

## 深層学習正解ラベルとしてのアパレルアイテム名利用とその有効性の検討

和泉 沙織 黒澤 義明 目良 和也 竹澤 寿幸  
 広島市立大学 情報科学部 広島市立大学大学院 情報科学研究科  
 [izumi, kurosawa, mera, takezawa]@ls.info.hiroshima-cu.ac.jp

### 1. はじめに

これまでファッションアイテムの検索には検索キーワードが用いられてきた。その際、アイテム名や商品紹介文等のテキスト情報が重要であった。しかし、商品の特徴全てを記述することはできないため、言語化されなかった情報～例えばアイテム画像に写っていたとしても～にはアクセスできず、アイテムが醸し出す印象等の主観的な検索は困難であった。

ところが、深層学習の利用により、画像から様々な対象を自動的に識別することが可能になってきた。一般物体認識の分野では、その認識能力がすでに人間を超えるほどになっている (Heら 2015)。

こうした研究の認識課題は、クラス間の境界が明確で排他的なクラス分けが一般的である。本研究が対象とするファッションアイテムでも、パンツ、スカート、トップスのような明確なカテゴリ分類であれば 9 割程度の精度で適用できるとの報告がある (堀ら 2016)。

その一方で、ファッションアイテムには排他的なクラス分けが困難な要素があるのも事実である。

例えば、構成要素『シルエット』(黒澤ら 2015)に着目すると、スカートには「タイト」「コクーン」という似たようなシルエットがあり、後者が「**より**丸みを帯びた」特徴を持つ。同様に、長めのスカートには「ミモレ」「ロング」「マキシ」と言ったシルエットがあり、後の方ほど「**より**長い」特徴を持つ。あくまで、「**より**」という相対的な関係であり、これらの間に明確な境界線は存在しないと言える。

先述のように、「タイト」「コクーン」に明確な境界は存在しない。しかし、両者が与える印象が微妙に異なるのも確かであり<sup>1</sup>、最終的にはこうした違いを検出したいと考えている。

今回、微妙な差異の検出の基礎とするため、我々の関心を 2 点挙げる。

まず、『このような特徴を深層学習は有効に扱えるか?』という点である。確かに、人間の側で特徴量を留意しなくてもよいという点が深層学習の利点で

はある。しかし、微妙に異なる相対的な情報は画像から得られるのだろうか?

2 つめは、『Web 上の命名は分類のための正解ラベルとして機能するか?』という点である。

深層学習には大量のデータが必要である。本研究の課題の場合、「画像とそのクラス」をセットで用意する必要がある。web 上のデータがそのまま利用可能であれば研究者が正解ラベルを用意する必要がなくなる。その一方、統一基準もない、精度の低いラベルという危惧もある。このため、web 上にある既存のデータを用い、その精度や限界を知っておくことは重要であると考えられる。

本研究は以上のような観点から、web 上に存在するアパレルアイテム画像の分類/検討を行う。

### 2. アパレルアイテムとアイテム名

アパレルサイトには、アイテムのページに「アイテム名」と「アイテム紹介画像」と「商品紹介文」がある。ここで「アイテム名」は、メーカーまたはブランドの担当者が記述しているため、商品の特徴を反映していると考えられる。

#### 2.1. アイテム名とアパレルアイテム構成要素

本研究は上記各種の特徴のことをアパレルアイテム構成要素と呼ぶ。黒澤ら (2015) は「素材」「柄」「シルエット」「印象」等の要素を挙げている<sup>2</sup>。

この構成要素ごとに分類の難易度が変わると考えられる。以下、代表的な例を説明する。

#### 2.2. 構成要素と明確な境界線の有無

要素間で明確に境界線が引ける要素と、あいまいな要素がある。難易度はこの区別によって異なる。

##### 2.2.1 素材

「素材」に関しては、『麻』『綿』と言ったように、排他的分類となる。しかし、画像には現れにくい特徴であるため、本研究の対象とはしない。

##### 2.2.2 柄

個々の柄については、「チェック」と「ストライプ」

<sup>1</sup> 商品紹介文中に「…コクーンシルエットで大人の女性らしさを演出…」等の記述が見られることから明らかである

<sup>2</sup> 鈴木ら(2005)、富田(2004)、小林(2003)、工藤ら(2001)、飯塚ら(1985)の研究を整理して 22 種類にまとめた。

は明確に分離できる等、それぞれ排他的な命名が可能である<sup>3</sup>。しかし、下位分類がある場合に、分類同士の境界が不明瞭になる場合がある。例えば、「花柄」と「小花柄」では、どこからが『小』なのか明確な基準は存在しない。しかし、「花柄」と「小花柄」では与える印象が異なると考えられる。したがって、「柄」について検討する。

### 2.2.3 シルエット

形状や丈に関する情報を持ち、「タイト」「ミニ」等、アイテム名中に頻出する。

形状と丈を1種類ずつ選ぶことが可能である。

丈に関して言えば、それぞれの名称を数直線上に並べ、相対的な位置関係は知ることができる（「ミモレ」<「ロング」<「マキシ」）。しかし、先述の通り、明快な境界は存在しない。この点、本研究の目的に合致するため、検討を試みる。

### 2.2.4 印象

主観的な要素であり、複数人の評価者がいた場合に統一した分類が可能かわからない要素である。このため、深層学習の正解ラベルがもっとも付与しにくい要素とも言える。

アイテム紹介文中のキーワードに着目することにより「印象」抽出する試み（黒澤ら 2017）もあるものの、現状では抽出数に難があり、深層学習に必要なアイテム数を用意できないため、今回は「印象」を対象とはしない。

## 2.3. コーディネート推薦研究との関連

構成要素を用いてコーディネートのおすすめを行う研究は多数ある（Liuら 2012, 神間ら 2011, 福田ら 2011, 佐藤ら 2009, 辻田ら 2009）。例えば、神間ら（2011）では、柄やシルエットの情報も利用している。本研究の利点は、先行研究があらかじめ用意した要素しか適用できないという点を解決し、自動構築できるようになるという点である。

また、本研究同様アパレルアイテムの深層学習を用いて分類を試みた研究（堀ら 2016）と比較すると、彼らの研究が「スカート」「パンツ」と言った『カテゴリ（サブカテゴリ）』に分類することに重点を置いているのに対し、本研究は「構成要素」に着目しているところが異なる。「花柄がコーデに華やかなアクセントをプラス」等のように、紹介文を利用した

<sup>3</sup> ただし、柄は重ねることができるため、アイテム名に複数の

印象評価への適用が容易である点が、本研究の利点である。

## 3. 実験と考察

ここでは、アパレルアイテム画像のクラス分類実験を行う。前述のように、構成要素「柄」と「シルエット」について、それぞれ実験を行う。トップスには「シルエット」の情報は少ないため、今回はボトムス（スカートとパンツ）のみを対象とする。

### 3.1. アパレルアイテム名とアイテム画像

アパレル販売サイトより、アイテム1着につき画像1枚とhtmlファイルをダウンロードした。

#### 3.1.1 アイテム名

アイテム名については「○○11月号掲載」等の直接関係ない情報が記述されることがある。また、略語、英語綴りがなされることがある。まず、整形のための処理を行った。

その後、「柄」辞書、「シルエット」辞書を用意し、アイテム名に同辞書のエントリが含まれるアイテムを実験対象とした。

ただし、分類クラスごとのデータ数にばらつきがあったため、閾値(600)を設け、それ以上のアイテムが用意出来る場合にはオープンテスト（学習500枚とテスト100枚、それぞれ別）、出来ない場合にはクローズドテスト（テスト用100枚は学習用に含まれる）での実験を行うこととした。実験データについて表1と表2に示す。黄色に塗られたクラスが、クローズドテストで実施されるクラスである。

#### 3.1.1 アイテム画像

500枚は深層学習に十分な量ではないため、data augmentationを行った。1枚の画像を256x256のサイズに変更したうえ、透視変換(x5)と回転(x3)により画像数を15倍にした。

堀ら（2016）では、画像からアイテムを中心とした領域を切り取る「画像の整形」という処理を行っている。しかし、本研究はそのような処理をしていない。したがって、アイテムだけが平置きされた画像や着装画像など、アイテム画像は様々である。

### 3.2. 深層学習フレームワーク

本研究は深層学習のフレームワークにcaffeを用

柄名（例えば、「ドットチェック」）を持つことはある。

いた (Jia ら 2014). また, 16 層の畳み込みネットワーク (VGG16 (Simonyan 2014)) を用い, 転移学習を行った.

前節に述べた data augmentation に加え, caffe の機能による. 鏡像(x2)とトリミング(x10)による画像増も行った. なお, トリミングの結果, ネットワークへの入力画像サイズは 228x228 となっている. 実験結果を表 1 と表 2 に示す.

### 3.3. 考察

#### 3.3.1 「柄」について

表 1 は全体的に高い数値を示している. また, 表 2 に比べても高い. 初期の予想通り, 区別のしやすい構成要素であり, こうした要素の存在は本研究の 1 つめの関心を満たすと言えよう.

次に, 表 3 の confusion matrix を使って考察する. 対角の色が濃いほど適切な分類であること, それ以外の着色された部分が不適切な分類であることを示している.

色合いに着目すると, 「チェック」「タータンチェック」「ストライプ」「ピンストライプ」のあたりが色づき, 相互に誤認識をしていることを示している.

実際の画像を目視してみると, 実際には下位の分類 (「タータンチェック/ピンストライプ」) なのに, 上位の分類 (「チェック/ストライプ」) として命名されていることがわかった. 実際, 商品紹介文中には下位の分類が正しく記載されていることもあった. この点~本研究の 2 つめの関心~, アイテム名だけに着目した正解ラベル付与が不十分であり, 改良が必要であることを意味している.

#### 3.3.2 「シルエット」について

実験の結果 (表 2), 「スキニー」と「ショート」の結果が 8 割程度と高くなっている. これは, 両者に似たクラスがないのが原因と考えられる. 2.2.3 で『形状と丈を 1 種類ずつ...』と述べたものの, 「スキニー」はフルレングスのパンツであることが多く, 丈の情報がほぼいらぬこと, また, 「ショート (パンツ)」は形が画一的であることが理由である.

表 4 では先の分析同様, 上位下位関係による誤認識 (例えば「ワイド/タックワイド」) がいくつか見られる. ただ, 多くは形に関するところで, 長さは「ロング/マキシ」だけとなっている. 「ミニ」は隣り合った「ミディ」と誤認識されることは少なく (1 個), 形

の「台形」と間違えることが多い (11). 一方, 「ミディ」を「ミニ」と誤認識することも少なく (2), 「ギャザー」 (13), 「タックフレア」 (18) と誤認識していることもわかる. この点, 長さの情報があまり学習に利用されていないことを示唆させる. 利用する画像を着衣画像にだけに絞るとどのような結果になるか, 今後検討が必要である.

### 4. おわりに

本研究は, web 上のアパレルアイテム画像とアイテム名から深層学習を用いた自動分類を行うことを試みた. 特に, 一般物体認識とは異なる, あいまいなクラス分類が可能かについて, あいまいさの程度を分類し, 結果の分析を試みた. 複数の課題が露見したので, 今後はその解決を試みたい.

### 謝辞

この研究は, 国立研究開発法人科学技術振興機構 (JST) の研究成果展開事業「センター・オブ・イノベーション (COI) プログラム」. 及び広島市立大学特定研究費 (先端学術研究費平成 26 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を推薦するシステム”, 情報処理学会全国大会講演論文集, pp.161-163.
- He K., Zhang X., Ren S., and Sun J. (2015). “Deep residual learning for image recognition,” arXiv:1512.03385, 2015
- 堀和紀, 坂原誠, 岡田将吾, 新田克己 (2016), “深層学習を用いたファッションコーディネート因子の抽出”, 人工知能学会全国大会
- 飯塚弘子, 内本八重子, 香川幸子 (1985). 服飾デザイン論. 文化出版局.
- Jia, Y., Shelhamer, E., and Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. (2014), “Caffe: Convolutional Architecture for Fast Feature Embedding,” arXiv:1408.5093
- 神間唯, 丸谷宣史, 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
- 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, 村上大志, 竹澤寿幸 (2015), “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”, 言語処理学会年次大会.
- 黒澤義明, 竹澤寿幸 (2017), “アパレルアイテム商品紹介文中の“演出”に着目した感性語の自動推定”, 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.
- Simonyan, K. and Zisserman, A. (2014). “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv:1409.1556, 2014
- 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, 上俣正義, 細谷聡, 佐渡山兵兵 (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 1 「柄」 についての使用データと結果

クラス名	データ数	train	test	pred	正解数	再現率	精度	F値
ボーダー	1,513	500	100	95	68	0.680	0.716	0.697
マルチボーダー	116	116	100	105	98	0.980	0.933	0.956
レオパード	326	326	100	110	88	0.880	0.800	0.838
迷彩	883	500	100	104	80	0.800	0.769	0.784
幾何学	376	376	100	118	90	0.900	0.763	0.826
ドット	647	500	100	101	71	0.710	0.703	0.706
千鳥	397	397	100	101	80	0.800	0.792	0.796
グレンチェック	372	372	100	97	79	0.790	0.814	0.802
ギンガムチェック	786	500	100	105	80	0.800	0.762	0.780
タータンチェック	131	131	100	82	69	0.690	0.841	0.758
チェック	4,307	500	100	88	51	0.510	0.580	0.543
ピストライプ	406	406	100	103	70	0.700	0.680	0.690
ストライプ	5,481	500	100	92	63	0.630	0.685	0.656
ベイズリー	240	240	100	109	94	0.940	0.862	0.900
ボタニカル	285	285	100	99	82	0.820	0.828	0.824
リーフ	178	177	100	94	80	0.800	0.851	0.825
花	4,628	500	100	104	64	0.640	0.615	0.627
小花	152	152	100	93	86	0.860	0.925	0.891

表 2 「シルエット」 についての使用データと結果

クラス名	データ数	train	test	pred	正解数	再現率	精度	F値
ギャザー	2,035	500	100	111	48	0.480	0.432	0.455
プリーツ	3,014	500	100	91	67	0.670	0.736	0.702
ラップ	1,228	500	100	116	58	0.580	0.500	0.537
スカーチョ	682	500	100	135	55	0.550	0.407	0.468
ガウチョ	3,761	500	100	64	31	0.310	0.484	0.378
ワイド	5,910	500	100	111	43	0.430	0.387	0.408
タックワイド	887	500	100	122	53	0.530	0.434	0.477
タック	4,954	500	100	59	17	0.170	0.288	0.214
タックフレア	947	500	100	150	54	0.540	0.360	0.432
フレア	5,786	500	100	68	30	0.300	0.441	0.357
テーパード	2,899	500	100	111	57	0.570	0.514	0.540
台形	679	500	100	108	75	0.750	0.694	0.721
タイト	5,615	500	100	58	32	0.320	0.552	0.405
ペンシル	631	500	100	119	70	0.700	0.588	0.639
スキニー	1,932	500	100	110	87	0.870	0.791	0.829
ミニ	720	500	100	92	55	0.550	0.598	0.573
ショート	5,294	500	100	96	85	0.850	0.885	0.867
ミディ丈	794	500	100	84	42	0.420	0.500	0.457
クロップド	1,570	500	100	94	56	0.560	0.596	0.577
マキシ	1,054	500	100	109	63	0.630	0.578	0.603
ロング	1,507	500	100	92	37	0.370	0.402	0.385

表 3 柄 についての confusion matrix

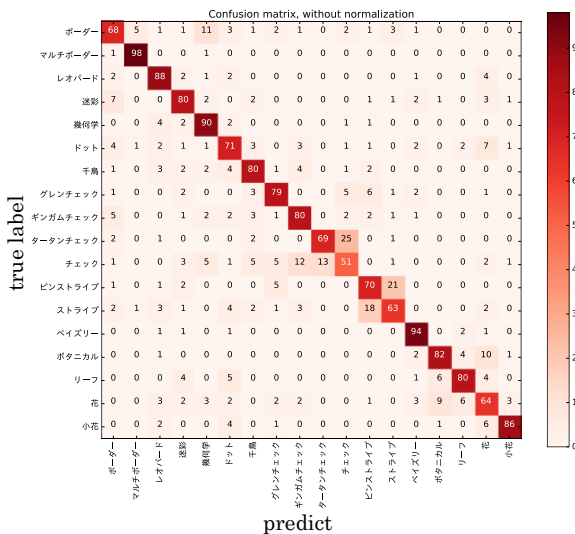


表 4 シルエット についての confusion matrix

