

オンラインファッションカタログを利用した 画像とテキストからの組み合わせ推薦

Recommendation of Coordinated Clothes
Based on Images and Texts of Online Fashion Catalog

堀 和紀^{*1} 岡田 将吾^{*1} 新田 克己^{*1}
Kazunori Hori Shogo Okada Katsumi Nitta

^{*1}東京工業大学総合理工学研究科知能システム科学専攻
Tokyo Institute of Technology

This research proposes a clothes recommendation system which uses three kinds of information such as image data, text data and coordination data extracted from the online catalog. Clothes data are gathered and clustered into several classes based on their genre and color features. When a user input an image of the clothes, our system selects several classes which matches up with it using these information, and shows images belonging to the classes to the user. Furthermore, by giving some texts which represent material or feeling of clothes, the user can find proper clothes.

1. 序論

近年ではネットショッピングを個人でも簡単に開設できるようになり、服の情報がインターネット上にはたくさん存在するようになった。アパレル EC サイトで有名な ZOZOTOWN[2] は、2011 年 3 月期の売上が 238.01 億円であったが、2013 年 3 月期の売上が 350.05 億円と、アパレル EC サイトはこれからも伸び続けていく市場であると思われる。

このようにネットショッピングは便利であるゆえに、市場規模がどんどん大きくなっていくことが予想される。一方で、ネットショッピングは便利ではあるが、市場が大きくなると情報が多くなり、自分のほしい服を探すのに時間がかかってしまうことがある。

この不便性を解消するために、推薦システムがいくつか作られてきた。最近では Relative Attribute Feedback[1] など、新しい推薦方法なども提案されている。Relative Attribute Feedback は、アイテムの属性に対してフィードバックをする方法である。Clothing Retrieval Based on Local Similarity with Multiple Images[3] は服の特徴の一部分を他の服にマージしたような服を推薦するシステムを開発した。しかし、Relative Attribute Feedback では推薦するためのデータ作成として、アノテーションに膨大な時間を有する。どちらのシステムも、洋服であれば、生地やブランド情報、重さなどは画像処理だけで判定できないこともある。また、システム内にある画像がシステムの入力となるため、自分の持っているものにあつものを推薦することはできない。

そこで、本研究ではインターネット上にある画像と文字情報を利用し、ユーザーの持っている服に似合うコーディネートを紹介するシステムを作ることを目的とする。インターネット上から服の画像と、服に対するコメントを取得し、推薦システムのデータベースを作成する。自分の持っている服の画像を推薦システムの入力として、インターネット上のコーディネート情報を用いて、コーディネートを紹介出来るようにする。推薦されたデータから、対話的に目的の服を取得できるようにする。画像に対するアノテーションは、コメントから自動的にタグ情

報を抽出し、人為的にアノテーションをせず、インターネット上の画像とテキストのコンテンツベースで推薦を行う。

2. 関連研究

Relative Attribute Feedback[1] はユーザーが属性のフィードバックを用いて繰り返し検索することができる。ユーザはまずいくつかのキーワードで検索する。例えば「黒くて高いヒールの靴」で検索すると、黒くて高いヒールの靴を画像で出力する。ユーザーはこれを見て、出力された画像それぞれの属性値に対してフィードバックを行う。データベース中の N 個の画像の集合 $P = \{I_1, \dots, I_N\}$ に対して、例えば I_1 に対して「より明るく」、 I_3 に対しては「よりフォーマルに」などとフィードバックする。すると、 P は、システムの現在の順位関数 $S_t : I \rightarrow \mathbb{R}$ で順位付けられる。 t は検索回数である。順位関数は過去の $(1, \dots, t-1)$ フィードバックすべてを学習し、 P を順位づけたものを返す。トップランクの $K < N$ 個の画像 $T_t = \{I_{t1}, \dots, I_{tK}\} \subseteq P$ をユーザーに見せて、フィードバックをもらう。これの作業を繰り返し、ユーザーに靴を推薦していく。

Clothing Retrieval Based on Local Similarity with Multiple Images は、全体の服の特徴から、他の服の一部分の特徴をマージした服を推薦するシステムを作成した。例えば、ファーがついていないダッフルコート、ファーが付いている服があるとする。ダッフルコートのデザインはそのまま、更にファーを付けた服を探したい場合、まずダッフルコートのファーを付けた部分を選択し、ファーが付いている服からファーの部分を選択する。すると、デザインはダッフルコートのままで、更にファーが付いた服が推薦されるようなシステムである。

3. システムの概要

この推薦システムを利用する対象者は、自分の持っている服に似合うコーディネートを探している人である。画像を用いてシステムが簡単に推薦し、テキストで更に細かい推薦が出来るようにした。

ユーザーが自分の服の写真を推薦システムへ入力すると、推薦システムがデータベースからその服にあったコーディネートいくつかを推薦する。推薦された服を見て、素材感などを更

連絡先: 連絡先: 堀和紀, 東京工業大学大学院総合理工学研究科知能システム科学専攻, 神奈川県横浜市緑区長津田町 4259 J2-53, TEL & FAX:0298-54-5204,

に細かく検索するために、テキストボックスが設けられており、ユーザーが「軽い」などのテキスト入力すると、推薦された服のなかで「軽い」素材の服が、ユーザーにいくつか推薦される。

推薦システムの機能は、入力された画像を opencv で画像処理し、特徴量をいくつか作成する。その特徴量が、推薦システムのデータベースの中の服の画像と特徴量が似ている服を取得する。データベースにあるコーディネート情報から、似ている服とコーディネートで使われている服をすべて取得する。取得した服の集合を C として、 C 中の服をいくつかランダムでユーザーに推薦する。ユーザーは推薦された服を見て、素材感などの細かい情報をテキストで推薦システムに入力する。すると推薦システムは、 C 中の服で情報が一致する服を取得し、いくつかランダムでユーザーに推薦する。

ディネートに使われている服が分類されたラベル同士をデータベース上で紐付けていく。

推薦モジュールは、ユーザーが写真を撮った服から特徴量を作成し、データベース上のどのラベルに属するかを判定し、そのラベルに紐付いているアイテムをランダムに推薦する。ユーザーは推薦された画像をみて、更にテキストをシステムに入力すると、細かい検索が出来るような仕組みになっている。

4. データの収集

ここでは、今回推薦システムに使用したデータについて説明する。使用したデータは、ZOZOTOWN(<http://zozo.jp/>) の服の画像と説明文 50,000 セットと、コーディネート情報 347 セットである。服のデータ、コーディネートのデータそれぞれについて説明する。

ZOZOTOWN の服はすべてカテゴリで分けられており、「トップス」「ジャケット」「パンツ」「スカート」などのカテゴリからさらに「トップス」であればさらに「T シャツ」「ブラウス」「ポロシャツ」などのサブカテゴリに分かれている。ZOZOTOWN の服のページには、服の画像とその服のブランド、金額、説明文などの情報がある。その中から、本研究では、画像と説明文を服の情報として用いる。画像のサイズはすべて $500 \times 600[px]$ であり、50,000 枚の画像のうち、顔が含まれる画像は 29,522 枚、顔が含まれない画像は 20,478 枚である。顔が含まれるか含まれないかの判断は第 5.1 節に記述するように、opencv の顔認識を用いて判別した。説明文には、服の情報として画像では取得できないような「着回し」、「肌触り」、「軽い」などの情報が含まれている。今回は服の画像、カテゴリ、説明文を収集して使用する。

ZOZOTOWN のコーディネートのページは、ZOZOTOWN に含まれる服で構成されていて、誰がコーディネートしたか、コーディネートのポイントの説明と、コーディネートに使った服の一覧が表示されている。本研究では、コーディネートページの服の一覧が抽出した服を 2 つ以上含む場合のみ、コーディネートデータとして収集し、使用する。

5. データベース作成手順

5.1 服の領域抽出

服の画像には、全身が写った画像、服のみが写った画像が混ざっている。それぞれに対して取得したい服の領域を抽出方法を説明する。まず、顔検出システムを用いて、顔の領域を抽出し、顔領域がある場合、服の画像は全身画像と判別し、顔領域がない場合、顔がない画像と判別する。なお、顔領域が複数ある場合、中心線に最も近い領域を顔領域とする。

全身画像の場合の服の領域は、上衣領域は顔領域の 2 倍の幅 2.5 倍の高さ、下衣領域は顔領域の 2 倍の幅、3.5 倍の高さとし、顔領域の下を上衣領域として、その下を下衣領域として抽出する。

顔がない画像の場合、画像のサイズ $500 \times 600[px]$ 中の $(140, 80)$ から $(350, 440)$ に大体の服の領域があることから、 $(140, 80)$ から $(350, 440)$ のサイズで抽出した。

5.2 画像から特徴量抽出

第 4 章で取得した服の情報を分類するために、服の画像をベクトル化する方法をそれぞれ次に示す。まず、画像の RGB 値を特徴量としてベクトル化する方法を記述する。第 5.1 節の方法で服の領域を抽出し、抽出後の画像に含まれる画素の RGB 値 $(0 \sim 255)$ のヒストグラムを作る。ヒストグラムのピンの数

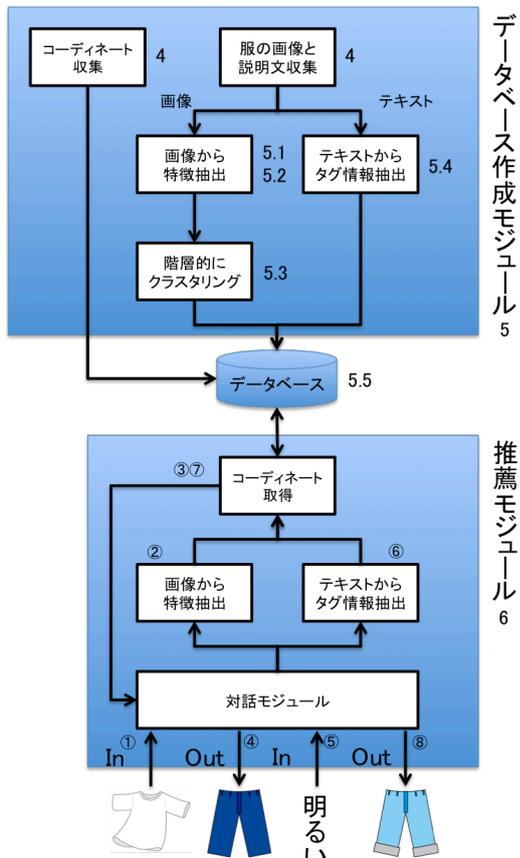


図 1: システム構成

推薦システムの構成を図 1 に示した。モジュールの横の番号は、本論文の章番号と対応している。

本論文の推薦システムは、あらかじめ作成したデータベースによってコーディネートを推薦する。まず、データベースの作成方法から説明し、次に作成したデータベースを用いた推薦方法を説明する。データベースを作成するために、アイテム画像とその説明文収集モジュールによって、ネットショップからアイテムの画像とその説明を取得する。取得したデータからいくつか特徴量を作成し、ベクトル化する。作成したベクトルによってクラスタリングし、階層的に分類する。コメントは形態素解析をし、word2vec[4] を用いて類似語を取得する。画像のクラスタリングで分類されたラベルと、テキストから取得したタグ情報をデータベースに保存する。最後に、コーディネート収集モジュールによってコーディネートデータを取得し、コー

は,RGB 値 0~255 を 5 つに分けて,RGB それぞれに対して 0~50, 51~101, 102~152, 153~203, 204~255 の 5 つのビン, 合計 15 個のビンを作成した. ヒストグラムのそれぞれの要素数を全体の画素数で割り, 正規化する. このようにして出来たヒストグラムを,RGB を連結してベクトル化する. 今回はヒストグラムを RGB の 3 つ作成し, 要素をそれぞれ 5 つ作ったので,15 の要素を持つベクトルを作成した.

次に, エッジを用いて特徴を抽出する. 第 5.1 節の方法で服の領域を抽出し, グレースケールに変換する. グレースケール画像から, ケニーのエッジ検出を用いてエッジを検出した. 特徴量は, ケニーのエッジ検出でエッジと判定された画素の座標 $[x, y]$ と隣の座標である $[x+1, y]$ がエッジと判定されるものをカウントする. カウント数を特徴量に使うが, 正規化するために, 抽出された画像の画素数でカウント数を割ったものを特徴量として用いる.

5.3 階層的にクラスタリング

クラスタリングの手法は, kMeans を用いた. クラスタリングは kMeans を用いるが, すべての画像を同時にクラスタリングして分類してもうまく分類できない. そこで, クラスタリングを階層的に行う. 階層的にクラスタリングした図を, 図 2 に示した. まず, ZOZOTOWN 上の服のデータはカテゴリに分けられていて, 例えば, tops, pants などのような大きなカテゴリでまず分けられ, カテゴリの中で更に細かくサブカテゴリに分けられている. tops のカテゴリでは, サブカテゴリは polo-shirt, knit-sweater などである. ここでは, tops, pants などのカテゴリを第 1 層, polo-shirt, knit-sweater などのサブカテゴリを第 2 層と呼ぶことにする. 階層的にクラスタリングするとは, 第 2 層のカテゴリを第 5.2 項で作成したうちの RGB の特徴量で分類し, 20 個のラベルをつける. この分類されたラベルを第 3 層とする. そして, 第 3 層のラベルごとに第 5.2 項で作成したエッジの特徴量で分類し, 3 つのラベルをつける. この分類されたラベルを第 4 層とする.

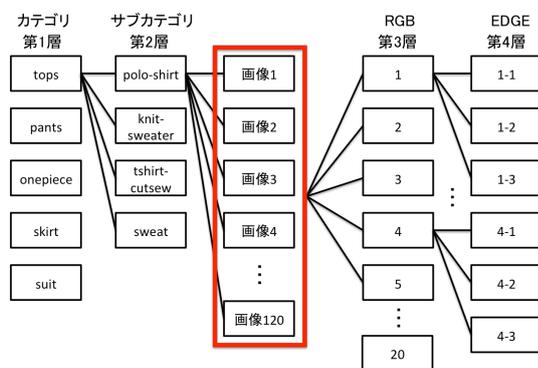


図 2: 階層的にクラスタリング

5.4 テキストからタグ抽出

ここでは, 第 4 章で取得したコメントから画像のタグを抽出する方法を説明する. まず, 第 4 章から取得したコメントを服ごとに形態素解析し, 出現した単語の集合を i をサブカテゴリ, j を服の番号として, W_{ij} とする. また, サブカテゴリ i 内の服の単語の集合 W_{ij} をすべて集めた集合を W_i とする. この時, 特徴語とする単語は名詞と形容詞のみの品詞だと仮定して, それ以外の品詞は無視する.

更にそこから word2vec で wikipedia[5] のデータを用いて単語間の類似語を取得する. これをすべてのサブカテゴリにつ

いて行う.

5.5 データベース

5.5.1 データベースの構成

ここではデータベースの構成について説明する.

第 5.3 節で階層的に分類した最深層のラベルと, 第 1 層, 第 2 層のカテゴリから, (第 1 層)_(第 2 層)_(最深層のラベル) でカテゴリごとの最深層が一意に特定できる. この特定したラベルを, 特定ラベルと呼ぶ. (特定ラベル例: tops_sweat_3-1, pants_denim-pants_2-4)

データベースは item, coordinate, center の 3 つのテーブルで構成されている.

item テーブルはそれぞれの服のデータが入っており, カラムは path, label, tag, sim で構成されている. path は服の画像の保存場所が入っており, 服を一意に識別できる. label は分類した時の, 特定ラベル, tag は第 5.4 節で取得したタグが List 形式で入っており, sim は tag の類似関係を示している.

coordinate テーブルはコーディネートデータが入っており, カラムは from, to で構成されている. from と to にはどちらも特定ラベルが入っている.

center テーブルは第 5.3 節で kMeans で分類するとき分類の中心となるベクトルを保存したものであり, カラムは label, center で構成されている. label クラスタリングの際に振り分けられたラベルの番号である. center はクラスタリングのラベルの中心座標が入っている.

5.5.2 データベース作成方法

第 4 章のデータを用いて, データベースを作成する方法を説明する. カラムの表現を簡単にするために, テーブル名. カラム という形式で記述する. 例えば, item テーブルの path カラムを参照する場合, item.path という表記をする.

まず, item テーブルの作成方法を説明する. item.path に服の画像を保存している path を入れる. 第 5.3 節によって階層的にクラスタリングして作成した特定ラベルをそれぞれ item.label に保存する. そして, 第 5.4 節によって作成したタグを, item.tag にそれぞれ保存し, 類似関係を item.sim に保存する.

次に, center テーブルを作成する方法を説明する. 第 5.3 節によって階層的にクラスタリングしたときに作成したラベルを center.label, その中心点を List として center.center に保存する.

最後に, 第 4 章のデータを用いて, coordinate テーブルを作成する. コーディネートデータに含まれる服の全体集合を I とする. コーディネートの番号 i を用いて I_i をコーディネート i が含む服の集合とし, I_i に含まれる服の番号 j を用いて, I_{ij} はコーディネート i に含まれる j 番目の服とする. 第 5.3 節で分類した時に, 服 I_{ij} が含まれている特定ラベルを取得する. 特定ラベルを L_{ij} とする. すべての i, j に対して, 特定ラベル L_{ij} を coordinate.from として, L_{ij} 以外の L_i に含まれる特定ラベルを coordinate.to として保存する.

6. 推薦手順

システム側の流れを図 1 を用いて説明する.

ユーザーが服を推薦システムに入力する.

画像から特徴を抽出する.

入力画像から特徴量抽出する場合, 第 5.2 節と同じ特徴量を抽出する. 入力画像の特徴量と, データベース内の center.center のベクトルとユークリッド距離が最小とな

るアイテムの center.label を取得し、特定ラベルを作成してゆく。

服の特徴から、コーディネートを取得し対話モジュールへ送る。

入力画像の特定ラベルを用いて、coordinate テーブルから入力画像の特定ラベルと紐付く特定ラベルを取得する。特定ラベルに属すデータベース上の画像をすべて対話モジュールに送信する。

対話モジュールがユーザーにコーディネートを紹介する。

送られてきた画像を特定ラベルごとにランダムで5つずつ推薦する。

ユーザーはコーディネートを見て、さらにテキスト形式で検索する。

図3の画像の下にテキストボックスが設けられており、そこに推薦されたコーディネートの中から絞り込むようなテキストを入力する。

送られたテキストからタグ情報を抽出する。

テキストを形態素解析し、単語をタグの集合として抽出する。このとき、取得するタグは名詞と形容詞のみを取り出し、その他は無視する。抽出したタグとそのタグと類似関係にあるタグを全て取得する。

タグ情報から、で取得したデータの中でタグ情報があるものを取得して対話モジュールへ送る。

で取得した服の中からで取得したタグを含む服を絞り込み、対話モジュールに推薦する。

対話モジュールがユーザーにコーディネートを紹介する。

図3に、推薦システムの推薦結果例を示した。

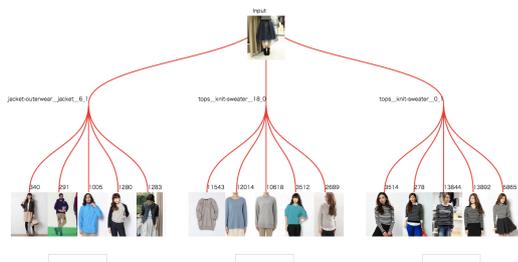


図3: 推薦システム-画像から推薦

上に入力画像があり、入力画像から特定ラベルごとに枝分かれしてコーディネートが推薦されている。特定ラベルごとに分かれた画像の先に、テキスト入力欄があり、ユーザーはそこにテキストを入力できる。

7. 評価

本研究で作成した推薦システムを18歳から25歳の女性10人に使ってもらった。

第6章の流れを10人のユーザーに体験してもらい、「ZOZOTOWNとの比較」、「改善方法」、「コーディネート情報を使った効果はあったか」、「感想」をアンケートした。「ZOZOTOWNとの比較」に関しては、ZOZOTOWNはファッションに詳しい

人でさえ検索に手間が掛かるが、本研究の推薦システムはだれでも簡単に直感的にシステムが使えるのが良いという評価を得た。「改善方法」を聞いたところ、「もっと服の分類を増やしてほしい」、「服は個人の好みがある。」という意見をいただいた。「コーディネート情報を使った効果はあったか」に対して、「コーディネートが推薦されることは新しい」、「ある程度服を推薦されると服選びが絞りやすくなる。」「感想」という問いに対してまとめてみると、「システムが簡単で使いやすい。」「自分の持っている服から検索できるのは新しい。」「素材感の情報まで検索できるのは面白い。」などの意見をいただいた。

8. 結論

本研究では、インターネット上の画像とテキストを用いてユーザーにコーディネートを紹介するシステムを作成した。推薦システムを作成するにあたり、推薦システムのデータベースと、推薦モジュールを作成した。そして、作ったシステムを実際に使ってもらい評価をしてもらった。推薦システムを作成したことによって、自分の持っている服の画像を推薦システムの入力として、インターネット上のコーディネート情報を用いて、コーディネートをコンテンツベースで推薦出来るようになった。この時、画像のタグなどは自動でインターネットから抽出するため、人為的なアノテーションを必要としない。推薦システムを作成してわかったことは、システムの入力を自分の画像とすることによって、検索における負荷を軽減することができ、ファッションの検索が簡単で使いやすいものになった。テキストで検索するときに、素材感まで検索できる推薦システムは面白いと評価をもらうことができた。

今後の課題として、データ数が足りないことと、コンテンツベースでやるためには画像からの特徴量を増やし、分類をもっと工夫する必要がある。推薦システムとして人の嗜好に合わせるためには、テキストのフィードバックだけでなく、Binary Relevance Feedback などを用いるのも良いと思う。服の推薦の仕方を、実際にマネキンが着ているような推薦ができればより良いと思う。

参考文献

- [1] A.Kovashka, D.Parikh, K.Grauman, University of Texas at Austin, and Toyota Technological Institute Chicago. WhittleSearch: Image Search with Relative Attribute Feedback. CVPR, 2012.
- [2] ZOZOTOWN, <http://zozo.jp/>
- [3] Masaru Mizuochi, Asako Kanazaki, Tatsuya Harada. Clothing Retrieval Based on Local Similarity with Multiple Images. ACM MM, 2014.
- [4] word2vec, <https://code.google.com/p/word2vec/>
- [5] wikimedia, <http://dumps.wikimedia.org/>