

# 商品説明文のテキスト特徴量を用いた キャラクターグッズの売却時間予測モデルの提案

玉井貴大, 丹一真, 小西哲平, 鷲尾隼弥, 深川大路, 宿久洋 (同志社大学文化情報学部)

## 研究背景

- メルカリ内で最も取引件数の多いカテゴリ：キャラクターグッズ<sup>[1]</sup>
- 推し活の普及でフリマサービスはグッズ取引の重要な手段になった
  - フリマサービス利用の理由：「購入予算を増やしたい」などの経済的動機<sup>[2]</sup>
  - ・グッズを購入する予算を増やすために、商品を短時間で売却したい
- ➔ 商品を短時間で売却するために出品者側の出品設定の工夫が必要
- 商品説明文のわかりやすさが売り上げや早期に買い手が見つかるかに影響する<sup>[3]</sup>

- 商品説明文のわかりやすさを定量的に評価：テキスト特徴量
  - テキスト特徴量：テキストデータを数値やカテゴリ情報に変換した分析可能な指標

## 研究目的

キャラクターグッズにおける商品説明文のテキスト特徴量が売却までにかかる時間に与える影響を定量的に明らかにする

[1] mercari. 2024. "メルカリ、サービス開始11周年記念インフォグラフィックスを公開". [https://about.mercari.com/press/news/articles/20240627\\_11th\\_infographics/](https://about.mercari.com/press/news/articles/20240627_11th_infographics/). 最終閲覧日2024/11/30.

[2] mercari. 2023. "リユース人気カテゴリー「推し活（エンタメ・ホビー）」における消費動向調査". [https://about.mercari.com/press/news/articles/20231108\\_category\\_survey\\_1/](https://about.mercari.com/press/news/articles/20231108_category_survey_1/). 最終閲覧日2024/11/30.

[3] 廣川田拓実, 櫻井暁, & 伏見卓哉. (2023). 転移学習を利用した多クラス分類タスクに基づく適切な商品説明文の抽出. In Web インテリジェンスとインタラクション研究会 予稿集 第19回研究会 (pp. 89-91).

## モデル概要

使用データ ○ キャラクターグッズ・キャラクターグッズ以外のカテゴリ：それぞれ120,596件ずつランダムで抽出

テキスト特徴量 ○ 商品説明文を文法軸・内容軸・属性軸で評価<sup>[4]</sup>

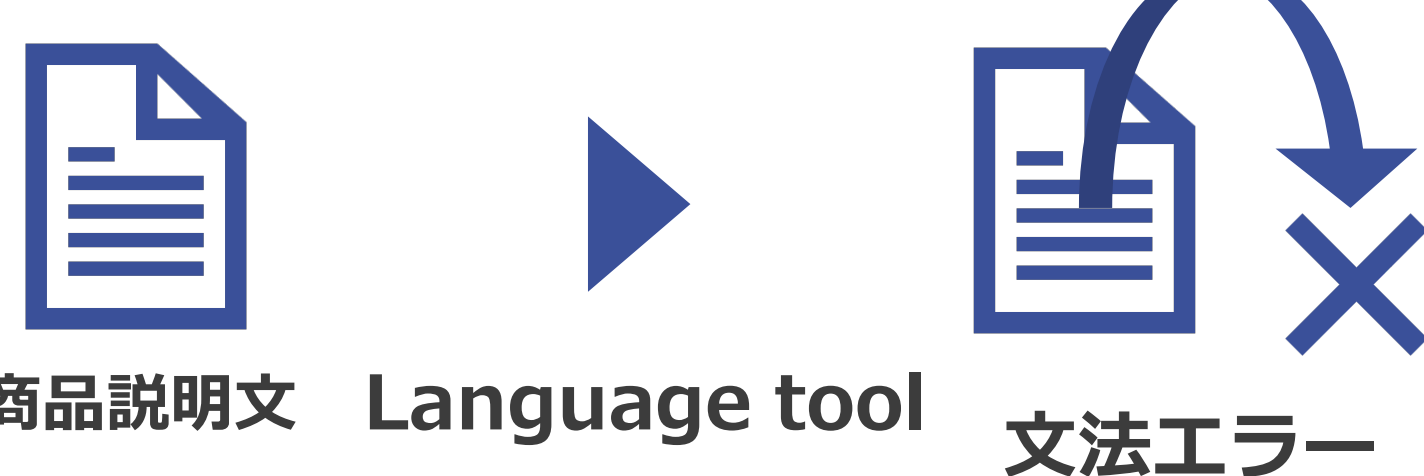
文法軸：商品説明文の文法的な正確さや読みやすさを評価

内容軸：商品説明文の内容の質と理解しやすさを評価

属性軸：購買意欲を高めるための商品説明文の特徴を評価

### 文法軸

#### 破綻性



MeCabによって形態素解析した情報にLanguage toolの文法ルールを適用しエラーを検出

$$\text{文法エラー率} = \frac{\text{文法エラー数}}{\text{総単語数}}$$

#### 可読性

- 文の平均の長さ
- 単語の平均の長さ

### 内容軸

#### 一貫性



トピック①[高品質, きれい]  
トピック②[スケール, ポーズ]

文1:[①0.7, ②0.3]  
文2:[①0.6, ②0.4]  
文3:[①0.8, ②0.2]

1文ごとに、どのトピックに属するかを確率で表現

文1と文2の  
コサイン類似度：0.99  
文2と文3の  
コサイン類似度：0.94

➢ 一貫性  
コサイン類似度の平均  
0.97



MeCab

BERT

#### 流暢性

各トークンの出現確率を計算

パープレキシティを計算

$$\text{パープレキシティ} = \exp\left\{-\frac{1}{N} \sum_{i=1}^N \log(P(w_i|w_{<i-1>}))\right\}$$

N: 商品説明文内の単語数  
P(w<sub>i</sub>|w<sub><i-1></sub>): 前の単語列から次の単語w<sub>i</sub>を予測される確率

商品説明文の自然さを評価  
文中の各単語が文脈に基づいて  
どれだけ予測しやすいか

### 属性軸

#### 人気度



MeCab

- ・ MeCabとNEologDによる形態素解析
- ・ 固有単語ごとの登場件数を抽出

MeCabにより商品説明文の固有単語を判定

Ex.)商品説明文の例

大人気「鬼滅の刃」の炭治郎フィギュアです  
公式商品で、彩色も綺麗！炭次郎の魅力を完全再現！  
新品未開封でコレクションに最適です

「鬼滅の刃」  
「炭次郎」  
登場した固有単語

それぞれ登場した  
固有単語に対して  
登場件数で重みづけ

人気度

➢ 人気度 = 商品説明文の固有単語の種類 × 固有単語の登場件数

表1. キャラクターグッズにおける固有単語登場件数上位3件

固有単語	登場件数
鬼滅の刃	22,270
amiibo	19,088
あつまれ どうぶつの森	9,515

#### 語彙の多様性

$$\text{TTR} = \frac{\text{異なり語数}}{\text{総語数}}$$

#### 丁寧さ

$$\text{丁寧さ} = \frac{\text{敬語を含む文の出現回数}}{\text{総文数}}$$

#### 文字の種類割合

- 漢字の割合
- ひらがなの割合
- カタカナの割合
- アルファベットの割合

## 使用モデル LightGBM

- 勾配ブースティングアルゴリズムに基づく機械学習モデル
- 誤差を小さくするようにデータを分割
- 売却までにかかる時間を目的変数とし、各特徴量をLightGBMに適用

説明変数 ○ 出品時の基本情報もその他の特徴量として考慮

テキスト特徴量 + その他の特徴量

目的変数 ○ 更新時間：商品情報が変更された時間

売却までにかかる時間 = 更新時間 - 出品時間

(※ 売却済みの商品の内)

#### その他の特徴量

- 価格
- 発送の方法
- 商品状態
- 発送元の地域
- 送料の負担
- 発送にかかる日数

## 結果・考察

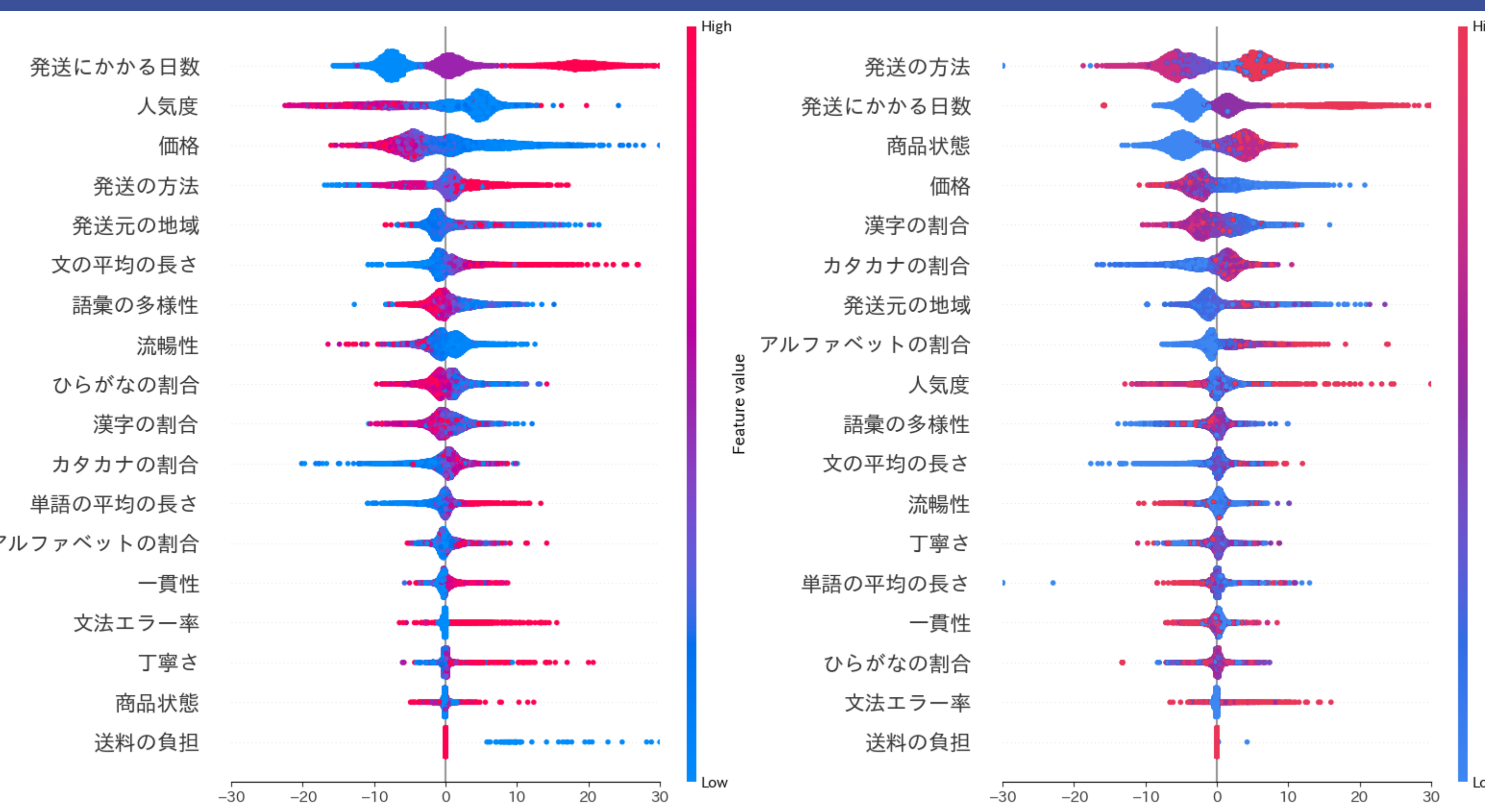


図1. SHAPでの各特徴量の影響度 (左:キャラクターグッズ, 右:キャラクターグッズ以外の商品)

### SHAP値を用いた特徴量の解釈

- ◆ キャラクターグッズでは、人気度が高いと売却までにかかる時間は短くなる傾向
- ◆ キャラクターグッズでは、豊富に語彙を用いて、敬語を交えないほうが、売却までにかかる時間は短くなる傾向
- ◆ 商品説明文全体において、1文はできるだけ短いほうが売却までにかかる時間は短くなる

#### 考察

キャラクターグッズにおいて、作品名などの固有単語を交えながら、簡潔でカジュアルな表現の説明文が早期売却につながる

## 今後の展望

- モデルの精度向上
  - ・ 説明力の高い特徴量を導入 (Ex.) 可読性指数の導入
- 施策の提案
  - ・ 商品説明文の修正機能の実装
  - ・ 早期売買のための商品説明文のレコメンド機能

## 謝辞

本研究では、株式会社メルカリの協力のもと、国立情報学研究所提供の「メルカリデータセット」を使用しました。深く感謝申し上げます。