

# 商品画像に着目したオンラインフリーマーケットにおける購買行動予測に関する研究

松本 淳太郎<sup>1</sup> 大知 正直<sup>1,a)</sup> 山下 雄大<sup>1</sup> 榊 剛史<sup>1,b)</sup> 森 純一郎<sup>1</sup> 坂田 一郎<sup>1</sup>

概要：本研究では、現在急速に成長しつつある C2C 事業、オンラインフリーマーケットサービスの「メルカリ」について、商品情報から消費者の購買行動が予測可能かどうかの検証を行う。Convolutional Neural Network および Random Forest を組み合わせることで、消費者の行動を予測する手法を提案する。また、評価実験より消費者の購買行動には画像情報が影響することを示す事ができた。本研究は急成長しているオンラインフリーマーケット事業の消費者行動について新たな知見を提供するものである。

## Purchasing Behavior Prediction in Online Flea Market Considering on Product Images

MATSUMOTO JUNTARO<sup>1</sup> OCHI MASANAO<sup>1,a)</sup> YAMASHITA YUTA<sup>1</sup> SAKAKI TAKESHI<sup>1,b)</sup>  
MORI JUNISCHIRO<sup>1</sup> SAKATA ICHIRO<sup>1</sup>

### 1. はじめに

近年、消費者間取引、すなわち C2C (Consumer to Consumer) 事業に変化が見られる。web の発達や、多機能型携帯電話 (通称スマートフォン、以下、スマートフォンという) の台頭により、C2C 事業がスマートフォンのアプリケーションとして多くの人々に身近なサービスとなりつつある [4]。本研究では、C2C 事業のうち、売買するものは特定しないサービス、例えばネットオークションサイトのヤフオク! や、メルカリ、楽天フリマのように消費者同士で本や服、インテリア、家電など様々なものを売り合うオンラインフリーマーケット (Flea Market) に着目する。

オンラインフリーマーケットサービスが流行し始めているものの、その流行はここ 2 年ほどの話であり、サービス運営サイドも、出品者や消費者などの利用サイドも、十分な知見とノウハウを得られていないのである。そのため、本研究ではオンラインフリーマーケットサービスを運営、

利用していく上での知見を探ることを目指す。特に、サービスを利用している消費者の購買行動を予測するための手法を提案することを目指す。本研究では、オンラインフリーマーケットサービスで最も規模が大きいものの一つである「メルカリ」を分析対象とする。

### 2. 提案手法

#### 2.1 着想

消費者の購買行動の予測モデルを構築する上で、まずはメルカリでのユーザの行動系列を分析する。メルカリでは毎日約 50 万の商品が売買されており<sup>\*1</sup>、利用者の 3 分の 2 以上がスマートフォンからアクセスしている [3]。多くの人はメルカリへのアクセスは非常にしやすいため、良い商品は非常に早い段階で売買される。事実、本研究で使用するデータに、何ページ目の商品一覧ページで売買が成立したかをヒストグラムにしたデータがある。それが図 1 である。図 1 を見ればわかるように、ほとんどの商品は早い段階で売買が決定している。3 分の 1 の商品は商品一覧ページの 10 ページ目までに掲載されたものである。

以上のことを踏まえて、オンラインフリーマーケット

<sup>1</sup> 東京大学  
The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo  
113-8654, Japan

a) ochi@ipr-ctr.t.u-tokyo.ac.jp

b) sakaki@ipr-ctr.t.u-tokyo.ac.jp

<sup>\*1</sup> <https://www.mercari.com/jp/info/>

表 1 本章で使用する特徴量一覧

特徴量	説明
日時	商品が出品された日時 .
価格	商品の価格 .
カテゴリー	商品カテゴリー . 本研究ではメンズのアウター 11 カテゴリーを使用 .
「良い」評価数	消費者がする出品者への評価のうち , 「良い」評価数 .
「普通」評価数	消費者がする出品者への評価のうち , 「普通」評価数 .
「悪い」評価数	消費者がする出品者への評価のうち , 「悪い」評価数 .
出品数	出品者が今までにメルカリで出品した商品数 .
商品サイズ	商品 (メンズアウター) のサイズ .
商品状態	商品の使用状態 . 「新品 , 未使用」 「未使用に近い」 「目立った傷や汚れなし」 「やや傷や汚れあり」 「傷や汚れあり」 「全体的に状態が悪い」 の 6 通り回答できる .
配送料負担	送料を負担するのが出品者か購入者か
配送方法	「クロネコヤマト」 「ゆうパック」 「らくらくメルカリ便」 などの発送方法 .
配送元地域	発送する地域 . 47 都道府県で回答 .
配送日の目安	配送までどの程度の時間がかかるか . 「1~2 日」 「2~3 日」 「4~7 日」 から選ぶ .
画像情報	画像を BVLC GoogleNet Model の学習済みモデルを使用し , ベクトル化したもの .
タイトル	商品タイトル . Word2Vec の Doc2vec モデルを使い , ベクトル化 .
ブランド	商品ブランド . Word2Vec の Doc2vec モデルを使い , ベクトル化 .
説明文	商品の説明文 . Word2Vec の Doc2vec モデルを使い , ベクトル化 .

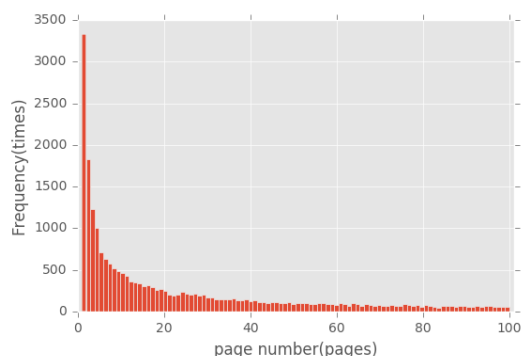


図 1 メルカリ (カテゴリー : メンズアウター) で商品が売れた際のページ数に関するヒストグラム

サービス「メルカリ」において、人気のある商品は非常に早い段階で売れてしまう。それを理解している消費者はそういった商品を探し当てた際になるべく即決で買っている、と予想できる。そのためメルカリにおいて消費者が購買行動に進む際に重要となってくるのが商品への「直感的印象」であると考えられ、特に商品一覧ページに記載されている商品画像と価格が消費者購買行動に強く影響すると考えられる。

## 2.2 オンラインフリーマーケットにおける購買行動予測

オンラインフリーマーケットサービス「メルカリ」において消費者の購買行動を把握するために、本研究では機械学習を用いて購買行動の予測を試みる。具体的には、ある商品を、その商品が持つ特徴から「売れる商品」と「売れ

ない商品」に分類する二値分類問題と捉える。

ここで、機械学習に「売れる商品」と「売れない商品」を学習させるために、学習データにおける「売れる商品」と「売れない商品」の定義付けを明確にする必要がある。本研究では、「売れる商品」を「出品されてから 2 時間以内に売れた商品」と定義し、「売れない商品」を「出品されてから 18 時間経っても売れなかった商品」と定義する。

このように、本研究では、オンラインフリーマーケットサービス「メルカリ」の消費者購買行動を予測するために、研究手法としては機械学習を用いてメルカリの商品情報を特徴量とし、ラベルを「売れる商品」「売れない商品」の 2 つとし、予測する。その際、メルカリのサイトデザイン上の特性として、タイトル、価格、そして画像の情報が消費者の購買行動に重要な意味を持つと予想し、特に画像情報を重点的に着目して予測する。

## 2.3 特徴量設計

表 1 は本研究で予測に使用する特徴量の一覧である。

画像情報の特徴量を抽出するのに用いる手法は、深層学習の一種、CNN (Convolutional Neural Network) を用いる。CNN は画像認識の分野で広く使われており [1]、画像関係の研究で CNN が多用されている [1]。例えば Krizhevsky らは、約 1,500 万の画像を 22,000 のラベルに分類したデータに対してこの手法を用いて、画像認識にて 6 割強の予測精度を出した [1]。本研究では、Andreas らの研究と同様に、GoogleNet を用いてベクトル化を行った [2]。

日時の情報に関して、日にちの情報は商品が出品された月、月の上旬か中旬か下旬かの情報をベクトルとして抽出した。時間の情報に関しては1日のどのタイミングで出品したかによって変わってくるため、24時間のどのタイミングで出品されたかをベクトル化した。

価格は商品の価格をそのまま特徴量として組み込んだ。

カテゴリに関しては、本研究ではメンズのアウトターという大きなカテゴリの中で、「テラードジャケット」「Gジャン・デニムジャケット」「レザージャケット」「ダウンジャケット」「ライダーズジャケット」「ナイロンジャケット」「フライトジャケット」「ダッフルコート」「ピーコート」「トレンチコート」「モッズコート」の11カテゴリを選んだ。

出品者情報として「良い」評価数、「普通」評価数、「悪い」評価数、そして出品数を用いた。商品売る出品者は売買が成立すると他者から評価を受けることができる。配達日を守ったかどうかや、再出品やダブルブッキングなどのマナー違反な行為があったか、コメントに対する返信の仕方など様々な出品者の行動を売買が成立すると購入者が評価できる仕組みとなっている。

商品サイズとしては、「FREESIZE」「XS以下」「S」「M」「L」「XL」「XL以上」の7種類を用いた。本研究ではカテゴリはメンズのアウトター11カテゴリのみを使用しているため、この商品サイズ情報はどの商品にも1つの要素となっている。

商品状態としては、「新品、未使用」「未使用に近い」「目立った傷や汚れなし」「やや傷や汚れあり」「傷や汚れあり」「全体的に状態が悪い」の6通りの回答全てを用いた。

配送料負担は、配送料を負担するのが購入者なのか出品者なのかの情報である。配送方法としては、「未定」「らくらくメルカリ便」「普通郵便(定形, 定形外)」「クロネコヤマト」「ゆうパック」「はこBOON」「レターパック」「クリックポスト」「ゆうメール」「ポストパケット/ゆうパケット」「ポストパケット」の11通り全てを用いた。配送元地域は出品者が住んでいる地域であり、47都道府県全てを用いた。配達日の目安は売買取引が成立してからどの程度の日を要して配送が開始されるかの情報である。

タイトル、ブランド、そして説明文は、自由記述となっている。そこで、それぞれの特徴量を Word2Vec モデルに学習させ、ベクトル化した。

## 2.4 学習器

本研究では、高い精度が出せる、特徴量に対して重要度を出しやすいこと、特徴量が大きいものの分類も簡単に行えることなどの有用性から、Random Forest を用いることとした [5]。

表 2 5000 円以上の商品データに関する予測

特徴量	Accuracy	Precision	Recall	F value
メタ+画像+テキスト	0.60	0.61	0.61	0.61
画像	0.57	0.58	0.58	0.57
テキスト	0.50	0.50	0.42	0.45
メタ	0.63	0.63	0.62	0.62
メタ+画像	0.61	0.62	0.61	0.60
メタ+テキスト	0.59	0.59	0.59	0.60

## 3. 評価実験

### 3.1 データ

本研究では2016年12月28日~2017年1月20日に、株式会社メルカリのオンラインフリーマーケットサービスサイト「メルカリ」のページを web スクレイピングし、商品情報を取得した。具体的には、メルカリよりカテゴリ毎にそれぞれの商品一覧ページにおいて一定時間において情報取得した後、各商品ページに進み、商品情報を取得した。

### 3.2 実験概要

評価実験においては、メルカリの消費者購買行動「売れる商品」「売れない商品」の予測に関して以下の特徴量を用いて予測する。以下では、画像情報を「画像情報」と、タイトル、ブランド、説明文のテキスト情報を「テキスト情報」と、価格などそれ以外の情報を「メタ情報」と呼ぶ。

- 全ての特徴量、画像情報、テキスト情報、メタ情報で予測をする。
- テキスト情報とメタ情報のみで予測をする。
- 画像情報とメタ情報のみで予測をする。
- メタ情報のみで予測をする。

以上の予測を行う。その際、価格帯を5000円以内の商品データに絞る予測、5000円から1万円以内の商品データに絞る予測、1万円以上の高価格帯商品データに絞る予測、の3種類を行い、違いを考察する。その理由としては、機械学習にかけた時、価格情報の影響が非常に高く出たため、本研究で画像情報やテキスト情報、出品者の出品時間帯など細かいデータの影響を見たかったため、あまり強く影響がでてしまう価格の影響を小さくするためである。

### 3.3 実験結果

メルカリの購買行動を予測する際、予測精度を測る尺度として、主に正解率 Accuracy、適合率 Precision、再現率 Recall と、F 値の4つの尺度を用いて評価する。

表2には価格5000円以内の商品に絞って予測した結果を、表3は価格帯が5000円以上1万円以下の商品に絞った結果を、最後に表4には1万円以上の商品に関する購買予測をした結果を示した。

表 3 5000 円から 1 万円までの商品データに関する予測

特徴量	Accuracy	Precision	Recall	F value
メタ + 画像 + テキスト	0.71	0.85	0.52	0.66
画像	0.59	0.60	0.58	0.59
テキスト	0.50	0.48	0.43	0.47
メタ	0.71	0.77	0.61	0.69
メタ + 画像	0.71	0.83	0.55	0.66
メタ + テキスト	0.69	0.82	0.49	0.62

表 4 1 万円以上の商品データに関する予測

特徴量	Accuracy	Precision	Recall	F value
メタ + 画像 + テキスト	0.70	0.86	0.48	0.61
画像	0.59	0.59	0.63	0.61
テキスト	0.49	0.48	0.47	0.49
メタ	0.72	0.82	0.54	0.65
メタ + 画像	0.71	0.82	0.54	0.64
メタ + テキスト	0.68	0.77	0.51	0.61

#### 4. 考察

表 2, 表 3, ならびに表 4 において, まずメタ情報がメルカリの購買行動に非常に大きな役割を果たしていることがわかる. F 値で見ると, 表 2 においても表 3 においても表 4 メタ情報のみで予測をした結果が最も良い値を出している.

メタ情報のうち, 特に重要と判定されたのが値段と出品者のメルカリ出品数であることがわかる. 値段は本研究の購買行動予測において, 最も重要な商品特徴量の 1 つと言えるだろう. また, メルカリアは C2C 事業サービスであり, 消費者は物を買う時, 見知らぬ人から買うことになるため, 出品者の信頼度を測る大きな要素であるメルカリでの過去の出品実績, ならびに良い評価などが大きく影響と考えられる.

次に, 画像情報とテキスト情報の購買行動への影響度を考察する. 表 2, 3, 4 にはそれぞれ画像のみで購買行動を予測した結果, ならびにテキスト情報のみで購買行動を予測した結果が載っている. メタ情報のみの予測には到底およばないものの, 画像情報に関しては 60 % 代の判定水準を保っており, 多少なりとも正しい予測をしていることがわかる. しかしながらテキスト情報は 50 % 付近に止まっており, テキスト情報だけではなかなか予測が難しいことがわかる. これは, メルカリアのアプリケーションサイトデザイン上の理由があると考えられる. また, 全ての価格帯において, メタデータとテキストデータの 2 つを合算して予測している予測精度より, メタデータ, テキストデータ, そして画像データの 3 つの要素で予測する方が少しだけ予測精度が上がっているのがわかる. これは, 新しい特徴量, すなわち画像データを足すことで予測精度が上がっている

ことになり, 画像データの予測精度の向上に寄与していると考えられる.

#### 5. 結論

本研究の貢献は以下の通りである.

- メルカリアの消費者購買行動予測において影響のある特徴量を示したこと.
- その中でも, 画像情報が多少なりとも予測に貢献することを示したこと.
- 現在成長を遂げているオンラインフリーマーケットの分野において, 主に出品者の出品の仕方に関する知見をデータ解析の観点からできたこと.

まず, オンラインフリーマーケットサービスのメルカリアに対し, 商品情報を特徴量とした商品に対する消費者の購買行動を予測することに成功し, その際, 消費者の購買行動に影響のある特徴量を取り上げることに成功した.

メタ情報, 特に値段の情報や, 出品数など出品者への信頼度を評価する情報, そして出品する時間帯が消費者の購買行動予測に影響がでることを確認し, そこから消費者が購買行動へ進む際, そういったメタ情報に敏感である可能性を示唆した.

次に, 多くの特徴量の中で画像情報を取り上げ, 購買行動予測の仕方を複数とることによって画像の情報が購買行動予測に多少の正しい貢献ができていたことが解明できた.

今後は, 具体的にどういった画像情報が影響しているのか, 例えば, 画面の明暗や画像の解像度, ロゴマークの有無など, 定量的, 定性的な分析を行い, フリーマーケットにおける購買行動予測をより精緻化していきたい.

#### 参考文献

- [1] Alex Krizhevsky, Ilya Sutskever, G. E. H.: Learning Visual Clothing Style with Heterogeneous Dyadic Co-occurrences, *NIPS* (2012).
- [2] Andreas Veit, Balazs Kovacs, S. B. e. a.: ImageNet Classification with Deep Convolutional Neural Networks, *Third International Conference* (2015).
- [3] Company, T. N.: スマホユーザーの 51 % にあたる 2,656 万人がオークションフリマサービスを利用. [http://www.netratings.co.jp/news\\_release/2016/02/Newsrelease201600223.html](http://www.netratings.co.jp/news_release/2016/02/Newsrelease201600223.html).
- [4] Oh, W.: C2C Versus B2C: A Comparison of the Winner's Curse in Two Types of Electronic Auctions, *International Journal of Electronic Commerce* (2002).
- [5] 大越里佳: ゲーミフィケーションが学習の継続性に与える影響 (2016).