

## ファッションアイテム構成要素自動抽出のためのデータ収集と分類

黒澤 義明 村上 太志 竹澤 寿幸  
 広島市立大学大学院 情報科学研究科  
 {kurosawa, murakami, takezawa}@ls.info.hiroshima-cu.ac.jp

## 1. はじめに

大規模複合ファッションアイテム通販サイトが増えてきている。それらサイト上で扱われている製品の量は非常に多い。例えば、スカートだけ見ても、ZOZOTOWN で 15,052 件、MAGASEEK で 7,611 件、Stylife で 4,423 件が登録されている (2015 年 1 月 20 日調べ)。

これらのアイテムは売切れればリンクが外され、新しい商品が開発されれば、新たなリンクが形成される。この量はどれくらいなのだろうか？ amazon で検索すると、最近 1 週間に登録されたレディーススカートが 2574 件にものぼる。

このような量のアイテムを全て見たうえで購買行動を取ることは現実的ではないため、何かしらの検索機能が必要となる。1 番に挙げられるのはキーワードにもとづく検索である。これに加えて各サイトで、様々な手段を講じている。例えば、amazon では「スタイル」、ZOZOTOWN ではスライダーを使った「サイズ」等が特徴的な方法として挙げられる。

しかし、これだけでユーザが望む検索方法を網羅していると言えるだろうか？ 例えば、『透け感』というキーワードで表されるような、うっすらと袖のあたりが透けたブラウスを探していると仮定する。しかし、システム担当者が『シア感』と書いたり、『オーガンジーを使った...』と書いたりした場合には、ユーザは目的のアイテムを購入することができない。サイト側としては大きな損失である。とは言え、考えられる全ての項目を手動で入力するのはコストが高い。

キーワード以外の検索でも、記述に関するコストは同様にかかる。先に紹介した amazon の「スタイル」は、『カジュアル』『エレガンス・キャリア』等 8 種類であり、さほど手間がかからないようにも思える。しかし、最大の『カジュアル』を選ぶと 7,739 件もヒットするため、サイト側の何らかの努力～例えば、アイテムのテイストに合わせて、『甘カジ』とか『大人カジュアル』等のキーワードを埋め込む～なしには、自ブランドのアイテムがユーザの目につくことはなく、そしてそのアイテムは売れないと考えるべきであろう。

本研究はこのような新たなキーワードを如何に追加するか、そして、この追加を如何に自動化するかについての研究である。この目的のため、ファッションアイテムの商品紹介文に注目する。

商品紹介文は、サイトの広報担当により、アイテムの特徴がコンパクトに表現されている。そこには、「今流行りのストライプで…」 「着やせ効果抜群の…」 等、現在のトレンドや着用効果も含む様々な情報が記述されている。このため、紹介文を適切に分類し、加工できれば、先述の埋め込みも容易になるわけである。

上記を目的として、黒澤ら(2014)は商品紹介文から構成要素の自動分類実験について報告している。しかし、実験がクローズドデザインで行われているため、今回はオープンテストでの実験を行う。さらに、ブランドによって学習結果が異なるかという観点を新たに付け加える。

## 2. 関連研究

## 2.1. アイテム・コーディネート推薦

アイテム推薦のために、構成要素に着目した研究がある。高寺ら(2000)は、アイテムに素材、細部形状等、構成要素について詳細なタグ付けを行い、感性検索システムの構築を試みている。しかし、このシステムは人手による詳細なタグを入力する必要がある。したがって、先に述べたコストの問題点に直面する。

コーディネート推薦でも、構成要素に着目した研究がある (Liu ら 2012, 神間ら 2011, 福田ら 2011, 佐藤ら 2009, 辻田ら 2009)。神間ら (2011) では、色やディテール (袖の長さや形等) などに着目している。ただ、これらの項目は固定であり、また、漏れなく用意されているとは限らない。パフスリーブがあっても、ドルマンスリーブがなければ、最新のトレンドに即した提案はできないことになる。固定された項目ではなく、拡張可能な枠組みが必要である。先述の通り、商品紹介文はその時々のトレンドを反映させているので、トレンドに合った推薦が可能となるはずである。

## 2.2. 分類のためのファッションアイテム構成要素

アイテムの構成要素として、如何なる要素が考えられるのだろうか？ 黒澤ら(2014)では、様々なファッションに関する文献(鈴木ら 2005, 富田 2004, 小林 2003, 工藤ら 2001, 飯塚ら 1985) から、14種類の構成要素を定義した。しかし、実際のアノテーション作業に伴い、明確でない要素が発見された(例えばコーデ。アイテムの組み合わせとアイテム単独での着用法が混在していた)。そこで、今回はコーデを「コーデ」と「着こなし」に分離する等改良を加え、足りない部分を補い22種類とした(表 1)。

## 3. 実験と考察

本研究は分類器としてナイーブベイズを用いる。シンプルで高速かつ高精度な分類器だからである。また、本研究とは対象は異なるものの、同様のショッピングサイトを対象とし、有効性を確認している研究があるからである(岩井ら 2013)。

### 3.1. アパレルデータ

次に、本研究で使用したデータについて述べる。

#### 3.1.1 要素アノテーション

まず、ファッションアイテム販売サイトから、9ブランド 3,191 アイテム 10,557 文の商品紹介文を収集し、先に述べた構成要素のどれに該当するか(1文に複数の要素の付与を許す)、4人の作業員により5,000 文程度ずつ手でアノテーションが行われた。この際、1文は必ず2人で評定を行い、一致しない場合には3人目の作業員が調停や訂正を行った。

この結果、出現数が100程度の構成要素は、自動分類課題には不適当と考え、除いた。また、「デザイン」は出現数が多いものの、作業員間の不一致が相当地に大きかったため、今回は対象から省いた<sup>1</sup>。内訳を最終頁表 6に示す。

#### 3.1.2 形態素解析

次に、MeCab<sup>2</sup>を用いて形態素解析を行った。形態素情報について最終頁表 7に示す。以下、このアノテーションを行った形態素解析済みデータを、アパレルデータと呼ぶ。

<sup>1</sup> 狭義のデザイン(飯塚ら 1985)を想定していた。しかし、文中に出現する『デザイン』という用語に惑わされたようであった。

<sup>2</sup> <http://code.google.com/p/mecab/>

表 1 アイテムの構成要素とその説明

素材	アイテムを構成する素材 繊細なオーガンジーでデザインした…
色	アイテムの色 爽やかな色使いが女性らしく…
柄	アイテムが持つ柄、もしくはプリント インパクトのあるストライプが目目を惹く…
技術	アイテムの形成に必要とされる技術 刺繍レースをあしらった…
シルエット	アイテムの外形 裾にかけてAラインになった、ゆったり…
ディテール	アイテムの細部 大きなポケットやステッチ使いが…
デザイン	構造上の特徴に関する記述 シンプルなトップスですので、いろいろな…
流行	アイテムまたは現在のトレンド、逆に定番 今年注目のデザインは…
コラボ	デザイナー等との共同制作や、キャラクターとの共演 ○○ちゃんとのコラボ商品です。
コーデ	アイテムの合わせやすさ いろんなボトムスと好相性なので…
着こなし	アイテムの着用法 羽織ったり、腰に巻いたりするも…
着心地	アイテムの着心地・履き心地 気軽に着られて肌触りも抜群な…
印象	アイテムの持つ印象 ワンランク上のカジュアルスタイルを…
効果	アイテムが与える効果 細身のシルエットで美脚効果抜群…
安心・不安	透け等への不安と、その対策 ベチパンツが付属していますので…
季節	アイテムを使用する季節 まだまだ暑い今時期にぴったりの…
目的	アイテムの使われるオケーション シンプルなので、お仕事用にも…
サイズ	アイテムのサイズ 大きめですのでゆったり着いただけます。
年齢	対象の年齢層に関する記述 大人めに着ていただけます。
取り扱い	アイテムの取り扱いに関する記述 シワになりにくい素材でできていますので…
原産国	アイテムの生産国に関する情報 イタリア産の糸で織られた記事を使用して…
性別	アイテムの性別を表す記述 シンプルなデザインですので、ユニセックスで…

### 3.2. 実験 1

leave-one-out cross validation により実験を行う。

アパレルデータ中の自立語から頻度データを作成し、ナイーブベイズを用いた実験を行う。ただし、「これからの季節に◎」のように、1部の記号には意味があるため、データに加えた。また、「色」「柄」のように、直前の語によっては「名詞・接尾一般」になる語も自立語扱いとした。

本研究は手続き上、1文に対し複数の構成要素を許しているため、それぞれの構成要素で学習を行い、全14種の分類器を作成した。本研究はナイーブベイズの分類器として broomie<sup>3</sup>を用いた。

<sup>3</sup> [http://code.google.com/p/broomie/wiki/broomie\\_tutorial\\_ja](http://code.google.com/p/broomie/wiki/broomie_tutorial_ja)

なお、表 6 から明らかなように、要素間の正事例出現回数が不均等であり、回数が多い要素の影響を受けやすい。このため、要素ごとに同数の負事例を用意する。「素材」では正事例が 3,239 件あるので、これと同数の負事例 3,239 件を「素材」を要素に持たない文からランダムに選び、6,478 文で実験する。ただし、負事例の選び方による偏りを減ずるため、試行 10 回の平均を求める。結果を最終頁表 8 に示す。なお、表中「無操作」は形態素解析結果に手を加えていないこと、一方、「表記ゆれ」は表記の統一を試みていることを示す。ブランドによって裾を『スソ』と記すなど不一致が見られたためである。

表 8 から「着こなし」の精度が低く、「効果」の再現率が低いという点も見られる。しかし、多くの項目で 90% を超える精度を得ており、手法は有効に機能していると考えられる。

### 3.3. 実験 2

#### 3.3.1 ブランド間の学習傾向の違い

次に、ブランド間の違いについて検討を行うため、9 ブランドのうち 8 種類を学習に用い、残った 1 ブランドでテストを行う。この方法が有効であれば、全く新しいブランドのデータに対しても本研究の手法が機能することを意味する。先程同様、負事例を同数用意し、10 回繰り返した結果の 1 部を表 2~表 4 に示す。

表 2 構成要素「素材」

		ブランド								
		A	B	C	D	E	F	G	H	I
正	精度	97.4	97.2	97.3	97.6	97.6	97.4	97.4	97.1	97.8
	再現率	76.2	76.4	76.3	76.7	76.7	75.8	77.0	76.3	75.9
	F	85.5	85.6	85.5	85.9	85.9	85.3	86.0	85.5	85.5
負	精度	80.4	80.6	80.5	80.8	80.8	80.2	81.0	80.5	80.3
	再現率	98.0	97.8	97.9	98.1	98.1	98.0	97.9	97.8	98.4
	F	85.5	85.6	85.5	85.9	85.9	85.3	86.0	85.5	85.5

表 3 構成要素「印象」

		ブランド								
		A	B	C	D	E	F	G	H	I
正	精度	80.6	85.4	87.2	87.5	87.3	88.5	89.2	89.7	87.6
	再現率	76.6	73.5	74.2	74.9	75.6	74.5	75.4	75.6	66.6
	F	78.5	79.0	80.2	80.7	81.0	80.9	81.7	82.1	75.7
負	精度	77.7	76.8	77.6	78.1	78.5	77.9	78.7	78.9	73.1
	再現率	81.5	87.5	89.1	89.4	89.0	90.3	90.9	91.3	90.6
	F	78.5	79.0	80.2	80.7	81.0	80.9	81.7	82.1	75.7

表 4 構成要素「効果」

		ブランド								
		A	B	C	D	E	F	G	H	I
正	精度	74.7	74.0	71.2	73.8	70.2	72.4	72.1	75.7	61.0
	再現率	36.9	33.9	33.9	33.7	31.8	32.7	36.2	37.7	18.6
	F	49.4	46.5	45.9	46.2	43.7	45.0	48.1	50.3	28.5
負	精度	58.2	57.2	56.7	57.1	55.9	56.6	57.5	58.6	52.0
	再現率	87.6	88.0	86.3	88.1	86.6	87.6	86.1	87.9	88.1
	F	49.4	46.5	45.9	46.2	43.7	45.0	48.1	50.3	28.5

表 5 単語削除後の構成要素「効果」

		ブランド								
		A	B	C	D	E	F	G	H	I
正	精度	79.4	78.8	79.0	77.9	74.7	77.1	79.3	78.9	59.9
	再現率	71.6	68.1	67.3	70.4	64.6	64.8	69.2	72.3	29.4
	F	75.2	73.0	72.6	73.9	69.3	70.3	73.9	75.4	39.5
負	精度	74.2	72.0	71.6	73.1	68.9	69.7	72.7	74.5	53.2
	再現率	81.5	81.7	82.2	80.0	78.2	80.7	81.9	80.7	80.3
	F	75.2	73.0	72.6	73.9	69.3	70.3	73.9	75.4	39.5

紙面の都合上、正例に対する考察のみ記す。

「効果」と「印象」を除く要素（例えば「素材」）では、表 2 に挙げるように精度の違いが、ブランド間でわずか 1 ポイント程度に収まっていた。このことは、多くの新ブランドでも本研究の手法が有効であることを意味する。

一方、「印象」では 10 ポイント程度の差が生じているため、適用には多少の注意が必要であることがわかる。それでも、最低のブランド A でも 80% を超えているため、大きな問題ではない。

最大の問題は「効果」であり、精度が 70% 程度であることに加えて、再現率が 30% 程度である点である（表 4）。また、ブランド I が飛びぬけて低い。

#### 3.3.2 「効果」の誤り要因の検討

7 割もが検出に失敗したため、その事例の確認を行った。その結果、「くれる」「いただく」等の動詞・自立としても動詞・非自立としても両方解釈可能な語の影響を受けていることがわかった。

また、「枚」が悪影響を与えていることもわかった。これは『名詞・接尾一般』を加えたことによる。

さらに、「◎」も問題となった。「◎」は多くの構成要素中で使えるため、『ペプラムが細見せに◎』のような文が「効果」として分類できなかったのである。

確認のため、これらの語を対象外とした実験結果を表 5 に示す。

表 4 と比較して、精度は落ちるところもある一方で、再現率の格段の上昇が改善点である。

ただし、ブランド I の精度・再現率ともに依然低いままであり、再検討が必要と考えられる。ブランド I は表 3 の再現率も低めである。一方、統計値（表 7）上は特に違和感はない。しかし、何か別の特徴を有しているかもしれない。再検討が必要である。

## 4. おわりに

本研究はナイーブベイズを用い、ファッションアイテムの紹介文を自動分類することを試みた。

leave-one-out cross validation の結果は有効であり、ブランド間の差異が大きな問題にならないことを確認した。ただし、「効果」の再検討は必要である。

さらに、これまでの課題では構成要素分類を中心としていたため、今後は分類結果から有効なルールを抽出し、そのルールに基づいたアイテム推薦やコーディネート推薦を行うことを試みたい。

### 謝辞

この研究は、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。

### 参考文献

福田未央, 中谷善雄(2011). “服の着用経験に基づき服自身がコーディネートに推薦するシステム”, 情報処理学会全国大会講演論文集, pp.161-163.  
 小池恵里子, 伊藤貴之, 渡辺知恵美(2013), “提示量を適正化した女性向け商品検索支援システム”, DEIM Forum 2013.  
 飯塚弘子, 内田八重子, 香川幸子(1985). 服飾デザイン論. 文化出版局.  
 岩井秀成, 池田郁, 土方嘉徳, 西田 正吾(2013), “レビュー文を対象としたあらすじ分類手法の提案”, 電子情報通信学会論文誌. D, 情報・システム J96-D(5), pp.1222-1234.  
 神間唯, 丸谷宜史, 梶田将司, 間瀬健二(2011). “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 2011-HCI-142(26), pp. 1-7.  
 小林茂雄(2003). 装いの心理. アイ・ケイコーポレーション  
 公益社団法人 日本通信販売協会(2013a). “第5回 インターネッ

ト通信販売利用実態調査 報告書”.  
 公益社団法人 日本通信販売協会(2013b). “第5回 全国通信販売利用実態調査 報告書”.  
 工藤勝江, 山本由紀子, 東陽子, 古御堂誠子, 野原美香(2001). 文化ファッション大系 ファッション流通講座⑤ コーディネートテクニック アパレル編 I (商品 知識), 文化出版局.  
 黒澤義明, 小川湧真, 竹澤寿幸(2014), “ファッションアイテム紹介文の構成要素自動分類”, 言語処理学会年次大会.  
 Liu, S., Feng, J., Song, Z., Zhang, T., Lu, H., Xu, C., and Yan, S. (2012), "Hi, magic closet, tell me what to wear!" in Proceedings of the 20th ACM international conference on Multimedia, pp. 619-628.  
 佐藤彩夏, 渡邊恵太, 安村通晃(2009). “姿を利用したファッションコーディネート支援システム suGATALOG の提案と評価”, 情報処理学会論文誌, 53(4), pp.1277-1284.  
 瀬古沢照治, 三橋宏行, 小澤幸夫(2008). “アパレルオンラインショッピングにおけるワン・トゥ・ワンのリコメンドシステム”, 電気学会論文誌 C, 128(8), pp.1333-1341.  
 鈴木洋子, 天野豊久 (2005). 文化ファッション大系 ファッション流通講座⑦コーディネートテクニック 演出編, 文化出版局.  
 高寺政行, 古川寅雄, 清水義雄, 上候正義, 細谷聡, 佐渡山亜兵(2000). “感性データベースを用いたアパレル製品 検索システムの開発”, 先進繊維技術科学に関する研究報告 平成 11 年度 成果報告 6: pp.67-68.  
 富田明美(2004). アパレル構成学 着やすさと美しさを求めて. 朝倉出版.  
 辻田眸, 北村香織, 神原啓介, 塚田浩二, 椎尾一郎(2009). “Asal-coordinator: 履歴情報を利用したファッションコーディネート支援.” ヒューマンインタフェースシンポジウム 2009 論文集, pp.85-88.

表 6 構成要素ごと出現回数

	素材	色	柄	技術	シルエット	ディテール	流行	コーデ	着こなし	着心地	印象	効果	季節	年齢
文数	3,239	1,116	1,593	1,284	2,973	2,489	1,861	3,336	1,093	1,061	5,017	1,467	1,577	514

表 7 形態素解析統計

	ブランド								
	A	B	C	D	E	F	G	H	I
文数	1,137	1,219	1,190	1,235	1,178	1,136	1,143	1,188	1,131
アイテム数	371	456	423	298	336	342	258	347	360
snt/item	3.06	2.67	2.81	4.14	3.51	3.32	4.43	3.42	3.14
形態素数	18,040	15,103	15,680	21,478	19,417	15,330	21,795	20,928	15,990
mrph/snt	15.87	12.39	13.18	17.39	16.48	13.49	19.07	17.62	14.14

表 8 leave-one-out cross validation による結果

		素材	色	柄	技術	シルエット	ディテール	流行	コーデ	着こなし	着心地	印象	効果	季節	年齢
無操作	精度	95.4	96.6	94.8	79.3	85.9	83.4	96.7	86.9	74.4	84.1	83.6	83.7	94.6	95.3
	再現率	73.7	82.7	83.9	80.1	75.9	84.3	88.5	79.0	72.7	80.0	77.2	64.5	88.8	94.9
	F	83.2	89.1	89.0	79.7	80.6	83.8	92.4	82.8	73.5	82.0	80.3	72.6	91.6	95.1
表記ゆれ	精度	96.0	97.6	96.9	85.9	89.1	84.5	96.3	88.4	75.5	84.6	84.3	81.0	93.1	93.7
	再現率	74.3	82.2	84.7	78.4	74.9	83.2	88.2	76.7	77.9	82.3	77.7	68.4	89.9	94.9
	F	83.8	89.2	90.4	82.0	81.4	83.8	92.0	82.1	76.7	83.4	80.9	74.1	91.5	94.3