

ECサイトにおける画像の類似性測定 ～ 『SUDACHI』の開発と商品推薦エンジンへの応用 ～

○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

商品推薦はいくつかのシーンで行なわれる。自分が何を探しているか分からない時、どれを選べばよいか分からない時に推薦する場合もあるし、何かを購入した時に、その商品とセットで買うとよいものを推薦する場合もある。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事が多い。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。特にイメージで選ばれるアパレル商品の類似性を測定する方法として、画像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

商品レコメンデーションの方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事は難しい。

アパレルなどのデザインで選ぶ商品に関しては、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、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人物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十分な解析ができなかった。

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事ができるシステム『SUDACHI』を開発することにした。



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。

3. 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だち、背景部分の除去を行い、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

4. 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一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。

5. 推薦エンジンへの応用

『SUDACHI』を用いてファッション商品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カテゴリ商品が色の類似性の高い順に並んでおり、左下の象限には別カテゴリで色の類似性が高い商品が順に並んでいる。



図3. 『SUDACHI』を用いた推薦エンジンの例 (見つかる.jp)

6. 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。『SUDACHI』を用いる事で、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事が考えられる。

参考文献

- 1) 山川：顧客満足度を上げる商品レコメンデーション；日本行動計量学会第36会大会発表論文抄録集,2008,
- 2) 神寫 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.