

ファッションアイテム紹介文の構成要素自動分類

黒澤 義明 小川 湧眞 竹澤 寿幸

広島市立大学大学院 情報科学研究科

{kurosawa, ogawa, takezawa}@s.info.hiroshima-cu.ac.jp

1. はじめに

インターネットの発展により、多くのユーザがオンラインショップでの購買行動を行っている。公益社団法人 日本通信販売協会によるインターネット通信販売利用実態調査(2013a)によれば、ユーザの45.3%が「本・雑誌・コミックス」を購入しており、以下「食品/飲料(アルコール類をのぞく)」39.9%、「レディースファッション・靴」26.8%と続く。

3位の「レディースファッション・靴」が、通販では1位となる点は興味深い(日本通信販売協会 2013b)。つまり、カタログなどによる購入者が一定数いることを意味している。

1つの可能性として、インターネットは短時間に感性に訴えかけることが困難という点が挙げられよう。自ら能動的に検索し、情報を取捨選択する必要があるからである。その一方で、カタログは必要な情報がコンパクトにまとまっており、受動的な情報のみで購買行動へ移行することができるという利点がある。

同様の考えは、女性の購買行動の曖昧さに着目してシステムの構築を目指した研究にも見られる(小池ら 2013)。彼らは、「受動的な商品検索システムが有効ではないか」との目的意識から、ユーザの「興味を引けるようにある程度の多様な商品群を一斉に表示…」するシステムを構築している。つまり、彼らのシステムは、アルゴリズムによる1つの選択肢が最適とする立場ではなく、ゆったりと曖昧に結びついた商品群を推薦する立場に依拠する。曖昧な結びつきを求めるユーザに、次に示すような遷移構造を持つサイトの増加も無関係ではなからう(図1)。

同図左ではアイテムの紹介だけでなく、そのアイテムを使ったコーディネートがいくつか表示されており、興味を持った写真からさらに別アイテムへ遷移できるようリンクが張られている。曖昧から具体へのリンクと考えられる。このリンクは存在するコーディネートに基づいて張られているのに対し、本研究は印象や錯覚といった実在しない媒介項を用いて、ゆるくリンクすること、そのための分類を行うことを目的とする。

たとえば、「脚長」等が挙げられる。鈴木ら(2005)はハイウエストやストライプの「脚長」効果を指摘している。そこで、ストライプスカートを検索したユーザに同等のアイテムを示す(図2)。

このような表現をする利点は、アイテム分類上、遠くにある商品を紹介できることにある。既存のシステムでは階層上近くの商品を推薦することになりがちである(図3左)。先のタイトスカートからは隣の(そして色合いの似た)プリーツスカートが選ばれる。その一方で、新たに「脚長」を媒介した図3右の表現からは、遠くの商品を推薦しうる。結果として、ワンピースとスカートとを同時に根拠を持って閲覧可能になる(図2)。これは通販カタログや雑誌によくあるように、1つのテーマをもとに構成されたページと似た効果をもたらす。同様に、「春のおでかけ」に着目したり、「まんねり撃退」に注目したり等、システムとしては曖昧に思えるリンクを張ることができれば、ユーザにとって有用なシステムとなると考えられる。

しかし、全てのリンクを手動で張ることは、相当なコストである。リンクを自動的に構築するため、本研究は、アイテム販売ページに書かれた商品紹介文に注目する。

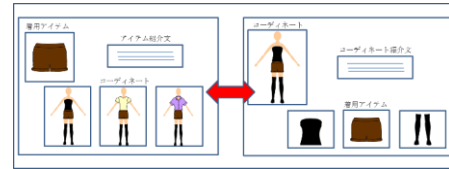


図1 アパレルサイトのページリンク(例)

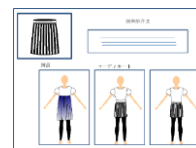


図2 “脚長”によるコーディネート紹介

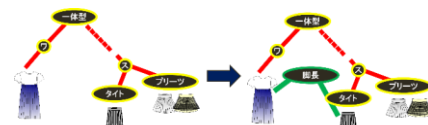


図3 システムの内部表現(イメージ)

紹介文は、サイトの販売担当による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、「今はやりのストライプで…」 「着やせ効果抜群の…」等、現在のトレンドや着用効果が述べられている。このため、分類することができれば、図2に挙げたようなページを、ユーザの嗜好に基づいた自動作成も容易となる。

そこで、本研究はレディースファッションのアイテムに関し、そのアイテムを紹介する紹介文の自動分類を試みる。

2. 関連研究

2.1. アイテム推薦

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、学習を行った上で感性検索を行うシステムを構築している。このシステムでは、「入力フォームを設計し…」とあるように、詳細なタグをすべて人手により付与する点が問題である。本研究はこの問題解消のため自動化を目指す。

瀬古沢ら(2008)は、階層分析法 AHP を用いて感性検索を行っている。ただ、アイテムの特徴記述を、たとえば「デザイン性」を9段階に評価することで行うため、構成要素そのものが異なる。

先に挙げた小池ら(2013)は、検索クエリーから(1つではない多くの)アイテムを表示するシステムを開発している。ただし、インタフェース及び提案のためのアルゴリズムに重点を置いており、構成要素については特段の議論はない。

2.2. コーディネート推薦

アイテム推薦の研究も多く存在する(Liuら 2012, 神間ら 2011, 福田ら 2011, 佐藤ら 2009, 辻田ら 2009)。今回は、構成要素に着目した研究について整理するととどめる。

表 1 構成要素比較

	共通する構成要素										独自	
	色柄素材	ディテール	シルエット	目的	流行	技術	コーデ	印象	効果	季節		
本研究	-	色柄素材	ディテール	シルエット	目的	流行	技術	コーデ	印象	効果	季節	着心地
神間ら(2011)	代表アイテム	色柄素材	ディテール(含む挿他)	シルエット	-	-	-	◇	◆	配色	-	-
Liuら(202)	アイテム(階層構造)	色柄素材	(襟、袖)	-	◆	-	-	◇	-	-	-	-
高寺ら(2000)	衣服	色柄素材	細部形状	-	対象用途	-	染色加工	-	-	-	-	生地

2.3. 分類のためのアイテム構成要素

本研究は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)を参考に特徴記述のための構成要素とした(表 2)。たとえば、飯塚ら(1995)による服装デザインの設計過程—形態の考察欄に「素材」、「技術」、「色彩」、「ディテール」があり、「技術」が独立している点などである。

他の研究と最も異なる点は「効果」である。情報推薦の立場からは注目を集めていなかった視点である。本研究は、分類がこの要素にうまく機能するか特に検討する。

3. 実験と考察

本研究はナイーブベイズ分類器を利用する。シンプルで高速かつ高精度な分類器であることに加え、対象は異なるとは言え、同様のショッピングサイトを対象とした研究でも有効な結果が得られていたからである(岩井ら 2013)。

3.1. アパレルデータ

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、複数のアパレルサイトから、1,325 アイテム 5,004 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手動でアノテーションを行った。なお、1 文に複数の要素の付与を許した。表 6 に出現数を記す。実験時にはこれらが正事例となる。

3.1.2 形態素解析

次に、MeCab¹を用いて形態素解析を行った。形態素中に「スカート」等のアイテム名を含む語を分割した。形態素「ストライプタイトスカート」は 3 形態素「ストライプ」「タイト」「スカート」に分割される。さらに、形容語に 3 拡張を行った。

1 つめは形容動詞の合成である。たとえば、「大人ガーリー」と「ガーリー」とでは微妙にニュアンスが違うことが予想されるため、1 形態素(大人ガーリー)とした。2 つめは形容詞の合成である。「かっこ」「の」「よい」を「かっこのよい」とする等、1 形態素とする操作を行った。3 つめは接尾辞である。動詞に続く場合(『着まわす』『やすい』を『着まわしやすい』等)と、名詞に続く場合(『女性』『らしい』を『女性らしい』等)がある。

以上の手続きを行い、収集した全 5004 文を形態素に分割した。全形態素数は 77,688 個であり、1 文当たりの形態素は 15 個程度であった。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。

表 2 採用したアイテムの構成要素と説明

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…
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…
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…
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…
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…
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	アイテムを持つ柄、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…
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…
印象	アイテムの持つ印象
	ワ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…
その他	上記に該当しない記述
	是非、ご利用ください…

3.2. 実験 1

アパレルデータ中の助詞等を除いた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。複数の構成要素を許しているため、各要素で学習を行い、全 14 種の分類器を作成する。本研究は broomie²を用いた。

なお、表 6 から明らかのように、アパレルデータは要素間の正事例出現回数がアンバランスであり、回数が大きい要素の影響を被る。このため以下の実験では、要素ごとに負事例を別に用意し、学習させる。たとえば「季節」ならば正事例が 382 個あるので、これと同数の負事例 382 個を、「季節」を要素に持たない文からランダムに選び、764 個のデータで実験する。ただし、偏りが生じる可能性を考慮し、試行を 10 回繰り返した平均を求める。

精度と F 値を求めた結果を表 7 に示す。なお、分類結果の判定は、クロズドテストで行った。

データ数の多い「ディテール」と「印象」で若干低めの数値となっているものの、全要素で 80%を超える精度を得ている。辞書等が何もない実験での精度が高い点は評価できる。「ディテール」については、アパレル用語辞書等を用い、その語に重みを与える操作で精度向上が可能と予想されるからである。

一方、形容詞については適切な辞書が見当たらないという問題がある。そこで、次の実験 2 で精度向上を試みる。

3.3. 実験 2

3.3.1 実験 2-1

ここでは、大規模コーパス中の出現頻度をもとに、形容語辞書を自動作成し、この辞書を用いた分類実験を行う。本研究では「Yahoo!知恵袋データ(第 2 版)」を用いた(以下、Y データ)。

上記データには、2004 年 4 月～2009 年 4 月に Yahoo!知恵袋³に投稿された質問(約 1600 万)と回答(約 5000 万)が収録さ

¹ <http://code.google.com/p/mecab/>

² http://code.google.com/p/broomie/wiki/broomie_tutorial_ja

³ <http://chiebukuro.yahoo.co.jp/>

れている。また、質問はカテゴリに分類されているため、対象を限定しやすいという利点がある。本研究はカテゴリ「健康、美容とファッション」下のサブカテゴリ「ファッション」に分類されている投稿に注目する。

ここで「ファッション」中に現れやすく、他のカテゴリに現れにくい用語は、ファッションに関連性の高い用語と考えられる。したがって、「ファッション」内の語と他のカテゴリに出現する語の比を求めることにより、有効な指標となると考えられる。

ただし、Yデータはサイズが大きく、形態素解析を行うコストが大きい。そこで簡易のため、今回はデータに対して検索を行い、出現回数を形態素の出現頻度とする。

検索は、形容詞については語幹が3文字以上である場合に語幹をクエリーとし、2文字の場合には、基本形とした。そして、形容動詞については、語幹が2文字以上を収集した。両者ともに、1文字の語については収集しなかった。

以上の手続きをアパレルデータ中に出現する464種の形容詞・形容動詞に適用する(以下、形容語)。形容語の出現頻度をカテゴリ「ファッション内」約27万文の質問とそれに対応する応答と、それ以外との出現頻度を計数した。その後、両者の比を求めた。以下に、出現比の上位下位10件ずつを示す(表3)。

実験はこの出現比に基づき、対象を限定し、重みに変更を行う。今回は0.5以上の値を持つ51語に重みを付加する一方、0.01より小さい78語を一律0.1倍にする。実験結果を記す(表7)。

実験の結果、要素「印象」については、わずかに0.7ポイントの上昇が見られるだけである。また、他の構成要素についてもほとんど変わらない値を示している。この点については後述する。

3.3.2 実験2-2

実験2-2では、464の形容語の重みを×1、×5...と変更して実験を行う。実験の結果を表7に示す。

表3 Yデータ中の形容語の出現比

形容語	出現頻度		出現比
	ファッション内	以外	
フレッピー	216	15	14.4000
大人カジュアル	67	5	13.4000
着回しやすい	48	4	12.0000
今年らしい	236	33	7.1515
着こなしやすい	64	19	3.3684
メンズっぽい	74	25	2.9600
秋冬らしい	18	7	2.5714
大人ガーリー	5	2	2.5000
大人可愛い	73	30	2.4333
合わせやすい	3316	1538	2.1560
⋮	⋮		
忙しい	410	164459	0.0025
見逃せない	2	1125	0.0018
パワフル	15	8612	0.0017
ヘルシー	16	10816	0.0015
ソフ	1905	1402352	0.0014
幸福	45	37282	0.0012
リズムカル	2	1686	0.0012
キャッチー	2	1875	0.0011
ボクシー	1	1060	0.0009
おいしい	273	487935	0.0006

要素「印象」の精度は82.4%から91.5%に上昇しており、一定の効果があると言える。しかし、F値が下がっているため、再現率の値が悪く、漏れが多いことを意味する。

3.3.3 実験結果の比較

以下に、判定した文数の比較を行う(表4)。

表4 実験ごと判定文数

人の分類	印象	機械の判断					
		実験1		実験2-1		実験2-2	
		印象	負	印象	負	印象	負
印象	1593.5	375.5	1602.3	366.7	1164.3	804.7	
負	339.5	1629.5	333.9	1635.1	119.2	1849.8	

※ 実験2-1, 2-2は重み10倍

実験1と実験2-1との比較では、わずかに10文程度(1593.5→1602.3)が精度向上に貢献している。一方、実験1と実験2-2では、大きく減少している(→1164.3)。

出現比の効果はわずかではあるものの確実に正事例を検出する。しかし、単なるリストでは精度を上げることに貢献しても、漏れを生むことになる。

3.4 実験3

実験1・2共に、「効果」の値(特にF値)は良いとはいえない。しかし、「効果」については本研究が独自に設定した要素であり、さらに精度向上を目指した実験を行う。

「効果」に含まれる文は、典型的には『ペプラムがお腹をカバーします』のように、効果を表わす動詞とその対象が必要となる。したがって、ナイーブベイズのような語の組み合わせを考えない手法では限界がある。

そこで、単純な手法として、3gramをナイーブベイズへの入力とする。すなわち、形態素の組「ペプラム-お腹-カバー」のような組を用意し、その出現頻度を計数する。

3gram化した結果、頻度が2回に満たない組を除いたところ、正事例779、負事例734となった。

表5 正誤分類結果

人の分類	印象	機械の判断	
		印象	負
印象		745	34
負		0	734

表から明らかのように、分類器が「印象」と判断した文はすべて正しいことがわかる。その一方で、34文については機械の判断が正しくない。「アイテム名-アイテム名-アイテム名」が「効果」と評価される等、組み合わせについての知識がないことによる悪化が確認された。この点、係り受け等による制約により、必要な形態素組の数を減らすことができ、一層の効果が期待できると考えられる。また、このような制約により、3gramを使うと組み合わせ数が大きくなる問題も解消できる。この実験では、約2,000の形態素で記述できた入力が15万を超える形態素組が必要となっている。多数のアイテムを扱う際には、減数が必至であろう。

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。Yahoo!知恵袋データを用いて構築した形容語による重みづけの結果、安定した結果を得ることができた。また、要素「効果」については、3gramを利用した手法により、高い精度が得られることを確認した。

今後の課題としては、形容語として形容詞と形容動詞以外について検討することであろう。ファッションの商品紹介文の中には、オノマトペも一定数含まれているし、形容動詞以外にも印象に関連した名詞は多数ある。これらについても出現比を計算し、詳細な実験を行いたい。

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。また、本研究の実施に際し、ヤフー株式会社から国立情報学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

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.
 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). “Asa1-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 アパレルデータ内の要素ごと文数

	季節	着心地	目的	流行	シルエット	ディテール	素材	色	柄	技術	コーデ	印象	効果	その他
文数	382	295	164	471	733	1731	1129	545	563	344	918	1969	931	375

表 7 実験 1, 実験 2-1: 重みごと精度・F 値

		季節	着心地	目的	流行	シルエット	ディテール	素材	色	柄	技術	コーデ	印象	効果	その他
精度	重み 1	92.3	89.4	88.3	93.3	93.9	81.6	93.7	95.2	93.8	93.0	89.9	82.4	86.5	82.8
	5	92.3	90.5	88.3	93.3	93.9	81.7	93.7	95.0	93.8	92.9	89.8	82.6	86.5	83.1
	10	92.3	91.4	88.2	93.3	93.8	81.6	93.6	95.0	94.0	92.7	89.8	82.8	86.4	83.4
	20	92.2	92.9	88.3	93.3	93.6	81.4	93.6	95.0	94.5	92.6	89.8	83.1	86.5	83.6
F 値	重み 1	90.3	88.2	87.7	91.7	83.4	82.7	89.3	89.4	87.2	88.4	79.7	81.7	77.3	70.0
	5	90.3	88.2	87.7	91.8	83.3	82.7	89.3	89.7	87.0	88.3	79.8	82.1	77.4	69.1
	10	90.3	87.9	87.6	91.7	83.3	82.7	89.2	89.7	86.8	88.2	79.9	82.1	77.4	68.2
	20	90.4	87.2	87.4	91.7	83.3	82.6	89.1	89.7	86.0	88.3	80.1	81.9	77.6	66.2

表 8 実験 2-2: 重みごと精度・F 値

		季節	着心地	目的	流行	シルエット	ディテール	素材	色	柄	技術	コーデ	印象	効果	その他
精度	重み 1	92.3	89.4	88.3	93.3	93.9	81.6	93.7	95.2	93.8	93.0	89.9	82.4	86.5	82.8
	5	91.0	91.0	91.4	92.8	92.0	80.5	90.8	93.8	91.5	91.6	87.6	88.5	84.8	83.2
	10	89.4	91.7	91.9	91.5	90.3	79.2	89.4	93.1	89.4	89.9	86.3	90.7	83.3	83.0
	20	86.5	92.0	92.4	90.2	88.4	77.7	87.6	91.7	87.2	87.5	85.1	91.5	81.8	83.1
F 値	重み 1	90.3	88.2	87.7	91.7	83.4	82.7	89.3	89.4	87.2	88.4	79.7	81.7	77.3	70.0
	5	90.6	87.1	84.4	91.2	84.1	81.2	87.5	89.6	85.6	88.1	80.5	76.9	76.2	70.7
	10	90.3	85.9	81.3	90.3	84.3	79.6	86.4	89.2	84.4	87.5	80.4	71.6	75.8	71.4
	20	88.8	84.2	79.0	89.0	84.2	77.9	85.1	87.6	83.2	86.8	79.9	67.4	75.0	73.0

※ 重み 1 が実験 1 に相当