

類似画像検索システム『SUDACHI』を用いた商品レコメンデーション

"SUDACHI":The similar image search engine as a recommendation system.

(キーワード：レコメンデーション,類似性,特徴量)

(KEYWORDS: recommendation, similarity, feature)

○山川義介 (株式会社 ALBERT), 北研二 (徳島大学高度情報化基盤センター)

1. はじめに

次世代インターネットのコアになる技術の一つがレコメンデーション(推薦)だと言われ、インターネット通販(EC)において、なくてはならない機能となっている。商品レコメンデーションは色々なシーンで行なわれる。自分がどれを選べばよいか分からない時に推薦される場合もあるし、何かを購入しようとした時にその商品とセットで買うとよいものを推薦される場合もある。しかし、店頭での推薦を考えると、手に取った商品の類似商品をお薦めされることが多い。家電であればメーカー違いであったり、アパレルであれば若干形や模様の違う物だったりする。実際に筆者の調査結果[1]においてもランキングの次に魅力的な推薦は、「類似商品」であるという結果が出ている。イメージで選ばれることが多いアパレル商品の類似性を測定する方法として、画像の色、模様、形の特徴量とカテゴリの特徴量を用いて、効率的に商品をお薦めするシステムを開発したので報告する。

2. 目的

Amazon が採用している商品レコメンデーションの手法として有名な協調フィルタリング[2]は、商品の属性データを使わない人ベースのアルゴリズムで、「購買履歴の似ている人同士は似ているので、似ている人々が購買したものは、その人も欲しいだろう」という仮説に基づいている。しかし、購買履歴が少ない場合は分析データ量が少なく、初めて訪れた人や新商品には対応できないなどの欠点がある。

一方、コンテンツベースフィルタリングは、商品の属性の類似性に着目するモノベースなので、購買履歴は不要であり、初回訪問者や新商品に対応することも可能だが、嗜好性の高い商品などはその類似性を定義するのが難しく、また商品点数が多かつ書籍や衣料などスペックで表しにくい商材に適用することは難しい。

アパレルなどのデザインで選ぶことが多い商品に関しては、その色や模様などの詳細属性データはメーカーやECサイトでも保有しているケースは少なく、人が目で見て手作業で登録しない限り、精緻な検索には耐えられないというのが現状である。

一方、デジタルカメラの顔認識等において、画像認識や画像の類似性測定などの技術はめざましい進歩が遂げられている。しかし、ECサイトにおける商品画像の形態にはさまざまなものが存在し、図1「P」の様に単に商品のみが写っているものもあるが、「Q」の様に背景部分に景色が写っていたりモデルが着用

していたりする画像もあり、従来の画像認識技術ではどこが商品でどこが背景であるかの分離が難しく、商品のみを解析することができなかった。

以上の状況を鑑み、筆者らは背景部分を除去し、対象となる商品の部分のみから特徴量抽出を行ない、多数の商品の中から、色、テクスチャ(模様)、形状、カテゴリの重み付けをした上で類似画像を検索することができるシステム『SUDACHI』を開発することにした。



P



Q

図1. 様々な商品画像

3. 『SUDACHI』の概要

『SUDACHI』は、多数の商品画像の中から、背景を除去した上でユーザーの指定した商品と類似した商品を抽出して提示するシステムである。『SUDACHI』では以下の4種類の類似性に基づき、商品全体の類似性を決定することにした。

- ①商品の色に基づく類似性
- ②商品の模様に基づく類似性
- ③商品の形に基づく類似性
- ④商品の属性に基づく類似性

それぞれの類似性を定量的に扱うために、各商品に対する画像や商品属性から数値的な特徴量を抽出し、抽出された特徴量間の距離に基づき、画像と画像との間の類似性を求めている。上記の類似性①～③を決定するために用いる特徴量は、画像処理により抽出されるものであり、順に色特徴量、テクスチャ特徴量、形状特徴量と呼ばれるものを用いている。類似性④は、他の類似性とは異なり、画像そのものではなく、画像に付随するメタ的な情報から決定される。

4. 背景除去

『SUDACHI』は、背景除去のステップが最も特徴的である。商品画像から色や模様、あるいは形状等の特徴量を抽出する際に

は、背景部分を除去し対象となる商品の部分のみから特徴量抽出を行なうことが望ましい。背景部分も含め、画像全体から特徴量抽出を行なったのでは、高精度な類似商品推薦システムを構築することは困難である。

『SUDACHI』では、画像からの特徴量抽出に先立ち、背景部分の除去を行ない、対象となる商品部分のみの部分画像を抽出する処理を行なっている。画像のほぼ中心に商品が写っている場合は、自動的な背景除去処理が可能である。しかし、どの部分が対象となる商品かプログラムによって自動的に判別するのが困難な場合には、背景除去は半自動的に処理できるようになっている。



図2. 背景除去の例

5. 特徴量の抽出

色や模様（テクスチャ）的な情報を表す特徴量として、『SUDACHI』では、セグメント特徴量と呼ぶものを用いている。セグメント特徴量は、画像の領域分割(segmentation)によって得られた領域ごとに抽出される特徴量である。領域分割では、色やテクスチャといった特徴の様な連結画素をグループ化することにより、元画像を類似した色やテクスチャを持った複数の領域に分割する処理を行なう。さまざまな領域分割の手法が提案されているが[3]、『SUDACHI』ではグラフに基づく領域分割手法[4]を改良することによりきわめて高速な領域分割処理を実現している。また、色に対するセグメント特徴量は領域の平均色を、テクスチャ特徴量は Wavelet に基づく特徴量を採用している。セグメント特徴量に基づく商品画像間の類似度計算では、Earth Mover's Distance (EMD) と呼ばれる距離尺度を用いている。

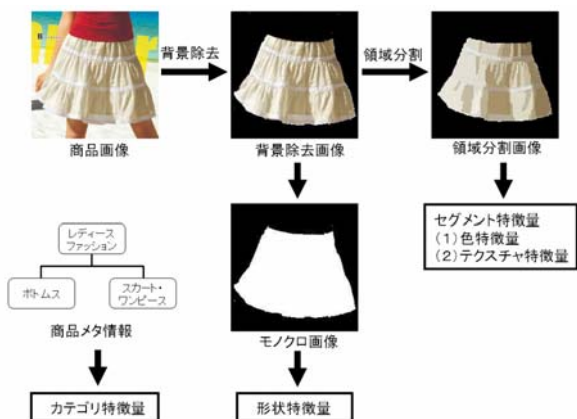


図3. 『SUDACHI』の特徴量抽出処理

6. 商品レコメンドーションへの応用

『SUDACHI』を用いて水着の類似商品を推薦するシステムを作成した。右上の象限には類似タイプの商品が、色の類似性の高い順に並んでおり、左下の象限には別タイプで色の類似性が高い商品が順に並んでいる。また、右下の象限には色違いの同じタイプが提示されている。



図4. 『SUDACHI』を用いた推薦エンジンの例 (見つける.jp)

7. まとめ

EC における商品の選択行動は、カテゴリの絞り込みかテキストによる検索がメインであったが、サイト内の回遊性を高める意味では、横断検索が有効だと考えられる。『SUDACHI』を用いることで、色や模様などのイメージで選択でき、かつサイト横断的な検索が可能となるので、消費行動のパラダイムシフトが起きることが考えられる。

参考文献

- 1) 山川：顧客満足度を上げる商品レコメンドーション；日本行動計量学会第36会大会発表論文抄録集, 2008,
- 2) 神鷹 敏弘；なんとなく協調フィルタリング-複数の順序応答に基づく推薦, 人工知能学会全国大会論文集, Vol. 64, pp. 219-226, 2004
- 3) J. Shi and J. Malik: "Normalized cuts and image segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.8, pp.888-905, 2000.
- 4) P. F. Felzenszwalb and D. P. Huttenlocher: "Efficient graph-based image segmentation", International Journal of Computer Vision, Vol.59, No.2, pp.167-181, 2004.