

EC サイトの Apparel 商品に対する 購買意欲の度合いの推定に関する検討

宮内 拓実* ・ 齊藤 直輝** ・ 浅水 仁***

A Note on Purchase Intent Level Estimation of Apparel Product on EC-site

Hiromi MIYAUCHI* ・ Naoki SAITO** ・ Satoshi ASAMIZU***

Abstract — A purchase intent-level estimation approach of apparel products is presented in this paper. Our method derives visual features from the apparel product images on EC-site and estimates the purchase intent level using these visual features via conventional machine learning approaches. The experimentation in this paper is evaluated the effectiveness of our method.

Key words: Purchase intent-level, apparel products, visual feature extraction.

1. はじめに

情報通信技術の発達に伴い Electronic Commerce (EC) 市場の規模が拡大している。特に Apparel 業界では、図1に示す通り、衣服カテゴリの EC 市場規模が増加している[1]。これに伴い、Apparel 商品を取り扱う EC サイト上には大量かつ多様な商品が存在するため、ユーザが所望する商品を探し出すためには多くの時間と労力が必要となる。

この問題に対して、情報推薦技術[2]の活用が有効であると考えられる。具体的には、EC サイト上の商品の中からユーザが望むものを自動で推薦する技術が実現することで、購入を望む商品を探し出す時間と労力が少なくなることが期待できる。さらに、EC サイトの運営側にとっても、購買意欲が高い商品を効率的にユーザへ提示できるため、売上の向上が期待できる。ここで、上記の推薦技術を実現するためには、まず EC サイトに掲載されている商品に対するユーザの購買意欲の度合いを推定する手法を実現する必要がある。

そこで本稿では、EC サイト上に存在する Apparel 商



図1. 衣服カテゴリにおける EC 市場規模の推移

品の画像を用いたユーザの購買意欲に関する度合いの推定を実現するため、推定に利用する画像特徴量と推定器について検討を行う。具体的には、Apparel 商品画像から算出される色や形状、意味内容を表現可能な3種類の特徴量をそれぞれ利用して購買意欲の度合いを推定し、その性能を比較する。さらに、推定器として、機械学習に基づく2種類の推定器を採用し、各々の推定性能を比較する。以上の検証により、EC サイト上の Apparel 商品画像に対する購買意欲の度合いの推定における今後の検討課題を明らかにする。

* 釧路工業高等専門学校 専攻科

** 北海道大学 総合 IR 室

*** 釧路工業高等専門学校 創造工学科

以降、2章ではアパレル商品に対する購買意欲の度合いを推定する手法について説明し、3章で画像特徴量と推定器による推定性能を検証するために行った実験結果を示す。最後に4章でまとめとする。

2. 購買意欲の度合いの推定手法

本章では、アパレル商品画像を用いてユーザの購買意欲の度合いを推定する手法について説明する。本手法では、推定対象となるアパレル商品画像から色、形状、および意味内容を表現可能な特徴量をそれぞれ算出し、これを推定器へ入力することで、商品に対する購買意欲の度合いを推定する。以降では、2.1節で画像特徴量の算出方法について説明し、2.2節で推定器として用いる機械学習に基づく手法について説明する。

2.1 画像特徴量

本手法では、色、形状、および画像の意味内容に関する3種類の画像特徴量をアパレル商品画像から算出する。以降では、各特徴量の算出方法について説明する。

1. 色に関する特徴量 (768次元)

本手法では、カラーヒストグラムから得られる数値情報を色に関する特徴量として採用する。具体的には、RGBのチャンネルごとにビン数が256のヒストグラムを作成し、これらを並べたベクトル ($256 \times 3 = 768$ 次元) を「色に関する特徴量」とする。

2. 形状に関する特徴量 (8,100次元)

画像スケールの変化や照明変化に頑健であり、人物および物体の検出に広く用いられる画像特徴量の一つであるHistogram of Oriented Gradients (HOG) 特徴量[3]を「形状に関する特徴量」として用いる。HOG特徴量は、グレースケール画像における画素値の勾配を角度毎にヒストグラム化することで求められる。具体的には、画像において 32×32 画素のブロック毎に勾配方向を0から $180[\text{deg}]$ まで $20[\text{deg}]$ 毎にヒストグラム化することで、8,100次元の特徴量を取得する。

3. 画像の意味内容に関する特徴量 (2,048次元)

画像を直接入力可能な深層学習による分類器の一つである畳み込みニューラルネットワーク

(Convolutional Neural Network: CNN) [4]における中間層の出力値は、画像の意味内容を表す特徴量表現として、様々なタスクで有効であることが報告されている。そこで本手法では、一般画像分類の大規模なデータセットであるImageNet[5]で学習されたCNNモデルの一つであるInception-v3[6]の第3プーリング層から出力された2,048個の数値データを並べたベクトルを「画像の意味内容に関する特徴量」として算出する。

ここで、上記の3種類の画像特徴量は、アパレル商品画像が持つ情報を詳細に表現するため、高次元なベクトルとなる。一方で、大量のアパレル商品画像に対してユーザが購買意欲の度合いを付与することは困難であるため、推定器の学習データは非常に少量になると考えられる。一般的に、機械学習による推定器では、学習時にサンプル数よりも特徴量の次元数の方が大きい場合、「次元の呪い」により推定器が学習データに過剰に適合してしまう「過学習」が発生する可能性が高まる。したがって、算出された画像特徴量をそのまま利用して推定器を学習すると、過学習により汎化性が低下し推定精度が大幅に減少する。そこで本手法では、算出された画像特徴量に対して主成分分析 (Principal Component Analysis: PCA)[7]による次元圧縮を施すことで、推定器の学習およびテストで利用する特徴量の次元数を減少させ、過学習による推定性能低下の影響を抑制する。

2.2 機械学習に基づく推定器

本手法では、前節で算出した画像特徴量を用いて、Support Vector Machine (SVM) [8]および k 近傍法 (k -NN) [9]によりアパレル商品画像に対する購買意欲の度合いを推定する。SVMは、特徴空間上で2つのクラスを分離するための超平面をサポートベクターの-marginが最大となるように決定することで、推定を実現する方法である。また k -NNは、特徴空間上においてテストサンプルの近傍に存在する訓練サンプルが属するクラスに基づいて、推定結果を決定する手法である。これらは、代表的な機械学習による推論手法であるため、本手法において推定器として採用した。



図 2. 本実験で用いるアパレル商品画像の例

3. 実験

本章では、2章で説明した手法によるアパレル商品画像に対するユーザの購買意欲の度合いの推定性能を検証するために行った実験の結果を示す。本実験では、楽天グループ株式会社が情報学研究リポジトリ¹を通して公開している楽天市場のデータ²の中で、ジャンルが「メンズ T シャツ」の商品画像 180 枚を利用した。図 2 に本実験で用いるアパレル商品画像の一例を示す。ただし、各画像は衣服に関係しないテキストや背景を除去して利用した。実験では、これらの商品画像に対して、7 名の参加者（参加者 A~G）がそれぞれ付与した購買意欲の有無を推定した。各参加者が付与した購買意欲がある画像と無い画像の数を表 1 に示す。

本実験におけるアパレル商品画像に対する購買意欲の有無の推定では、色、形状、意味内容に関する画像特徴量をそれぞれ用いて、SVM および k -NN により推定する方法を採用した。したがって、本実験では 6 種類の方法で推定を行い、それらの性能を評価した。推定精度の評価は Leave One Out 法により行い、評価指標として次式に示す正解率を用いた。

$$\text{正解率} = \frac{\text{正しく推定された画像数}}{\text{推定対象の画像数}} \quad (1)$$

なお、SVM のカーネル関数は線形カーネルとした。また、 k -NN の近傍数は 1 から 17 の中から推定結果の正解率が最も高くなる値を採用した。

表 1. 各参加者により付与された購買意欲がある商品画像と無い画像の数

	購買意欲あり	購買意欲なし
参加者A	17	163
参加者B	18	162
参加者C	43	137
参加者D	61	119
参加者E	91	89
参加者F	44	136
参加者G	89	91

各参加者により付与された購買意欲の有無の推定結果を表 2 に示す。実験結果より、画像の意味内容に関する特徴量を用いた k -NN による推定が、他の特徴量および推定器と比較して最も高い正解率で推定可能であることが確認された。ユーザが購買意欲の有無を判断するタスクは、色や形状等の低次の情報のみではなく、意味内容等の高次の情報を考慮する必要がある複雑なものであると考えられる。したがって、より複雑な情報を表現可能な「画像の意味内容に関する特徴量」がアパレル商品画像の購買意欲の推定に有効であったと考えられる。ここで、本実験で用いた SVM は線形カーネルを採用しているため、特徴空間上での線形構造のみ考慮可能である。一方で、 k -NN は特徴量空間上の近傍の構造を考慮しているため、非線形構造を考慮した推定が可能である。購買意欲の有無の推定においては、算出された特徴量が非線形な構造を有していると考えられるため、これを考慮可能な k -NN による推定が有効であったと考えられる。

ここで、本実験で利用した手法では、過学習を抑制するために PCA による次元削減を施している。しかしながら PCA は教師なしの次元削減手法であるため、次元削減後の特徴量が、購買意欲の度合いの推定に有効であるとは限らない。したがって、推定精度向上のためには、画像特徴量および推定器に関する検討に加えて、教師あり次元削減手法の導入についても検討する必要があると考えられる。また、本稿の手法で採用している推定器は非常に古典的なものであり、近年提案

¹ <https://openforum.nii.ac.jp/dsc/idr/index.html>

² <https://doi.org/10.32130/idr.2.1>

表 2. アパレル商品画像に対する購買意欲の有無の推定結果

画像特徴量	推定器	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	参加者 G	平均
色	SVM	0.78	0.81	0.64	0.65	0.48	0.65	0.49	0.64
形状	SVM	0.83	0.81	0.62	0.63	0.62	0.64	0.58	0.68
意味内容	SVM	0.83	0.80	0.63	0.61	0.64	0.66	0.56	0.68
色	k -NN	0.76	0.74	0.77	0.67	0.59	0.76	0.54	0.69
形状	k -NN	0.85	0.89	0.75	0.65	0.64	0.76	0.56	0.73
意味内容	k -NN	0.91	0.90	0.76	0.66	0.68	0.76	0.61	0.75

されている深層学習に基づく推定器など、複雑なタスクにおいても高精度な推定を可能とする手法が種々提案されている。これらの手法の採用についても検討を実施する必要があると考えられる。

4. まとめ

本稿では、ECサイト上に存在するアパレル商品画像に対する購買意欲を推定する手法の実現に向けて、画像特徴量および推定器に関する検討を行った。実験結果より、購買意欲の推定は非常に複雑なタスクであり、より高次の情報を表現可能な特徴量および高度な推定器の導入の必要性が示唆された。また、過学習抑制のために導入した次元削減手法についても、教師ありの手法に拡張された手法を利用する必要があると考えられる。

謝辞

本研究では、国立情報学研究所のIDRデータセット提供サービスにより楽天グループ株式会社から提供を受けた楽天データセット³を利用した。

参考文献

- [1] 経済産業省, “令和元年度 内外一体の経済成長戦略構築にかかる国際経済調査事業(電子商取引に関する市場調査) 一報告書一”
- [2] 土方嘉徳, “嗜好抽出と情報推薦技術,” 情報処理学会論文誌, vol. 47, no. 4, pp. 1-10, 2006.
- [3] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886-893.
- [4] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
- [6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, “Rethinking the Inception architecture for computer vision,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818-2826.
- [7] S. Wold, K. Esbensen and P. Geladi, “Principal component analysis,” *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37-52.
- [8] C. Cortes, V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [9] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

³ https://rit.rakuten.com/data_release/