

深層学習を用いた服飾品自動分類の コーディネート推薦における有効性の調査

倉永将宏¹ 椋木雅之¹

概要：オンラインでの服飾品購入は便利だが、コーディネートに関するアドバイスは不足している。本研究では、深層学習で識別した服飾品に関する種々の情報が、コーディネート推薦においてどの程度有効であるか調査する。評価は SIGIR eCom'22 の服装推薦のコンテストを利用する。このコンテストで上位となった手法は、手作業で付けられた服飾品に関する属性情報を利用している。この属性情報を VGG16 で認識結果に置き換え、コーディネート推薦の精度を比較することで、深層学習による服飾品分類のコーディネート推薦における有効性を評価する。

キーワード：感性情報, 画像分類, e ビジネス

1. はじめに

オンラインでの購入では、24 時間どこからでも豊富な品揃えの中から衣服を選ぶことができ、利便性が高い。しかし、一般にオンラインショップでは店員が直接接客対応することは稀で、購入者は衣服選択、特に衣服の組み合わせ（コーディネート）に関するアドバイスを受けることができない。コーディネートの自動推薦システムがあれば、購入者の助けになるとともに、オンラインショップの販売促進にもつながる。そのため、コーディネート推薦システムに関する多くの研究が行われている。

コーディネート推薦システムでは、ユーザの過去のコーディネートを中心に推薦を行うものがある。例えば、長尾ら[1]は、季節や気温、天候、イベント毎のユーザの過去のコーディネートを記憶しておき、類似の状況で過去のコーディネートを推薦するシステムを提案している。森本ら[2]は過去のコーディネートにおける服装の色や柄、種類などを元に、ベイジアンネットワークを構築し、コーディネート推薦を行う。こうした手法は、ユーザへの個人適応が行えるが、十分なデータの蓄積に時間がかかる。一方、桂ら[3]は、オンラインショップに掲載されているコーディネート画像を元に TPO に相応しいコーディネートを推薦している。また、岩田ら[4]は、ファッション雑誌に掲載されているコーディネート画像により推薦を行っている。これらの手法では、推薦に値するコーディネートを既存のデータから自動で取得するため、一般性のある高品質な推薦が行えると考えられる。しかし、これらの手法では、推薦の前提となるユーザが持つ服飾品を手動で入力する必要があり、ユーザ側の立場からは手間がかかる。

この問題に対して、画像認識技術を使ってユーザの持つ服飾品の情報を自動で識別できれば、手間の解消につながる。近年、深層学習の技術が向上し、様々な分野の画像認識に活用されている。服飾品画像の認識についても多くの研究が行われている。それらの研究の多くでは、Fashion-

MNIST[5] と呼ばれるデータセットが使われている。Fashion-MNIST は、"T-Shirt" や "Trouser" といった 10 種類の服飾品に関する画像データセットである。例えば、VGG16 という深層学習の画像識別器に対して Fashion-MNIST で学習し、識別を行った場合、93.5%という高い識別結果が得られると報告されている[5]。

そこで本研究では、深層学習で識別した服飾品に関する種々の情報が、コーディネート推薦においてどの程度有効であるか調査する。深層学習手法としては、上述の VGG16 に基づく手法を利用する。コーディネート推薦の評価には、SIGIR eCom'22[6] として開催されたコーディネート推薦のコンテストを利用する。このコンテストにおいて、上位となった Selman らの手法[7]では、手作業で付けられた服飾品に関するメタデータを利用している。このメタデータを VGG16 で認識結果に置き換えた場合での、コーディネート推薦の精度を比較することで、深層学習による服飾品分類のコーディネート推薦における有効性を評価する。この際、VGG16 の学習にも eCom'22 で提供されたデータを用いる。

以下、2 章ではコーディネート推薦と服飾品自動分類の従来研究を紹介する。3 章では深層学習を用いた服飾品分類と評価の方法について述べる。4 章では実験を行い、5 章では結果のまとめを行う。

2. コーディネート推薦と服飾品自動分類の従来研究

2.1 コーディネート推薦

コーディネートとは、シャツやズボン、バッグ、靴などの服飾品の組み合わせ（服装）のことである。ファッションにおいては、個々の服飾品の良し悪しだけでなく、それらの組み合わせ方も重要とされている。服飾品の組み合わせ方は多数の可能性があり、その中から適切な組み合わせを選択することは容易でない。そのため、コーディネートを決める際に支援するコーディネート推薦が望まれている。

¹ 宮崎大学 大学院 工学研究科

コーディネート推薦は様々な観点から行われる。例えば、個人の趣味嗜好に適応した推薦[1][2]や、イベントの性質や参加者などの TPO に合わせた社会通念上適切な推薦[3]、ファッションの専門家がプロデュースする推薦[4]など様々な研究が行われている。本研究では、コーディネート推薦を、予め与えられた正解となるコーディネートに基づき、それと同種のコーディネートを生成する処理とする。正解となるコーディネートを様々に変えることで、様々な観点からの推薦が行えると考えられる。

2.2 コーディネート推薦コンテスト

コーディネート推薦はオンラインショップの付加価値にもなるので、服飾関連企業が協力したコンテストが盛んに行われている。

SIGIR eCom'22 は、情報検索分野の最重要国際学会の一つである SIGIR 内で、2017 年から毎年開催されているワークショップで、特に e コマース（電子商取引）領域を中心に扱っている。SIGIR eCom'22 では、コーディネート推薦に関するコンテストが行われた。このコンテストは、スタイリストとファッションの専門家によって作成された一部の服飾品がない状態の不完全なコーディネートに対して、正解の服飾品を候補リストの選択肢から選び、スコアを競うというものであった。

まず、服飾品 398,670 点について、画像とその服飾品に関するデータ（メタデータ）が与えられた。服飾品画像は、背景無地で服飾品のみが映っているカラー画像で、服飾品毎に 1 枚提供される。服飾品画像の例を図 1 に示す。メタデータには、その装飾品の ID やカテゴリ、色、男性向け/女性向け（ジェンダー）などの情報が含まれている。これらのデータは、衣服のオンラインショップを展開している FARFETCH 社が提供したもので、人手で作成した正しい情報になっている。

コンテストでは、コーディネートの学習用データ、incomplete outfit（問題）、missing item（正解）、candidates（選択肢）が与えられる。学習用データは、スタイリストとファッションの専門家によって作成された適切なコーディネートで、コーディネートを構成する服飾品の ID のリストにより構成されている。incomplete outfit は、学習用データと同様に作成された適切なコーディネートから服飾品を一つ除いたリストである。除いた服飾品を missing item（正解）とする。また、ランダムに選択された服飾品 8-40 個のリストに、取り除いた服飾品を加えたリストが candidates となる。コンテストでは、incomplete outfit（問題）が与えられた時、candidates（選択肢）の中から正解である missing item をどれだけ正確に選択できるかで競われた。この選択の際に、学習用データが利用できる。コンテストは、2022 年 5 月から 7 月にかけて実施され、12 件の応募があった。本研究では、このコンテストのデータを利用する。



図 1 SIGIR eCom'22 のコンテストでのデータセットの一部[7]

2.3 協調フィルタリングを利用したコーディネート推薦

コーディネート推薦の手法も、様々なものが研究されている。本研究では、Selman ら[7]が SIGIR eCom'22 に出場した際の手法である、協調フィルタリングを利用したコーディネート推薦を用いる。Selman らは、アイテムベースの協調フィルタリングとアイテム属性モデルという 2 つの手法を組み合わせる手法で、このコンテストで 3 位を記録している。

協調フィルタリング(Collaborative filtering)とは、たくさんのユーザの好みを情報として管理し、あるユーザと好みの類似した別のユーザの情報を利用して自動的に推測を行うことである。例えば、ユーザ A がアイテム X を好むとすると、アイテム X を好む別のユーザ B が好むアイテム Y を探し出し、ユーザ A もアイテム Y を好むのではないかと、という推論をコンピュータによって自動的に行う。

Selman らの手法では、まずアイテム（服飾品）ベースの協調フィルタリングにより候補を選択する。candidates の各アイテムと incomplete outfit の各アイテムが、学習用データの正解のコーディネート内で共起する回数を計算する。candidates のアイテムの中でこの共起回数が最大となったアイテムを候補とする。この時、候補の共起回数が 0 回または 1 回だった場合、その候補の信頼性は低い。そこで、その場合にはアイテム属性による協調フィルタリングに切り替える。アイテム属性とは、各アイテム（服飾品）のメタデータから、「カテゴリ-ジェンダー-色」の情報を取り出したものである。アイテム属性による協調フィルタリングでは、アイテムの共起回数の代わりに、アイテム属性の共起回数を計算し、最大となったアイテムを候補とする。これにより、学習用データの中に candidates と incomplete outfit の組み合わせのコーディネートが含まれていなくても、類似のアイテム属性を持ったアイテムのコーディネー

トから、候補を選ぶことができる。

2.4 服飾品の自動分類の従来研究

Selman ら[7]の手法は、服飾品のメタデータを利用して、カテゴリ-ジェンダー-色の3つの情報が必要である。したがって、実際にコーディネート推薦する際に、ユーザーに服飾品のこれらの情報を入力してもらうなどの手間がかかってしまう。そこで、服飾品の自動分類が望まれる。

服飾品の自動分類の従来研究では、Fashion-MNIST[5]という公開データセットが多く用いられている。Fashion-MNIST では訓練用に 60,000 枚、検証用に 10,000 枚の画像が提供されている。各画像は、 28×28 画素の濃淡画像である。また、メタデータとして T-shirt/top、Trouser、Pullover、Dress、Coat、Sandal、Shirt、Sneaker、Bag、Ankle boot の 10 カテゴリ分類の正解が与えられている。このデータセットは、深層学習のフレームワークである Pytorch などですぐに利用できるよう提供されており、VGG16 や GRU+SVM、MobileNet などの多くの画像分類の結果が公開されている[5]。例えば、深層学習を用いた画像分類手法である VGG16 を使った場合、93.5% の高い識別結果が得られるとされている。

3. 深層学習を用いた服飾品分類と評価

3.1 VGG16 による服飾品分類

SIGIR eCom'22 で上位となった Selman らの手法は、手作業で付けられた服飾品に関するメタデータを利用している。このメタデータを深層学習による画像分類結果に置き換えることができれば、実際のコーディネート推薦の場面でユーザーの入力の手間を減らすことができる。本研究では、実際に深層学習による画像分類を行い、その結果を利用したコーディネート推薦の精度を比較することで、深層学習による服飾品分類のコーディネート推薦における有効性を評価する。

本研究では、服飾品分類に深層学習による画像分類手法である VGG16 を利用する。VGG16 は、深さが 16 層の畳み込みニューラルネットワークであり、ImageNet と呼ばれる画像データセットで学習させた学習済みネットワークが提供されている。この事前学習済みネットワークは、画像をキーボード、マウス、鉛筆、動物など 1,000 個のオブジェクトカテゴリに分類できる。SIGIR eCom'22 で提供されているデータを利用して、VGG16 を学習し、服飾品分類を行う。

3.2 服飾品のカテゴリ分類

カテゴリは、シャツやズボンなど、服飾品（アイテム）の種類を表すものである。eCom'22 のデータでは、アイテムは 129 種類のカテゴリに分類されている。本研究では、VGG16 の学習に $K=10,000$ 枚の画像データを用いる。そのため、カテゴリ分類の対象クラスは、 $K=10,000$ 以上のアイテムを含んでいる以下の 13 クラスとする。

Tops, Trousers, Jackets, Dresses, Sweaters&Knitwear, Knitwear, T-Shirts&Vests, Tote-Bags, Shirts, Denim, Shorts, Skirts, Coats.

これらの 13 クラスで、398,670 個のアイテムのうち、323,686 個 (81%) のアイテムを含んでいる。

3.3 服飾品のジェンダー分類

ジェンダーは、そのアイテムが男性向け (MEN) か女性向け (WOMEN) か男女兼用 (UNISEX) かを表すものである。ジェンダー分類にも、 $K=10,000$ 枚の画像データを用いる。UNISEX はこの枚数に達しないため分類対象は、MAN と WOMEN の 2 クラスとする。

3.4 ジェンダー別の服飾品のカテゴリ分類

ジェンダーは 2 クラス分類であり、比較的精度良く分類が行えると考えられる。また、例えば男性向け (MEN) には Dresses のカテゴリのデータはほとんど存在しないなど、ジェンダーによってカテゴリに偏りがある。従って、ジェンダー別にカテゴリ分類を行うことで、カテゴリ分類の精度を高くできると考えられる。そこで、男性向け、女性向けそれぞれについて、カテゴリ分類を行う。学習は、 $K=3,000$ 枚の画像データを用いるため、男性向けのカテゴリ分類では以下の 8 クラスに分類する。

Trousers, Jackets, Sweaters&Knitwear, Shirts, Denim, T-Shirts&Vests, Shorts, Coats.

また、女性向けのカテゴリ分類では、以下の 10 クラスに分類する。

Tops, Trousers, Jackets, Dresses, Knitwear, Tote-Bags, Denim, Shorts, Skirts, Coats.

3.5 服飾品の色分類

色は、そのアイテムの主な色を表すものである。アイテムが複数の色を含む場合でも、主要な色を与えられている。 $K=10,000$ 枚以上のアイテムを含む以下の 9 クラスに分類する。

Black, Blue, Brown, Green, Grey, Neutrals, Pink, Red, White.

3.6 VGG16 の学習

3.2 節から 3.5 節に述べたそれぞれの服飾品分類のために、学習データを使って VGG16 を学習する。まず、事前学習済みの VGG16 ネットワークに対して、それぞれの分類するクラス数に合わせて最終層のノード数を変更する。その上で、学習データである画像を与えて学習する (fine tuning)。各クラス K 枚の画像データがあるが、その内 8 割を学習用、残りの 2 割を検証用とする。学習結果の識別率の評価には、検証用データを利用する。学習は 20 エポック行う。

3.7 コーディネート推薦による服飾品分類性能の評価

本研究では、VGG16 による服飾品分類結果がコーディネート推薦に有効であるかどうかによって、自動分類の性能を評価する。コーディネート推薦には、Selman らの手法を用いる。

アイテムの画像データに対して、学習したそれぞれの分

類器を適用する。次に、得られた分類結果を正解のメタデータ内のカテゴリ、色、ジェンダーの部分と置き換える。この際、分類対象以外のクラスはそのまま利用する。置き換え後のメタデータを用いて、Selman らの手法を適用しコーディネート推薦の正解率を求めて評価する。ただし、Selman らは、アイテムベースの協調フィルタリングで得られた候補の共起回数が 0 回または 1 回の時にアイテム属性による協調フィルタリングに切り替えていたが、本研究では、共起回数が 0 回の時のみアイテム属性ベースに切り替える手法も調査する。

eCom'22 のコンテストでは、incomplete outfit、candidates が与えられ、各推薦手法による結果が missing item と一致したかで評価している。しかし、コンテストの正解である missing item は公開されていない。そこで、300,000 件の正しいコーディネートを含む学習用データを 8:2 に分け、240,000 件を推薦時の学習用データとして用い、残りから新たに incomplete outfit (問題) と missing item (正解) を作成した。candidates (選択肢) も、missing item を含む 9-41 個のアイテムのリストをランダムに生成して作成した。これらを使って評価を行う。

4. 実験

4.1 各識別器の評価

3.2 節から 3.5 節のそれぞれについて、VGG16 を用いて学習し、分類器を作成した。学習済みの分類器の検証用データに対する識別率を表 1 に示す。

4.1.1 カテゴリ分類器の考察

表 1 の 1 行目の $K=10,000$ での識別率は 83.12% と、Fashion-MNIST での VGG16 の性能と比べて低かった。Sweaters&Knitwer は男性向けが 10,000 枚以上あるのに対し、女性向けが 1,000 枚以下しかなく、反対に Knitwear は男性向けが 1,000 枚以下で女性向けが 10,000 枚以上あるように、カテゴリのクラス分けの段階で大きく偏ってしまっていることが一因と思われる。

比較のために、 $K=1,000$ での識別率を表 1 の 2 行目に示した。この際、クラス数は 1,000 枚以上のデータを含む 32 カテゴリとした。識別率は、78.45% と更に低下した。クラス数が 32 クラスに増えたこと、画像枚数が $K=1,000$ と少ないことが影響していると考えられる。

4.1.2 ジェンダー分類器の考察

クラス数 2、 $K=10,000$ でのジェンダーの識別率は 92.90% と高かった。しかし、 $K=24,000$ に増やした場合では識別率は 84.64% に低下した。この原因はわかっておらず、追加の調査が必要である。また、 $K=1,000$ とした場合、クラス数 2 では識別率 83.25%、クラス数 3 では識別率 82.00% であった。クラス数 3 では、男性向け/女性向けに加えて男女兼用のクラスを加えている。画像枚数を減らしたことで、識別率は $K=10,000$ より低下した。また、 $K=1,000$ 同士で比較

すると、クラス数を 3 に増やしたことで、識別率がやや低下した。

4.1.3 色分類器の考察

クラス数 9、 $K=10,000$ での色の識別率は 84.28% と低かった。Neutrals というクラスが黒や白、グレーなどといった彩度のない無彩色の画像であるため、他のクラスの White や Black と混同したことが理由と考えられる。 $K=1,000$ でクラス数 12 とした場合の識別率は、81.13% とさらに低下した。学習における Neutrals の割合は $K=10,000$ の場合と比較して下がったが、クラス数が 9 から 12 へ増えたため識別率が低下したと思われる。

4.1.4 ジェンダー別カテゴリ分類器

男性向けのカテゴリ分類では、 $K=3,000$ 、 $K=6,000$ のいずれでも識別率が 90.83% と 93.29% となり高かった。クラス数 7、 $K=6,000$ の識別率はクラス数 8 $K=3,000$ の識別率に比べて少し高かった。クラス数が 1 つ減ったのと画像枚数が増えたためと考えられる。4.1.1 節で述べたジェンダー毎に分けないカテゴリ分類よりも 10 ポイント以上識別率が向上しており、ジェンダー別カテゴリ分類にした効果が見られる。

一方、女性向けカテゴリの学習は、 $K=3,000$ 、 $K=6,000$ のいずれでも識別率が 85.10% と 86.89% となり、4.1.1 節のカテゴリ分類器よりも上昇したが男性向けカテゴリ分類器より 5 から 6 ポイント程度低かった。これは、学習するカテゴリ数が女性向けの方が多かったからだと思われる。女性向けカテゴリの学習は $K=3,000$ から $K=6,000$ に増やした際、クラス数は同じであるが、少し識別率が良くなった。

表 1 分類器の識別率

分類対象	画像枚数 (K)	クラス数	識別率(%)
カテゴリ	10,000	13	83.12
カテゴリ	1,000	32	78.45
ジェンダー	24,000	2	84.64
ジェンダー	10,000	2	92.90
ジェンダー	1,000	3	82.00
ジェンダー	1,000	2	83.25
色	10,000	9	84.28
色	1,000	12	81.13
男向けカテゴリ	3,000	8	90.83
男向けカテゴリ	6,000	7	93.29
女向けカテゴリ	3,000	10	85.10
女向けカテゴリ	6,000	10	86.89

4.2 ジェンダー分類器とジェンダー別カテゴリ分類器の組み合わせによる評価

4.2.1 実験手順

4.1 節の実験結果より、13 クラスのカテゴリを直接学習

するより、ジェンダー別に分けてカテゴリ識別の方が識別率が向上することが分かった。しかし、この結果は対象とするアイテム（服飾品）が男性向けか女性向けか分かっている場合での識別率であった。実際にカテゴリ分類器を適用してアイテムのカテゴリを自動付加する場面では、ジェンダーは不明である。

そこで、本実験ではアイテムのジェンダーをジェンダー分類器で識別し、その結果に基づいて適用するジェンダー別カテゴリ分類器を選択し、カテゴリ識別を行う。具体的には、アイテムが与えられると、まず、ジェンダー分類器によりジェンダーを識別する。識別結果が男性向けだった場合、そのアイテムに男性向けカテゴリ分類器を適用し、カテゴリ識別結果を得る。この結果を「男-カテゴリ」と呼ぶ。同様に、ジェンダー識別結果が女性向けだった場合、女性向けカテゴリ分類器を適用する。この結果を「女-カテゴリ」と呼ぶ。ジェンダー分類器は、4.1.2 節の $K=10,000$ および $K=24,000$ のものを適用した。ジェンダー別カテゴリ分類器は、4.1.4 節の $K=3,000$ および $K=6,000$ のものを適用した。それらの組み合わせについて、カテゴリ識別の識別率を調査した。

4.2.2 実験結果・考察

結果を表 2 に示した。男-カテゴリの識別率は全体的に 90% 付近と高かった。特に 24,000 枚のジェンダー分類器を使った $K=6,000$ の識別率は、90% を超えていた。その他の組み合わせでも、ほぼ同じ程度の識別率が得られており、大きな違いは見られなかった。

女-カテゴリの識別率は全体的に 85% 付近だった。男-カテゴリと同様に 24,000 枚のジェンダー分類器を使った $K=6,000$ の識別率が最も高かったが、その他の組み合わせでも、ほぼ同じ程度の識別率が得られていた。男-カテゴリの識別率と比べ、女-カテゴリの識別率が 5 ポイント程低い傾向があった。これは、学習時のクラス数が男性は 7,8 クラスなのに対して、女性は 10 クラスと 2,3 クラス多かったためと考えられる。

表 1 の男向けカテゴリ、女向けカテゴリと比較しても、識別率は大きくは低下していない。また、表 1 のカテゴリ分類器よりもいずれも高い識別率を示している。この結果から、ジェンダー分類器の結果を使ってジェンダー別カテゴリ分類器を選択し、カテゴリ識別する手法は、識別率の観点から有効であると言える。

表 2 分類器の組み合わせによる識別率

分類対象	画像枚数 (K)	ジェンダー分類器の画像枚数	クラス数	識別率 (%)
男-カテゴリ	3,000	10,000	8	89.01
男-カテゴリ	3,000	24,000	8	88.60
男-カテゴリ	6,000	10,000	7	89.73
男-カテゴリ	6,000	24,000	7	90.13
女-カテゴリ	3,000	10,000	10	83.82
女-カテゴリ	3,000	24,000	10	85.07
女-カテゴリ	6,000	10,000	10	85.77
女-カテゴリ	6,000	24,000	10	86.40

4.3 コーディネート推薦による評価

4.1 節、4.2 節の結果より、カテゴリ、ジェンダー、色の分類は 80%~90% 程度の識別率で行えることが分かった。この識別率がコーディネート推薦において有効か評価するために、実験を行った。メタデータにおいて正解の分類を"手動"、4.1 節で作成した分類器で置き換えたものを"自動"、4.2 節の方法で置き換えたものを"ジェンダー別"と表記する。それぞれのメタデータを使った場合のコーディネート推薦の正解率を表 3 に示す。

4.3.1 実験結果・考察

表 3 1 行目のメタデータがすべて手動のものがすべて正解のときの正解率である。このとき、共起回数 0 は 0.6873、共起回数 0 と 1 は 0.6485 になっている。分類器を使用したときの正解率が、これらの値に下側から近いほど手動で分類するのと精度が変わらないと言える。

表 3 末尾のアイテムベースの協調フィルタリングのみのときの 0.6398 がアイテム属性を利用していないときの正解率である。分類器を使用したときのすべての正解率がこの 0.6398 を下回っていないので、少なくとも作成した分類器を使ったアイテム属性モデルは、コーディネート推薦結果を向上させられることがわかる。

メタデータを"自動"にした場合、1 行目のすべて"手動"と比べて正解率の差は 0.01 ポイント未満であり、いずれも有効であったと言える。中でも、カテゴリのみ自動 (3 行目) とカテゴリと色を自動 (8 行目) では、すべて"手動"よりもわずかに正解率が高くなっている。カテゴリ分類では、例えばデニム生地ズボンが"Denim"と"Trousers"に分かれて分類されている。色分類でも、白いアイテムが"White"と"Neutrals"に分かれて分類されている。正解の分類は、服飾品の製造者側が指定したもので、必ずしも服飾品の見た目を反映していない。識別結果としては誤りになっても、見た目に類似したものを同じクラスに分類できれば、コーディネート推薦結果では、正解率がむしろ向上する場合もあると考えられる。

共起回数0の場合と、共起回数0と1の場合を比べると、いずれも共起回数0の方が0.04ポイント程度高くなっている。Selmanらの論文では、共起回数0と1の場合にアイテム属性モデルに切り替えていたが、本実験では共起回数0の方が良い結果となった。これは、使用しているデータの違いが大きいと考えられる。

ジェンダー別でメタデータをすべて自動とした場合（9行目）と、2行目のすべて自動を比べると、ジェンダー別の方がわずかに正解率が高くなっているが、ほぼ同程度と言える。識別率ではジェンダー別の方が高かったが、コーディネート推薦での評価では、大きな違いはないと言える。

これらの実験結果を総合して、VGG16を用いたアイテム分類は、本研究で採用したコーディネート推薦において有効であると結論できる。

表3 コーディネート推薦の正解率

メタデータ			アイテム属性利用	
カテゴリ	ジェンダー	色	共起0	共起0と1
手動	手動	手動	0.6873	0.6485
自動	自動	自動	0.6803	0.6398
自動	手動	手動	0.6877	0.6492
手動	自動	手動	0.6853	0.6406
手動	手動	自動	0.6868	0.6479
自動	自動	手動	0.6805	0.6406
手動	自動	自動	0.6803	0.6402
自動	手動	自動	0.6875	0.6488
ジェンダー別	自動	自動	0.6825	0.6400
アイテムベースの協調フィルタリングのみ			0.6398	

5. おわりに

本研究では、深層学習による画像分類の一手法であるVGG16で識別した服飾品に関する種々の情報が、コーディネート推薦において有効であるか調査した。まず、VGG16により、服飾品のカテゴリ（種類）、ジェンダー（男性向け／女性向け）、色のクラス分けを学習し、その識別率を調べた。その結果、80%から90%の識別率が得られることが分かった。次に、ジェンダー別カテゴリ分類器を構築することで、カテゴリ識別の性能が向上するか調査した。実験の結果、実用場面に即して、誤りを含むジェンダー分類の結果に基づきジェンダー別カテゴリ分類器を選択し、カテゴリ識別を行っても、直接カテゴリ分類器を適用するより、高い識別率が得られることが分かった。更に、学習した分類器を使ってメタデータを自動付与した結果に基づくコーディネート推薦の性能を調査した。その結果、分類結果には誤りがあるにもかかわらず、正解である"手動"でのメタデータ付与と同程度の推薦の正解率が得られることが分かった。これらの結果から、VGG16による服飾品へのメタデ

ータ付与が有効であることが示唆された。

今後の課題として、データ数が少数のクラスへの対処が挙げられる。本研究では、VGG16を学習するために十分な数のデータが提供されているクラスのみを識別対象とした。例えば、カテゴリ分類では、データ中には129種類のカテゴリが含まれているが、識別対象としたクラス数は13であった。これら13クラスで全体の81%程度のアイテムを含んでいるが、残りのアイテムについては、メタデータの自動付与の対象とならない。学習済みのクラスの情報を利用して、未学習のクラスに対しても何らかの特性を付与してコーディネート推薦に繋げるなど、より適用範囲を広げることが今後の課題である。

参考文献

- [1] 長尾聡, 高橋伸, 田中二郎, "過去の行動から服のコーディネートを紹介する鏡状アプライアンス", ヒューマンインタフェースシンポジウム, pp. 973-976 (2007).
- [2] 森本泰貴, 藤本典幸, 萩原兼一, "ベイジアンネットワークモデルを用いた衣服コーディネート推薦システムの開発", 情報処理学会研究報告数理モデル化と問題解決 (MPS), pp.177-180 (2008).
- [3] 桂慶介, 加藤桃子, 島川博光, "TPOにふさわしくユーザが望む印象を与える服装の推薦", 情報科学技術フォーラム講演論文集 Vol.J14, No.4, pp. 445-446 (2015).
- [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 Vol.J9, No.3, pp. 179-180(2010).
- [5] Fashion-MNIST, <https://github.com/zaladoresearch/fashion-mnist/blob/master/README.ja.md>.
- [6] SIGIR eCom'22, <https://sigir-ecom.github.io/data-task.html>.
- [7] Delil, Selman, and Birol Kuyumcu. "Sefamerve Research at The SIGIR eCom'22: Outfit Recommendation Based on Collaborative Filtering", SIGIR eCom'22 (2022).