

VGG16 を用いたファッション雰囲気選別 Deep Learning モデル

登坂天成[†] 大槻明[‡]日本大学経済学部^{†‡}

1. はじめに

女性のファッションコーディネートに関する意識について、女性雑誌との関連性について分析した研究は存在するが[1]、例えば、女子学生がデートの際に実際にコーディネートの評価する相手は男子学生が多く、女性雑誌だけでは、男子学生の価値観を加味した正しい評価を行うことは難しい。ゆえに、本稿ではデート相手と同じ属性の価値観についてDeep Learningを用いて分析することにより、デート相手の価値観を考慮したコーディネートを紹介するモデルについて実装した。具体的には、女子学生が男子学生とデートに行くという状況を想定して、女子学生のコーディネートが男子学生にどのような印象を持たれるのかを推測（レコメンド）するコーディネート雰囲気選別Deep Learningモデルを実装した。

2. 雰囲気選別Deep Learningモデル

2.1 はじめに

Wear サイト¹から女性のファッションコーディネート画像を取得し、これらの画像を不特定多数の男子学生に提示して雰囲気をラベル付けしてもらうアンケートを実施した。次に、雰囲気選別を行うモデルを作成するために、学習前の VGG16[2]という画像処理モデルにラベル付けしたコーディネート画像を学習させる。そして、雰囲気を選別したい画像を新たに用意して、学習させたモデルで推測させることで、コーディネートの印象をレコメンドするモデルである。

2.2 本提案モデルにおけるロール

本研究の目的は、女子学生のコーディネートに対し、男子学生がどのような印象を持つのかを推測するファッション雰囲気選別Deep Learningモデルを実装することであるため、本モデルにおけるロールを表1に示す通り設

定した。なお、女性と男性の年齢に違いがある理由は、画像所得に使用したコーディネートサイトWEARでの条件設定が18~24歳しか設定ができなかったためである。

表1：本モデルにおけるロール

ロール	性別	年齢
印象を知りたい人	女性	20~24歳
印象を持つ人	男性	18~22歳

2.3 アンケートの実施

Wearサイトから、表1の条件（性別、年齢）に合致する女子学生のファッションコーディネート画像を約1,700枚所得した。そして、これらの画像に雰囲気ラベルをラベル付けするためにアンケートを行った。具体的には、2020年2~3月の間に、18~22歳の男子学生15名に1,700枚の画像を分担して雰囲気ラベルを付けてもらう形でアンケート回答を取得した。アンケートを取る際には「デート相手が画像の様なコーディネートをしてきた場合、あなたはどんな雰囲気を持ちますか?」という問いに、山本ら[3]の研究よりコーディネートの印象に類出する印象語を引用した5つの選択肢を与えた。この印象後5つを表2に示す。アンケートの結果、雰囲気選別が出来なかった画像を除き、収集した画像約1,700枚が表2に示す内訳（枚数）でそれぞれラベル付けされた。

表2：ラベル付けに使用した印象語

印象語 (カッコ内はChictopia datasetの名称)	雰囲気ラベルをラベル 付けた枚数
子供っぽい (Back-to-School)	209
個性的な (Bohemian)	335
カジュアルな (Casual)	800
品のある (Elegant)	87
男らしい (Menswear)	127

2.4 スクリプト全体像

2.3節の雰囲気ラベル付きコーディネート画像を学習させるために、図1に示すPythonスクリプトを実装した。図1の1_から始まるスクリプトは学習に必要なデータを読

Fashion Atmosphere Sorting Deep Learning Model Using VGG16

[†]Tensei Tosaka, [‡]Akira Otsuki^{†‡}Nihon University College of Economics¹ <https://wear.jp/>

み込むためのものである。2_~4_までは、いずれもコーディネート画像をモデルに学習させるためのスクリプトである。2_から始まるスクリプトは、VGG16モデルをゼロから新規に学習するモデルである。3_から始まるスクリプトは、学習済みVGG16を使って転移学習させるモデルである。転移学習とは最終層だけ変更・更新するモデルである。4_から始まるスクリプトは、学習済みVGG16をファインチューニングさせるモデルである。ファインチューニングとは学習済みモデルを学習し直すことであり、5_から始まるスクリプトは学習と検証で損失と正解率をプロットして過学習になっていないかなどの参考使用した。なお、2_、3_、4_はChictopia dataset¹のスキーマに合わせた学習モデルであり、それを前節で取得したコーディネート画像を配置したフォルダーで学習させるようにしたものが2_2、3_2、4_2である。2.6節の試行（予測）では、2_~4_のうち2_2_zero_learning.ipynbを学習部分に使用した。また、5_から始まるスクリプトは、2_2_で学習させたモデルを使って、新たなコーディネート画像の雰囲気を選別（レコメンド）するスクリプトである。



図1：スクリプト全体像

2.5 学習

2_2_zero_learning.ipynbでは、雰囲気ラベル付きコーディネート画像の学習を行う。なお、雰囲気ラベルはChictopia datasetの印象語ラベルを参考に表2の通り作成した。Chictopiaは米国を拠点とするオンラインファッションサイトで、多くの個人がコーディネートを投稿している。Chictopia datasetのラベル名を参考にした理由は、今後の研究で日本人以外も対象としたモデルを作成する際に、

Chictopia datasetのスキーマと連動して学習させることが出来るためである。Chictopia datasetは、外国人向けのコーディネート画像データセットであり日本人向けではないが、2_2_zero_learning.ipynbで学習させる雰囲気ラベル付きデータセットは、日本人向けの、さらに雰囲気ラベルが付けられたコーディネートデータセットであるため、このデータセットの構築自体にも意義があると考えられる。

また、2.4節でも述べたように、2_2_zero_learning.ipynbの学習部分にはVGG16を採用した。VGG16はILSVRCで2014年に2位になった一般物体認識モデルで、畳み込み13層と全結合3層、計16層のDeep Learningモデルである。2_2_zero_learning.ipynbでは、VGG16の学習前のモデルを用いており、学習条件は、学習データを計1258枚、検証データを300枚として、ミニバッチサイズは32、Epoch上限は50に設定した。

2.6 予測（レコメンド）

5_inference.ipynbは、図2に示すように前節で学習させたモデルを用いて新たに用意したコーディネート画像を分析することで予測（雰囲気の背別）を行う。つまり、印象語5つのうちの1つラベルを付けて出力する。

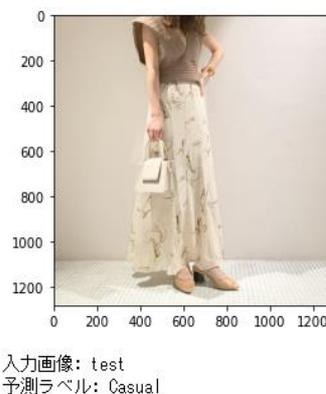


図2：予測（雰囲気の背別）のイメージ

3. 終わりに

本稿では、VGG16を用いたコーディネートの雰囲気選別Deep Learningを実装した。今後はコーディネート=アイテムセットとして考え、アイテム単体ごとに分析可能なスクリプトへと発展させていきたい。

謝辞

本研究はWearサイトのデータを利用することで行うことができました。データ利用を許可いただいた株式会社ZOZOTOWN様に御礼を申し上げます。

参考文献

- [1] 孫珠照, 小野幸一: 女子学生のファッション意識と女性雑誌との関連. ファッションビジネス学会論文誌, Vol.15, pp67-78, 2010.
- [2] Karen Simonyan, Andrew Zisserman: Very Deep Convolutional Networks for Large-Scale Image Recognition
- [3] 山本萌絵, 鬼沢武久: 衣服の印象を考慮した服飾コーディネートに関する研究

¹ <https://github.com/kyamagu/paperdoll/tree/master/data/chictopia>