

EC サイトにおける商品の特徴ラベル分類に関する精度と付与率の向上に向けた取組み

酒井 敏彦[†]

Toshihiko Sakai

三宅 悠介[†]

Yusuke Miyake

1. はじめに

電子商取引 (Electronic Commerce: EC) サイトの運営では、ユーザが求める商品に辿りつきやすくすることが求められる。そこで、商品に対して、該当商品の特性を表す適切なラベルを付与しておき、ユーザが検索時に求める特性を指定する、絞り込み検索の機能が EC サイトで提供されている。商品にラベルを付与する方法として、手動による方法と機械的に分類する方法がある。前者の方法を CtoC の EC サイトで実施する場合、商品の販売者が該当商品の特性を把握していると考え、販売者に付与の設定を依頼することができる。しかしながら、特性を表すラベルの乱立を避けるため、ラベルの種類を EC サイト運営側が指定する運用の場合、ラベルの追加削除の時期と付与の時期のずれといった、追従の課題が発生する。また、手動で付与することから、漏れや誤りも容易に発生しうる。そこで、商品のテキスト・画像等の情報を用いて、機械的に分類する方法が検討されている。

本研究では、手動による方法と比較し、機械的なラベル分類によって、ラベル付けの精度と付与率の向上を目指している。そこで、本報告では、機械的なラベル分類における課題のうち、CtoC の EC サイトに特徴的な以下の課題の解決を試みる。

課題 1: 商品のテキスト・画像単体では、ラベル进行分类するための情報が不足

課題 2: ラベルの信頼度及びラベル付与率が低い

課題 1 に対しては、実際にデータから課題 1 の傾向を確認した上で、テキスト・画像それぞれに対する分類器の分類結果の和集合を求めることで、二つの分類結果を統合するアプローチを検討する。課題 2 に対しては、テキストにおいて、ラベルの単語でのキーワードマッチを行うアプローチを検討する。課題 1 へのアプローチに対する評価では、テキストだけでは分類できない作品に対して、画像の分類結果との和集合を求めることで、分類の補完が可能となることを確認した。また、テキスト・画像単体よりも、Recall の値が向上することを確認した。課題 2 へのアプローチに対する評価では、ラベルの分類誤り及び付与漏れを防止できることを確認した。また、特定の商品カテゴリにおいて、F1-score が向上した。

なお、本報告では、ハンドメイド作品を扱う EC サイト

「minne」[4]における、プロダクション環境のテキスト・画像を用いた商品の特徴ラベル分類のデータに対して、提案する手法を適用し、その有効性について評価を行った。これ以降は、minne のサービスについて述べる場合は、商品を「作品」、販売者を「作家」と呼ぶ。

2. 関連研究

EC サイトでの商品分類において、テキストや画像を特徴量として用いる研究 [1], [7] が行われている。しかし、どのような特徴量が商品分類に効果的なのかは、扱う商品の種類や分類問題の設定、特徴量の元となる商品の属性情報の品質に依存してしまう。

CtoC の EC サイトでは、商品のタイトル・説明文や画像といった属性情報を販売者自身で設定する。よって、多様な属性の表現が存在することから、中には、精度良く分類を行うには、情報量が不足しているものも含まれる。また、同じ対象であっても、画像の数が多く品質も幅広いことから、精度良く分類するには、学習データとしての画像数が十分ではないことも多い。特に、本研究で扱うハンドメイド作品を対象とした CtoC の EC サイトでは、販売される作品自体も手作りの一点ものであることから、作品の種類増加に伴って、上述の課題が顕著化してしまう。前述の課題 1 に相当する、この課題から、各特徴量の少ない手がかりを統合し、分類の精度を向上させる方式が検討されている。

このような、テキストの特徴量と画像の特徴量を組み合わせる研究が行われている [2], [6], [10]。これらの研究では、商品のテキストや画像を各々特徴量へ変換し、早期に融合する手法を Feature-level fusion (Early fusion) と呼んでいる。一方、各々分類器を作成し、分類結果を基に融合する手法を Decision-level fusion (Late fusion) と呼ぶ。Decision-level fusion では、各々分類器を作成し、分類結果を出力した後に、多数決を取る手法 [3] や各々の分類器から出力される分類確率を多層マルチパーセプトロンに入力し、学習する手法 [8] がある。

しかし、上記の各融合手法においてもどのような手法が効果的なのかは、扱う特徴量や分類問題の設定に依存するため、特徴量の性質や問題設定を考慮し、適切な手法を選定する必要がある。加えて、機械的な商品分類の学習データとなるラベルが、販売者によって設定されたものを扱う場合、ラベル自体の信頼度及びラベル付与率が低いという、前述の課題 2 に相当する課題も考慮する必要がある。

[†] GMO ペパボ株式会社 ペパボ研究所
Pepabo R&D Institute, GMO Pepabo, Inc., Tenjin,
Chuo-ku, Fukuoka 810-0001 Japan

表 1 本研究で用いる特徴ラベル

大カテゴリ > 小カテゴリ	項目	特徴ラベル数	付与上限数	特徴ラベルの例
アクセサリー・ジュエリー > ピアス	素材	36	6	ゴールド類, チタン, レジン・樹脂 等
食器・キッチン > 汁椀・ボウル・鉢	素材	5	3	陶器, 磁器, 漆器, 木, ガラス
食器・キッチン > 汁椀・ボウル・鉢	手法・技法	9	3	陶芸, ガラス工芸, 木工 等

3. 提案手法

ハンドメイド作品を扱う CtoC の EC サイトでは、作家が作成する作品のテキスト・画像・特徴ラベルを基に分類を行うが、情報不足・ラベルの信頼度・付与率が低いことが課題となる。本節では、作品の特徴ラベル分類の詳細、各課題と提案手法の詳細について述べる。

3.1 本研究で扱う作品の特徴ラベル分類

本項では、ハンドメイド作品を扱う CtoC の EC サイトにおける作品の特徴ラベル分類について説明する。作品は、予め定義された大カテゴリ及び小カテゴリに分類されている。小カテゴリごとに作品の特徴を表現できる特徴ラベルが定義されており、作家が各作品に対する特徴を踏まえ、特徴ラベルを付与する。本研究で分類の対象とする特徴ラベルを表 1 に示す。今回は計三種類の特徴の項目における特徴ラベルを対象とした。一つ目として、特徴ラベル数の多い「アクセサリー・ジュエリー > ピアス」カテゴリの「素材」項目を対象とした。二つ目、三つ目として、視覚的に特徴ラベルの分類が容易であると想定し、「食器・キッチン > 汁椀・ボウル・鉢」カテゴリの「手法・技法」項目及び「素材」項目を対象とした。作品に対して、特徴ラベルを 0 個以上付与することが可能であり、特徴の項目ごとに付与可能な特徴ラベルの上限数が決まっている。本研究では、作品に対するマルチラベル分類を行う。

3.2 手法 1: テキスト・画像の分類結果の和集合

作品の特徴ラベル分類を行うための特徴量の一つとして、作品のタイトルや説明文に出てくる単語を特徴量として利用する。ハンドメイド作品を扱う CtoC の EC サイトでは作家がタイトルや説明文の作成を行うため、テキスト中に作品の特徴ラベルを分類するための情報が不足することが課題として挙げられる。例えば、ある作品の場合、「アクセサリー・ジュエリー > ピアス」カテゴリの「素材」項目の特徴ラベルにおいて、「ゴールド類」が正例となるが、「ゴールド類」の特徴ラベルを分類するための情報がなく、テキストのみでは、分類が困難なことがある。一方、同作品の画像から、部分的に金色の素材を使った箇所を目視できる場合には、特徴ラベルとして「ゴールド類」が分類可能と考えられる。

上記の相互補完的なデータの性質を踏まえ、本研究では、テキストのみでは、特徴ラベルの分類が困難な作品に対しては、画像を特徴量として利用し、テキスト・画像両方の分類結果の和集合を求める手法を提案する。

3.3 手法 2: 特徴ラベルの単語でのキーワードマッチ

CtoC の EC サイトでは、ラベルの信頼度及びラベル付与率が低いことから、作品の特徴ラベルの付与漏れ・分類誤りの考慮も必要となる。まず、作品の特徴ラベルの付与漏れについて説明する。作品の特徴ラベルは、作家が作品登録時に付与を行っているため、付与漏れが発生することが考えられる。例えば、ある作品の場合、説明文中に、「天然石」を使用している旨の文言があり、本来であれば、「アクセサリー・ジュエリー > ピアス」カテゴリの「素材」項目における「天然石」の特徴ラベルが付与されるべきであるが、付与されていないことがあった。

次に、テキストを特徴量として用いた分類器で、特徴ラベルの分類時に発生する分類誤りについて説明する。テキストを特徴量とする分類器による分類を行ったとしても、特徴ラベルの信頼度が低いことにより、一定数の分類誤りが発生することが想定される。そのため、テキスト中の情報を基に分類誤りを是正する方法が必要となる。

以上の課題から、本研究では、課題 1 に対して、テキスト・画像両方の分類結果の和集合を求め、課題 2 に対して、特徴ラベルの単語がテキスト中に出現する場合は、特徴ラベルを付与するキーワードマッチを組み合わせた手法を提案する。

4. 評価

4.1 データセット

2022 年 6 月 29 日時点の minne の作品において、表 1 に基づくカテゴリの作品を抽出した。抽出する際には、特徴ラベルが付与された作品のみに限定し、既に削除された作品は除外した。また、タイトル及び説明文が入力されていない作品も除外した。本研究では、抽出した作品のうち、実験に用いる作品数として、「アクセサリー・ジュエリー > ピアス」カテゴリは 10,000 件を用いた。「食器・キッチン > 汁椀・ボウル・鉢」カテゴリは約 2,000 件を用いた。前処理として、各作品のタイトルと説明文については、正規化、URL 及び記号の削除を実施し、形態素解析は MeCab[11] 及び NEologd[9] を使用し、名詞のみを抽出した。画像は、一作品に対して、複数枚付与可能であるが、本研究では、作品画像の一枚目が作品の特徴を適切に表現していると仮定し、一枚目の画像のみを利用した。画像のサイズは、224 ピクセル × 224 ピクセルにリサイズし、各ピクセルの値に対して、正規化を行った。学習に用いる画像は、ランダムにズームする zoom_range の値を

表 2 テキスト・画像の分類結果の和の手法名称

テキスト	画像	名称
keyword	Image	keyword-Image
TF-IDF	Image	TF-IDF-Image
TF-IDF-keyword	Image	TF-IDF-keyword-Image

0.2 に設定し、水平方向に入力をランダムに反転する設定を追加した。

4.2 分類に用いる特徴量

4.2.1 テキスト

テキストを特徴量とする 3 種類の手法について述べる。一つ目の手法は、作品のタイトル及び説明文に出現する単語において、特徴ラベルの単語が出現した場合に、特徴ラベルとして予測する手法（以下、keyword と呼ぶ）である。二つ目の手法は、作品のタイトル及び説明文に対して、TF-IDF で特徴ベクトル化を行う手法（以下、TF-IDF と呼ぶ）である。三つ目の手法は、TF-IDF の分類結果と keyword との結果の和集合を求める手法（以下、TF-IDF-keyword と呼ぶ）である。

分類器については、事前に予備実験を行い、scikit-learn のマルチラベル分類* で用意されている分類器のモデルで性能比較を行った。その結果、多層マルチパーセプトロン (MLP) が一番分類性能が高かったため、本研究では、MLP のみを用いた。MLP はデフォルトの設定を用いている。学習とテストの比率は 9:1 とした。TF-IDF の単語の最大次元数は 150,000 とした。

4.2.2 画像

今回は画像識別で用いられる CNN のモデルである ResNet50[5] を利用した。ResNet50 では、ImageNet データセットで学習されたモデルの重みを用いた。ResNet50 の出力層側にある全結合層を削除し、出力次元の平坦化を行い、各カテゴリにおける特徴ラベルのクラス数に応じた全結合層を追加した。本研究では、ResNet50 の重みは固定し、全結合層のパラメータを学習させる手法を Image と定義する。各画像は学習・検証・テストの比率を 8:1:1 とした。最適化には確率的勾配降下法を用い、損失関数はバイナリ交差エントロピー、学習率は 0.001、モメンタムは 0.9 とした。学習時のエポック数は 10 とした。

4.2.3 テキスト・画像の分類結果の和

テキスト・画像の分類結果の和集合を求める手法を用いる。手法名はそれぞれ、テキスト・画像での命名規則を合わせたものとする。表 2 にテキスト・画像の分類結果の和の手法名称の定義を記載する。

4.3 全体の結果及び考察

全体の F1-score(micro) の結果を表 3 に示す。テキストの行は、各手法における最良値を記載している。テキスト

表 3 テキストと画像の F1-score(micro) の結果

カテゴリ	食器・キッチン > 汁椀・ボウル・鉢		アクセサリー ・ジュエリー > ピアス
	素材	手法・技法	素材
特徴の項目 特徴ラベル数	5	9	36
テキスト	0.908 (TF-IDF)	0.905 (TF-IDF)	0.530 (TF-IDF-keyword)
画像	0.642	0.598	0.152

における「食器・キッチン > 汁椀・ボウル・鉢」カテゴリはいずれも TF-IDF が最良であった。「アクセサリー・ジュエリー > ピアス」カテゴリの「素材」項目については、最良であった TF-IDF-keyword の結果を記載した。

いずれの分類結果においても、画像よりもテキストを特徴量とした場合が分類性能が良いことがわかる。画像の分類性能が低い要因としては二つ考えられる。一つ目は、用いる画像数が少ないことが挙げられる。関連研究で用いている画像数は、数十万 [2]~数百万 [10] が用いられている。本研究では、最大で 10,000 枚の画像を用いており、用いた画像数が少ないことが挙げられる。本報告では、実験時間の制約から利用する画像数を抑えたため、画像数を増加させた場合の評価を今後進めたい。二つ目は、ImageNet で学習したクラスの粒度が本研究で分類したい特徴ラベルよりも大きいことが挙げられる。表 1 に本研究で分類したい特徴ラベルの例を示しているが、例えば、ImageNet で学習するクラスとして、食器や陶器に関するクラスは含まれていなかった。そのため、本研究で扱う食器のさらに細かい特徴までは分類することは難しいと考えられる。

特徴ラベルのクラス数及び特徴ラベル数について述べる。「食器・キッチン > 汁椀・ボウル・鉢」カテゴリの「素材」項目及び「手法・技法」項目が「アクセサリー・ジュエリー > ピアス」カテゴリの「素材」項目よりもテキスト、画像両方とも分類性能が向上している。要因としては、特徴ラベルのクラス数の違いが考えられる。テキスト、画像両方とも特徴ラベルのクラス数が少ないほど分類性能が向上していることがわかる。また、図 1 に「アクセサリー・ジュエリー > ピアス」カテゴリの「素材」項目における特徴ラベル毎の結果を示す。この結果から、ラベル数が比較的多い特徴ラベルは若干分類性能が向上していることがわかる。画像数を増加させた場合における、特徴ラベル数の増加に伴う分類性能についての評価を今後進めたい。

4.4 課題 1 に対する結果及び考察

作品のテキスト・画像単体では、特徴ラベルを分類するための情報が不足しているという課題 1 に対して、手法 1: テキスト・画像の分類結果の和集合を求めることが有効であったかを評価する。

「食器・キッチン > 汁椀・ボウル・鉢」カテゴリの「手法・

* <https://scikit-learn.org/stable/modules/multiclass.html>

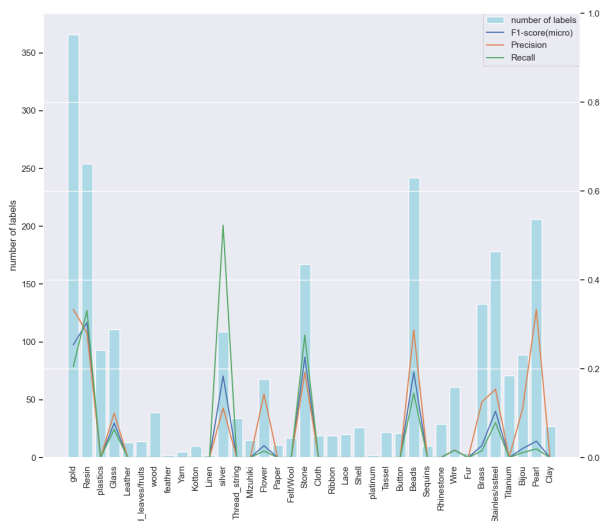


図 1 「アクセサリ・ジュエリー > ピアス」カテゴリの「素材」項目の 10,000 件の Image の結果

技法」項目における TF-IDF 及び TF-IDF-Image の特徴ラベル毎の結果を表 4 に示す。表の値は F1-score(micro) の結果を記載している。陶芸 (ceramic.art), 木工 (woodworking), ポーセリンアート (porcelain.art) のクラスにおいて, TF-IDF-Image は, TF-IDF と比べ, F1-score が下がっている。この結果の要因としては, 画像による誤判定の影響で Precision が下がったためである。陶芸 (ceramic.art) においては, Recall の値は上がっているため, テキストと画像の分類結果の和集合を求めることにより, Recall 向上は確認することができた。次に, 「食器・キッチン > 汁椀・ボウル・鉢」カテゴリの「手法・技法」項目における TF-IDF 及び TF-IDF-Image の Precision, Recall, F1-score の結果を表 5 に示す。この結果からも, TF-IDF-Image は, Recall の値は向上するが, Precision, F1-score は下がることがわかる。ただし, 陶芸 (ceramic.art) が正例の作品において, テキストでは木工 (woodworking) と誤分類したが, 画像では陶芸 (ceramic.art) と予測できた作品もわずかながら見つかった。以上のことから, テキスト・画像の分類結果の和集合を求めることによる, 相互補完的な分類を行うためには, 画像認識の精度改善が今後の課題として考えられる。

4.5 課題 2 に対する結果及び考察

作品の特徴ラベルの付与漏れ・分類誤りを考慮しなければならない課題 2 に対して, 手法 2: 特徴ラベルの単語によるキーワードマッチ, が有効であったかを評価する。評価には, キーワードマッチが有効であった「アクセサリ・ジュエリー > ピアス」カテゴリの「素材」項目における TF-IDF と TF-IDF-keyword の分類結果を用いる。ここで, TF-IDF では分類ができず, TF-IDF-keyword に

て分類ができた例を「分類誤り」, また, 正例としては付与されていないが, TF-IDF-keyword にて分類ができた例を「付与漏れ」と呼ぶ。これらの例は, 目視による個別の判断が必要となるため, 以下では, 分類誤りと付与漏れについて, それぞれ三件ずつを抽出し, その結果を考察している。

以下は「分類誤り」の例である。あるガラスを素材として用いた作品では, TF-IDF では「ガラス」を分類できなかったが, TF-IDF-keyword は正しく「ガラス」を分類できた。同様に「ビジュリー」, 「フラワー」といった特徴ラベルについても正しく分類することができた例が存在した。以上から, 分類器の分類結果に誤りがあったとしても, キーワードマッチにより, 分類誤りを防ぐことが可能となることを示した。同一のテキストに対して, TF-IDF を通して得られた特徴量を用いた場合に, キーワードマッチで検出可能な単純な例で分類誤りが発生してしまう理由については, 今後の調査課題である。

以下は「付与漏れ」の例である。例えば, 正例として, 「ゴールド類」, 「シルバー類」は付与されていない作品や, 同様に「ステンレス・スチール」, 「レジン・樹脂」, 「フラワー」が付与されていない作品について, キーワードマッチによる分類ができた。これらの作品においては, 文章中に特徴のキーワードが出現していること及び作品を目視した結果, 特徴ラベルを正例として付与することが正しいことを確認した。以上から, 作品登録時に特徴ラベルの付与が漏れていたとしても, キーワードマッチにより, 付与漏れを防ぐことが可能となることを示した。

5. まとめ

本研究では, 作品の特徴ラベル分類に関する精度と付与率の向上に向けて, テキスト・画像を用いた作品の特徴ラベル分類に取り組んだ。作家が作成する作品のテキスト・画像単体では, 特徴ラベルを分類するための情報が不足する課題に対しては, テキスト・画像両方の分類結果の和集合を求めることで Recall が向上することを確認した。また, テキストだけでは分類できない作品は画像分類器で分類が可能な例が存在することも確認できた。作品の特徴ラベルの信頼度・付与率が低いことに伴う分類誤り・付与漏れの課題に対しては, 特徴の単語によるキーワードマッチにより分類誤り・付与漏れを防止できることを確認した。

今後の課題としては, 三点挙げられる。一つ目は, 画像分類の性能向上に向けた検討である。実験に用いる画像数の増加及び転移学習における事前学習モデルのクラスと分類したいクラスの粒度が異なる場合の解決策の調査を進めていく。二つ目は, 特徴ラベルのクラス数の差異や特徴ラベル数の差異による考察である。このために, 今回対象

表 4 「食器・キッチン > 汁椀・ボウル・鉢」カテゴリの「手法・技法」項目における TF-IDF と TF-IDF-Image の特徴ラベル毎の結果

特徴ラベル	Precision		Recall		F1-score	
	TF-IDF	TF-IDF-Image	TF-IDF	TF-IDF-Image	TF-IDF	TF-IDF-Image
陶芸 (ceramic_art)	0.9184	0.7154	0.9548	0.9943	0.9362	0.8321
木工 (woodworking)	0.9090	0.7692	0.9090	0.9090	0.9090	0.8333
ポーセリンアート (porcelain_art)	1.0000	0.8888	0.7619	0.7619	0.8648	0.8205

表 5 「食器・キッチン > 汁椀・ボウル・鉢」カテゴリの「手法・技法」項目における TF-IDF と TF-IDF-Image の結果

	Precision	Recall	F1-score
TF-IDF	0.915	0.895	0.905
TF-IDF-Image	0.761	0.920	0.833

とした特徴の項目以外の他項目での実験を実施していく。三つ目は、Feature-level fusion や Decision-level fusion の他手法での比較実験である。

参考文献

- [1] Cevahir, A. and Murakami, K.: Large-scale Multi-class and Hierarchical Product Categorization for an E-commerce Giant, *In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 525–535 (2016).
- [2] Chen, L., Chou, H., Xia, Y. and Miyake, H.: Multi-modal Item Categorization Fully Based on Transformer, *In Proceedings of the 4th Workshop on ECNLP*, pp. 111–115 (2021).
- [3] Feng, Z., Tang, J., Liu, J., Yin, W., Feng, S., Sun, Y. and Chen, L.: Alpha at SemEval-2021 Task 6: Transformer Based Propaganda Classification, *In Proceedings of the 15th International Workshop on SemEval-2021*, pp. 99–104 (2021).
- [4] GMO ペパポ株式会社: minne, <https://minne.com/>.
- [5] He, K., Zhang, X., Ren, S. and Sun, J.: Deep Residual Learning for Image Recognition, *In Proceedings of the IEEE Conference on CVPR*, pp. 770–778 (2016).
- [6] Kiela, D., Grave, E., Joulin, A. and Mikolov, T.: Efficient Large-Scale Multi-Modal Classification, *In Proceedings of the AAAI Conference on Artificial Intelligence* (2018).
- [7] Peng, J., Feng, S., Wang, Y., Hou, H., Lian, F. and Kang, Z.: Multimodal Product Identification: Submission to Watch and Buy 2021 Challenge, *In Proceedings of the 1st Workshop on Multimodal Product Identification in Livestreaming and WAB Challenge*, pp. 9–13 (2021).
- [8] Ross, K., Hungler, P. and Etemad, A.: Unsupervised multi-modal representation learning for affective computing with multi-corpus wearable data, *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–26 (2021).
- [9] Toshinori, S.: Neologism dictionary based on the language resources on the Web for Mecab (2015).
- [10] Zahavy, T., Krishnan, A., Magnani, A. and Mannor, S.: Is a Picture Worth a Thousand Words? A Deep Multi-Modal Architecture for Product Classification in E-Commerce, *In Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 7873–7881 (2018).
- [11] 工藤 拓: MeCab, <https://taku910.github.io/mecab/>.