# ファッションアイテムの分散表現に基づくコーディネート理解

## 1　はじめに

近年，機械学習の発展を背景として人工知能 (Artificial Intelligence： AI) が注目を浴びている．その中でも特に画像認識の分野は画像検出や顔検出など実社会における応用範囲が広く，AI の適用に関する研究が積極的になされている．AI は単純なパターン認識では人間の能力を凌駕する一方で，人間の感性に関する分野への AI の適用はいまだに難しく AI 研究の重要な課題とされている．以上の点を背景として，最近，様々な国際会議において，ファッション関連技術を扱うワークショップが催されるようになり，ファッションに対する認識技術への関心が高まりつつある．ファッション分野は人間の感性の多様性や主観を多く含む対象なので工学的には敬遠されてきたが，認識技術の高度化・データセットの充実・産業界の需要の増加などの要因で研究事例も増加傾向にある．衣服のトレンドは時系列解析，衣服のコーディネートは組合せ最適化問題というように，ファッションは工学的に広汎な研究テーマを内包している．本研究は，複数の衣服やアクセサリーの画像で構成されたコーディネートのデータを用いて，衣服の組合せに関する問題を解くことを目的としている．

## 2　Polyvore Dataset

今回用いるデータセット Polyvore Dataset [1] について述べる．Polyvore は EC サイトの服の画像を，ユーザがコラージュして投稿するサービスで，同サイトの投稿を収集したものが Polyvore Dataset である．

Polyvore Dataset ではアイテムがおおよそ種別の順番に並んでおり，トップス，ボトムス，靴，アクセサリの順になっている．トップスの中でも，シャツや T シャツはアウターの前に来るように配置されている．また，アクセサリーも概ねハンドバッグ，帽子，メガネ，時計，ネックレス，イヤリングなどの順に並んでおり，トップスから順に index が振られていく．このため index が大きくなるに従って，アイテム種別は大まかに定まるが，一意には対応していない．

このデータセットには 21,889 のコーディネートがある．1 つのコーディネートに含まれるアイテムの数は各コーディネートによって異なるため, アイテム数が 8 未満のデータは使用せず，アイテム数が 9 以上であるデータに関しては，index 9 以上のアイテムを削除して，index 1〜8 のアイテムのみ使用した．


図 1: ズボンが集結しているエリアの様子

## 3　卒業論文までの実験内容

### 3.1　提案手法

ファッションを理解するために，まずコーディネートの理解に焦点を当て，以下の手法を提案し実験をした．

#### 3.1.1　CAE に基づく分散表現化

提案手法では，まず Convolutional AutoEncoder（CAE）によりアイテム画像の分散表現化をした．図 1 に得られた分散表現を t-SNE により 3 次元空間にマッピングしたものを拡大して，ズボンの画像が集結しているエリアを示す．可視化することによっても，うまく分散表現を得られていることが確認できた．

既存の CNN モデルを直接用いる手法も考えられるが，ファッションアイテム固有の特徴に着目するためには CAE の方が適している場合が多い．また，CAE を用いる場合は，分散表現から画像を再現できるという利点もある．提案手法による分散表現は，服の色や形といった性質に基づいて作られた潜在空間中において，性質の類似性が高いアイテムの推薦に利用可能である．よって適切な分散表現の獲得と，得られた分散表現の効率的な利用が必要である．

#### 3.1.2 問題 Question 1 および 2

コーディネートは明確な答えのない問題であるため，提案手法の性能を評価するために，Fill-in-the-blank 問題に属する独自の問題を提案した．以下に問題の詳細を示す．

コーディネート理解の確認のために，予測候補と選択肢の分散表現におけるユークリッド距離に基づく 4 択問題を設定した．4 択問題は テストデータの他のコーディネートから候補を 3 つ選び，正解のアイテムと合わせて 4 択とした．この 4 択問題の正解率 （accuracy) を評価指標とした．

また，不正解となる選択肢を他のコーディネートの index 1 から集めた場合と index 3 から集めた場合の 2 通りの問題を作成し，Question 1, Question 2 とした．Question 1 ではトップスやワンピースなど類似した衣服が選択肢を構成するため，問題としての難易度は高くなる．一方で Question 2 は，選択肢が複数のカテゴリのアイテムからなるため，難易度の低い問題となる．なおこの問題は 7 種のコーディネートアイテムに最も適したコーディネートアイテムを選ぶことに相当する．

#### 3.1.3 深層学習に基づく学習器

与えられたコーディネートの分散表現を入力，適切なファッションアイテムの分散表現を出力として学習する手法を提案した．具体的な学習として，データセットのコーディネートを正例として，index 2〜8 の分散表現を学習器に入力し， index 1 の分散表現を出力させた．ここで今回は具体的な学習器として MLP と LSTM の 2 種類の深層学習モデルを使用した．MLP では複数の分散表現を連結して入力し，LSTM では index 順を 1 つの系列として解釈して学習するものとした．

### 3.2 実験

分散表現は 2048 次元のものを使用し，実験をした．MLP に基づく提案手法での正答率は Question 1 では 0.268, Question 2 では 0.528 であった．Question 1 においては, 4 択問題をランダムに解答した場合のベースライン = 0.25 をわずかに上回っているが，ほぼ同等といえる値であった．LSTM に基づく提案手法での正答率は Question 1 では 0.291, Question 2 では 0.532 であり，共に MLP でのテスト結果より高い精度が得られた．

Question 1 においてベースラインをわずかに超える結果しか得られなかったため，提案手法の有効性を確認するため，問題の難易度を下げた Question 3 を設定した．分散表現の距離に基づく学習に意味があるのであれば識別結果の accuracy は向上すると考え，Question 1 では 4 択候補となるアイテムをランダムに選出していたが，Question 3 では true のファッションアイテムの分散表現から比較的距離が遠い 3 種の index 1 アイテムで選択肢を構成した．Question 3 は識別結果がより良かった LSTM に基づく提案手法で解いた．正答率は 0.916 となり，精度が大きく上がっているため，分散表現に基づいたファッションアイテムの学習ができていることがわかった．

表 1: 距離指標の違いによる Question1, Question2 の正答率の比較

| 距離指標 | Question1 | Question 2 |
|---|---|---|
| ユークリッド距離 | 0.291 | 0.532 |
| コサイン類似度 | 0.304 | 0.532 |

## 4 距離指標の追加による精度の比較

4 択問題を解くにあたって，出力された分散表現と選択肢のアイテムの分散表現の距離をユークリッド距離によって比較し，回答を選択していた．今回はユークリッド距離に加えてコサイン類似度も新たに距離の指標として使用した．

表 1 にユークリッド距離とコサイン類似度に基づいて Question 1, Question 2 を解いた結果を示す．わずかではあるが，ユークリッド距離よりコサイン類似度を距離指標として用いた時の方が，Question 1 の正答率は上がり，3 割を超える精度が得られた．

## 5 人間による Question 1 の精度

学習器による Question 1 の正答率は約 3 割という結果である．そこで，この結果を評価するために Question 1 を人間が解いた場合の正答率を，アンケート調査に基づいて算出した．被験者を女性，男性各 5 名ずつ（20 代 9 名＋森 直樹教授）として Question 1 を 計 50 問解くことで，人間にとっての Question 1 の難易度を測る．

表 2 に人間と学習器両方の Question 1 回答結果を示す．Question 1 の人間の正答率はおおよそ 50% という結果になった．ちなみに，学習器が正解したデータに限定した時の男性と女性の正答率はそれぞれ男性 : 0.45, 女性 : 0.44 である．つまり，学習器が正解したデータの中には人間の正答率が高い問題も低い問題もおおよそ半々で含まれており，高い相関はないことがわかる．

表 2: 人間と学習器の Question1 の回答結果の比較

| 問題 | 学習器 | 男性 | 女性 | 問題 | 学習器 | 男性 | 女性 |
|---|---|---|---|---|---|---|---|
| 1 | ○ | 4/5 ○ | 1/5 ○ | 26 | × | 1/5 ○ | 4/5 ○ |
| 2 | ○ | 4/5 ○ | 4/5 ○ | 27 | × | 1/5 ○ | 1/5 ○ |
| 3 | ○ | ○ | 3/5 ○ | 28 | ○ | 3/5 ○ | 2/5 ○ |
| 4 | × | 3/5 ○ | ○ | 29 | ○ | × | 2/5 ○ |
| 5 | × | × | 3/5 ○ | 30 | × | 4/5 ○ | ○ |
| 6 | × | 3/5 ○ | 3/5 ○ | 31 | × | × | 4/5 ○ |
| 7 | × | 4/5 ○ | 2/5 ○ | 32 | × | × | 2/5 ○ |
| 8 | ○ | 1/5 ○ | 1/5 ○ | 33 | × | 1/5 ○ | 2/5 ○ |
| 9 | × | 1/5 ○ | 2/5 ○ | 34 | × | 2/5 ○ | 1/5 ○ |
| 10 | × | 3/5 ○ | 4/5 ○ | 35 | × | × | 3/5 ○ |
| 11 | × | 2/5 ○ | 2/5 ○ | 36 | × | × | 4/5 ○ |
| 12 | × | 3/5 ○ | 3/5 ○ | 37 | × | 2/5 ○ | 1/5 ○ |
| 13 | × | 3/5 ○ | 3/5 ○ | 38 | ○ | 3/5 ○ | 3/5 ○ |
| 14 | × | 3/5 ○ | 3/5 ○ | 39 | × | × | 1/5 ○ |
| 15 | × | 3/5 ○ | 2/5 ○ | 40 | × | 3/5 ○ | ○ |
| 16 | × | ○ | 4/5 ○ | 41 | × | 1/5 ○ | 1/5 ○ |
| 17 | × | ○ | 4/5 ○ | 42 | ○ | × | × |
| 18 | × | 2/5 ○ | 3/5 ○ | 43 | × | 1/5 ○ | 2/5 ○ |
| 19 | ○ | 3/5 ○ | 4/5 ○ | 44 | × | 1/5 ○ | 3/5 ○ |
| 20 | × | 3/5 ○ | 2/5 ○ | 45 | × | 3/5 ○ | 3/5 ○ |
| 21 | × | 3/5 ○ | 3/5 ○ | 46 | × | 2/5 ○ | 1/5 ○ |
| 22 | × | 1/5 ○ | 3/5 ○ | 47 | × | 2/5 ○ | 2/5 ○ |
| 23 | × | 4/5 ○ | 1/5 ○ | 48 | × | 3/5 ○ | 4/5 ○ |
| 24 | × | 3/5 ○ | 4/5 ○ | 49 | × | ○ | 3/5 ○ |
| 25 | ○ | 1/5 ○ | × | 50 | ○ | 1/5 ○ | 4/5 ○ |

正答率 学習器：0.22，男性：0.46，女性：0.53

また，1 問だけ学習器だけが正解していた問題があった．図 2 にその問題を示す．男性の回答では選択肢 3 が多く，女性の回答では選択肢 4 が多かったが，データセットとしては選択肢 1 が答えである．選択肢 2 を回答としている人もいたため，回答を一意に定めにくい問題であった．学習器はおそらく，入力のデータから赤色がテーマであることを読み取り，選択肢 1 を回答とし，正解したと考えられる．人間の目から見ても Question 1 は難易度の高い問題であることがわかる．

以上のことから，Question 1 は人間でも即座に正解がわかるような難易度の問題ではないと言える．

# 6 追加実験

## 6.1 実験説明

得られた各アイテムの分散表現を用いてコーディネートを組み替える実験を新たにした．今回，データ内に含まれる完全にテイストの異なる二つのコーディネートを用意した．図 3 にその 2 つのコーディネートを示す．A はメンズライクなコーディネートで，B はフェミニンなテイストのコーディネートである．

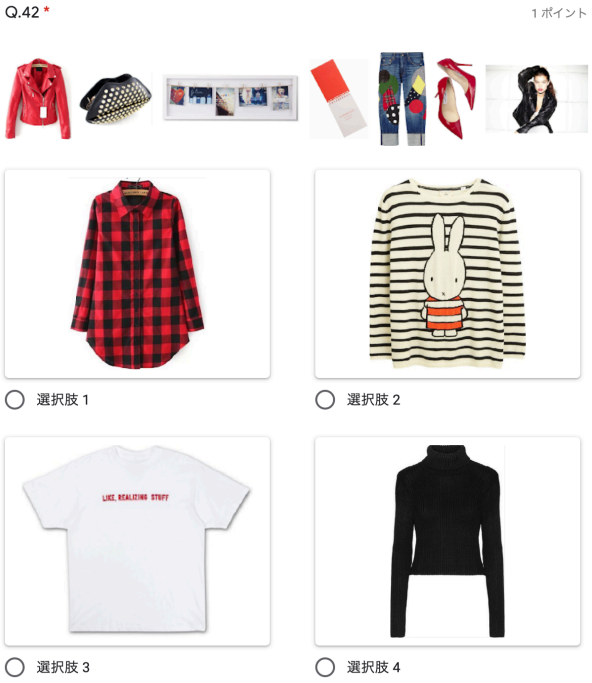


図 2: 学習器だけが正解した問題


図 3: 対照的な 2 つのコーディネート

以前までの実験と同様，トップスが欠落した状態で残りのアイテムを混ぜ合わせてコーディネートを組み直したものを LSTM に入力した場合，予測結果はどのように変化するのかを検証した．この実験によって，どのアイテムがコーディネートのアイデンティティとしての役割を担っていたのかということがわかる．

トップス以外のアイテムは A, B それぞれ アウター・ボトムス・シューズ・アクセサリ の 4 つである．よって，入力コーディネートは全て A のアイテム，全て B のアイテムの 2 通りを除いた $2^4 - 2 = 14$ 通りのコーディネートが存在する．また，回答の選び方は今までの実験と同様である．図 4 に例題を示す．選択肢の候補は A と B のトップスと，A と B のコーディネートとは関係のない他のコーディネートで使われていたトップスの 3 つ


図 4: 追加実験の例題

で構成した．

### 6.2 実験結果

今回もユークリッド距離とコサイン類似度を用いて回答を選んだ．だが，コサイン類似度を用いた場合の結果は，A のうち 1 つのアイテムを B のものに変えただけで B のトップスを回答としてしまうという B に大きくバイアスがかかった結果になった．よって，ユークリッド距離を用いた場合の結果を詳しく示す．

すべて A のアイテムを入力して得られた予測値と各選択肢のユークリッド距離を基準値とし，アイテムを混ぜ合わせたものを入力した時の予測値と各選択肢の距離は基準値との差で表す．

表 3 に結果を示す．A の中から 1 つだけアイテムを交換すると，回答はどれも A のトップスのままであった．B のアウターやスカートを交換した時，予測値と B のトップスの距離が縮まっていることから，この 2 つのアイテムが B の主要なアイテムであることがわかる．

次に，A の中から 2 つアイテムを交換すると，どのコーディネートの回答も依然として A のトップスのままであった．しかし，やはりアウターとスカートを交換した時，予測値と B のトップスの距離が最も縮まっていることがわかる．

最後に，A の中から 3 つアイテムを交換すると，3 つもアイテムが変わると流石に B のトップスを回答とするものが多い中，意外なことに B の主要アイテムが集まっている (B, B, B, A) のコーディネートの回答が A のトップスとなった．予測値と B のトップスの距離は今までで最も縮まっているが，予測値と A のトップスの距離も 0.4 とわずかながら縮まっている．このことから A のコーディネートの中でヘッドフォンとトップスが強く紐付いていることがわかる．

表 3: 追加実験の結果

基準値
(A tops，B tops，other tops) = (46.59, 49.95, 68.37)

| (outer, bottoms, shoes, accessories) | 回答 | 基準値との差 (A tops，B tops，other tops) |
|---|---|---|
| 1 つ交換 | | |
| (A, A, A, B) | A | (+2.1, +1.6, +2.6) |
| (A, A, B, A) | A | (+0.1, -0.6, +1.1) |
| (A, B, A, A) | A | (-0.5, -1.6, +1.9) |
| (B, A, A, A) | A | (-0.7, -2.7, +3.0) |
| 2 つ交換 | | |
| (A, A, B, B) | A | (+2.1, +0.4, +4.9) |
| (A, B, A, B) | A | (+1.5, -0.8, +5.0) |
| (A, B, B, A) | A | (-0.4, -2.5, +3.1) |
| (B, A, A, B) | A | (+1.8, -1.3, +6.6) |
| (B, A, B, A) | A | (-0.5, -3.0, +4.1) |
| (B, B, A, A) | A | (-0.6, -3.4, +4.5) |
| 3 つ交換 | | |
| (A, B, B, B) | B | (+2.3, -1.3, +8.0) |
| (B, A, B, B) | B | (+2.5, -1.1, +8.4) |
| (B, B, A, B) | B | (+1.8, -2.3, +8.0) |
| (B, B, B, A) | A | (-0.4, -3.7, +5.3) |

しかし，アイテムの影響力が感じられるような回答をしたコーディネートは先述した 1 つだけであったため，おおよそどのアイテムの分散表現も均等に考慮し，アイテムを予測していることがわかった．

## 7 まとめと今後の課題

今回，2 つの距離指標であるユークリッド距離とコサイン類似度による実験結果を比較をした．Question 1 においてはコサイン類似度を用いた場合の方が高い精度が得られた．しかし，追加実験では，ユークリッド距離を用いた場合の方が興味深い結果が得られたため，距離の測り方について今後検討していく必要がある．

また，Question 1 の難易度が高いことがアンケートにより判明したため，学習器の正答率が低いことも妥当であると考えられる．よって，正答の定義をよりコーディネートの冗長性に合わせたものに変えていく必要がある．データセットとしての正解以外にもコーディネートが成り立つアイテムは複数存在するため，正解とするアイテムの幅をもたせ，合わせると明らかにコーディネートが成立しないアイテムをはじくといったシステムの構築をすることが今後の課題となる．

## 参考文献

[1] Xintong Han, Zuxuan Wu, Yu-Gang Jiang, and Larry S Davis. Learning fashion compatibility with bidirectional lstms. In *ACM Multimedia*, 2017.