)$$

なお今回は、全タグ手法、特徴語手法の比較のため、両条件で提案コーディネートが変化するよう、重みの設定を行った。以下に重みの計算式(2)を示す。

$$\text{重み} = \frac{1}{(\text{各コーディネートの特徴語の出現頻度の和})^3} \quad (2)$$

上式に基づき、全タグ手法で求めたコサイン類似度の値に重みをかけ、特徴語手法の値とした。全タグ手法上位 100 件と特徴語手法 100 件を比較すると、特徴語手法の上位 100 件に含まれ、かつ全タグ手法と異なるコーディネート対は 36 件であった。本研究では、特にこの 36 件(全-特手法)について考察を行う。

## 4 印象評価

### 4.1 使用データ

使用データは特徴語が 2 単語以上出現し着用アイテムの情報が上半身、下半身 1 つずつの計 2 つ存在しているコーディネート対を選んだ。総数は 348 件であった。

### 4.2 評価方法

評価サイトを用いて、前述の 348 件が似ているかどうかの印象評価を行った。被験者は 5 人の女子大学生であった。評価サイトの例を図 4 に示す。



似てる ① ② ③ ④ ⑤ ⑥ ⑦似てない

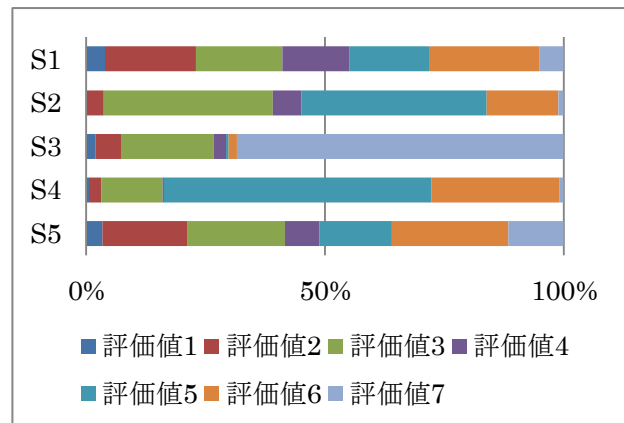
Next

図 4: 評価用サイトの例

### 4.3 評価結果

評価結果を表 3 に示す。なお、評価値は 7 段階評価で、「1:非常に似ている, 2:似ている, 3:やや似ている, 4:どちらでもない, 5:やや似ていない, 6:似ていない, 7:非常に似ていない」に設定した。7 段階に設定した理由は、評価値をばらつかせて、より詳細な評価を得るためである。表 3 の S とは、被験者のことを示す。

表 3: 評価結果



## 5 考察

全タグ手法の上位 100 件と特徴語手法の上位 100 件, 3.3 節の 36 件を比較し, 似ていると評価された結果を表 5 に示す. 表 4 の括弧内は件数を表し, 比率は全タグ手法, 特徴語手法は 100 件で, 全-特手法は 36 件で件数を割った値である.

表 5: 似ていると評価された件数および比率

	1人 以上	2人 以上	3人 以上	4人 以上	5人
全タグ手法	0.83 (83)	0.55 (55)	0.36 (36)	0.18 (18)	0.06 (6)
特徴語手法	0.78 (78)	0.52 (52)	0.33 (33)	0.14 (14)	0.06 (6)
全-特手法	0.72 (26)	0.39 (14)	0.22 (8)	0.08 (3)	0.06 (2)

全タグ手法では似ていないとされていたコーディネート対 36 件に対して, 特徴語手法では 5 人の内 1 人以上が似ていると評価したコーディネート対が 26 件発見された. したがって, 本研究の特徴語手法は有用であったと言える.

また, 全タグ手法で求めたコサイン類似度の値の高い上位 5 件について表 6 に示す. ここでの評価値は, 「1: 似ている, 2: 似ていない」を示す.

表 6: 全タグ手法で値の高い上位 5 件

類似度	評価値
0.625	1
0.589	1
0.572	1
0.556	2
0.510	2

表 5 より全タグ手法では高い値を示し, 似ているとされていたコーディネートでも, 印象評価では似ていないと評価されることがある. よって, 全タグ手法での値が高いという理由だけで, コーディネートを提案することは妥当ではない.

## 6 おわりに

本研究では, 特徴語手法を用いて印象評価を行い, 全タグ手法との比較, 考察を行った. 考察の結果, 全タグ手法の値のみでコーディネート提案することは妥当ではないことがわかった. また, 特徴語手法を用いると, 全タグ手法では似ていないとされていたのに対し, 似ていると評価されたコーディネートが 26 件発見された. よって, 特徴語手法が有用であることがわかった.

## 7 今後の課題

今後の課題としては, より多くの画像を用いて, 同様の実験を行うことや特徴語を形容詞, 形容動詞と限定せず, その他の自立語やオノマトペで実験し結果が異なるか考察する必要がある.

### 謝辞

この研究の一部は, 広島市立大学特定研究費 (一般研究平成 24 年度~) の補助を得ている. また, 本研究の実施に際し, ヤフー株式会社から国立研究所に提供された「Yahoo!知恵袋データ(第 2 版)」を利用している.

### 参考文献

- [1] 公益社団法人 日本通信販売協会, “第 20 回全国通信販売利用実態調査報告書”, 2013.
- [2] 公益社団法人 日本通信販売協会, “第 5 回インターネット通信販売利用実態調査報告”, 2013.
- [3] 神間唯, 丸谷宜史, 梶田将司, 間瀬健二: “ファッションイメージキーワードに基づいたコーディネートシステムの提案”, 情報処理学会研究報告, HCI, ヒューマンコンピュータインタラクション研究会報告, pp.1-7, 2011.
- [4] 岩田具治, 渡部晋治, 澤田宏: “ファッション雑誌を用いたコーディネート推薦システム”, 情報科学技術フォーラム講演論文集 9(3), pp.179-180, 2010.
- [5] 福田未央, 仲谷義雄: “服の着用経験に基づき服自身がコーディネート推薦するシステム”, 情報処理学会全国大会講演論文集 2011(1), pp.161-163, 2011.
- [6] 富田明美, “アパレル構成学 着やすさと美しさを求めて”, 朝倉出版, 2004.
- [7] 飯塚弘子, 内田八重子, 香川幸子, “服装デザイン論”, 文化出版局, 1985.
- [8] 小林茂雄, “装いの心理 服飾心理学へのプロムナード”, アイ・ケイコーポレーション, 2003.
- [9] ファッション用語辞典 apparel-fashion wiki, <http://apparelwiki.symphonic-net.com/>.