

商品へのカテゴリ付与誤り事例に対する修正作業支援の検討

井上翔太 稲田和明 張信鵬

株式会社 MonotaRO

{shota.inoue, kazuaki.inada, xinpeng_zhang}@monotaro.com

概要

Eコマースで取り扱う商品に付与されるカテゴリ情報は、商品検索をはじめとするEコマースの様々なサービスを提供する上で重要な情報である。商品の新規登録やカテゴリ構造の変更の際に、商品に対するカテゴリ情報の付与が適切かを確認する必要があるが、人手による保守のコストが大きい。一方で、商品のカテゴリを自動的に分類する研究が存在するが、Eコマースで実用できるレベルの非常に高精度なモデルの構築は現実的ではないため、人手による確認が必要不可欠である。そこで本研究では、商品へのカテゴリ付与・整備作業の支援を目指して、商品とカテゴリの合致度を判定するモデルを作成し、現状のモデルの問題点や改善点について分析する。

1 はじめに

Eコマース上で取り扱われる商品の情報を正しく登録・整備することは、健全かつ有用なサービスを提供するために必要不可欠である。特に商品へ付与されるカテゴリ情報は、キーワードマッチによる商品探索やカテゴリ情報を活用した商品推薦など、Eコマースの様々なサービスに使用される最も重要な情報の1つである。商品へのカテゴリ情報の付与は商品の新規登録やカテゴリ構造変更の際に必要となり、単純な作業ミスや確認漏れ、あるいは判断が難しいケースで、不適切なカテゴリを割り当ててしまうことがある。カテゴリ付与の誤りは、Eコマース内の様々なサービスの性能低下を招き、ユーザー体験の悪化を引き起こす原因となる。

そのためEコマースの管理者は、取り扱う商品に対して常に正しいカテゴリ情報が登録されている状態を保持したいが、商品数が膨大になると商品情報の登録・整備に対する人手での確認は困難になる。たとえばモノタロウ¹⁾のカテゴリは、24種類の最上

1) <https://www.monotaro.com/>

位カテゴリの元で最大深さ8の木構造として存在しており、深さ3のカテゴリ数は約1,000種類で、それらのカテゴリに平均34,000の商品が紐づいていることから、カテゴリ名やカテゴリ構造を変更した際に発生する、商品に付与されているカテゴリの精査に多大なコストが掛かることがわかる。一方で、商品に対する適切なカテゴリを分類する研究が存在する[1]が、実際のEコマースで利用するためには非常に高い精度が必要であるため、自動分類のみで実現することはリスクが高く現実的ではない。

そこで本研究では、商品情報の登録・整備の作業支援を目指して、商品とカテゴリが合致しているかを判定するモデルを作成し、その出力を詳細に分析する。まずカテゴリ情報の変更が発生した事例を人手で分析することで、カテゴリ情報の付与に誤りがある商品を効率的に抽出するとともに、カテゴリ構造の変更についての理解を深める。その後、商品とカテゴリの合致度を複数のモデルで算出した上で、カテゴリ情報の付与に誤りの可能性が高いと考えられる商品群に対し、現状のモデルの問題点や改善点について議論する。

2 関連研究

Eコマースの商品情報整備として、商品へ付与するカテゴリを自動で分類する研究が存在する[1, 2, 3, 4]。商品カテゴリの分類には自然言語処理やWebマイニング技術が利用され、その分類方針として主にカテゴリ階層情報を利用しない分類と、カテゴリツリーの階層構造を利用する分類の2種類に分けられる[4]。本研究では、カテゴリ階層情報を利用しない分類手法を参考に、商品-カテゴリ間の合致度を判定するモデルを構築する。

またEコマースにおける商品カテゴリ分類と類似するタスクとして、国際特許の分類がある。星野ら[5]は、特許分類の階層性を考慮してHard Negativeを作成し、Contrastive Learningを利用した教師ありSimCSE[6]で、入力の特許文書と似た特許を検索す

る枠組みを提案している。本研究では星野らを参考に、事前学習済み BERT および SimCSE で商品とカテゴリに含まれるテキスト情報をベクトル化し、そのベクトル間のコサイン類似度を商品-カテゴリ間の適合度として扱う。

商品情報を常に正しい状態に整備・保持することは、アノテーション誤り検出 [7, 8] のタスクと類似する。ただし、カテゴリ構造の整備などの E コマース側の変更によって適切な状態が変化する可能性があり、1 度正しい状態を付与できれば完了ではない点が一般的なアノテーションとは異なる。本研究では、E コマースで求められる正しい商品情報の保持を目的とするフレームワークの構築を目指し、カテゴリ情報の付与誤り事例について分析を行う。

3 商品とカテゴリ間の合致度の計算

本研究では、商品とカテゴリ間の合致度合いを測るモデルとして、以下の3つの手法を採用する。

N-gram: 商品のタイトル・説明・属性を連結した文字列、およびカテゴリツリー上の末端2つのカテゴリ名を連結した文字列に対する、それぞれの文字 N-gram 集合 ($n=1,2,3$) 間の類似度を、商品とカテゴリ間の合致度を表すスコアとして用いる。商品側の文字列情報に比べてカテゴリ側の文字列情報が短くなるのが予想されるため、N-gram 集合の類似度計算手法に Simpson 係数を採用する。

BERT: 文字 N-gram 集合の類似度では、テキストの表層しか考慮できず意味的な合致度を考慮できない。そこで商品とカテゴリに含まれるテキスト情報の意味をとらえるために、事前学習済みの BERT を採用する。商品の入力テキストとして、タイトル、説明文、属性名とその値を、カテゴリの入力テキストとして商品に付与されたカテゴリ名を最上位階層から順にすべて連結させた文字列を使用する。また、BERT の出力として得られる商品とカテゴリのベクトル間のコサイン類似度をそれらの間の合致度を表すスコアとして用いる。

SimCSE: E コマースサイトの商品とカテゴリの対応関係を学習するために、上述の BERT モデルに対して教師ありの SimCSE [6] によるファインチューニングを行う。教師あり SimCSE の学習データの作成にあたって、モノタロウにおけるカテゴリのツリー構造と、特定のカテゴリを指定してそのカテゴリ以下の商品一覧を検索結果として提供する「カテゴリ検索」の仕組みを利用する。具体的には、カテゴリ

検索における各カテゴリのカートイン数上位 10 商品をそのカテゴリに合致する正例とし、そのカテゴリと同じ階層の兄弟カテゴリからランダムにサンプルした商品をそのカテゴリに合致しない Hard Negative として採用する。

4 カテゴリ付与誤りの分析

4.1 事前調査: カテゴリ付与誤りの調査

商品-カテゴリ間の合致度に関する分析を効率良く実施するために、過去にカテゴリが変更されたことがある商品を調査した。カテゴリ付与誤りが存在した事例の探索に加えて、商品とカテゴリの適切な関係についての理解を深める。

調査では、モノタロウにおいて 2022/12/01 から 2023/11/30 の間でカテゴリに変更があった商品の移動前後のカテゴリを抽出した。抽出したカテゴリからランダムに選択した 50 事例に対する人手による分析結果を表 1 に示す。

表 1 のとおり、明らかな誤りとみなせる「誤配置」は全体の約 30%であった。また「用途修正」は、より適切な利用シーンを示すカテゴリに変更されている場合であり、最も多い 40%を占めていた。その他にも、より詳細なカテゴリに変更されている「詳細化」や、過剰なカテゴリの細分化を避けるためにより一般的なカテゴリに変更されている「一般化」などの事例が存在した。なお「その他」は、カテゴリ構成自体の大幅な変更が行われた事例や、正誤の判断が困難な事例に割り当てた。

4.2 商品-カテゴリ間の合致度の分析

商品-カテゴリ間の合致度付与に関する分析用のデータとして、表 1 の「誤配置」の 16 事例のうち、最上位カテゴリの異なる以下の3つのカテゴリに紐づく全商品に対して、3 章の各手法で商品-カテゴリ間の合致度合いを計算した。

- オフィス家具/照明/清掃用品 > 清掃用品 > ウェス > 紙ウェス > カットタイプウェス
- 物流/保管/梱包用品/テープ > 保管用品 > コンテナ > コンテナ用オプション > フタ
- 医療・介護用品 > 救急・衛生 > 絆創膏 > 一般用絆創膏

なお BERT の事前学習済みモデルとして、BERT

表 1: カテゴリ変更が発生した商品におけるカテゴリ付与誤り事例とその割合

ラベル	事例数	事例 (代表の商品名と移動前後のカテゴリ)
誤配置	16 (32%)	商品名: フローリングワイパースタンド 移動前: オフィス家具/照明/清掃用品 > 清掃用品 > ウェス > ディスペンサー・スタンド 移動後: オフィス家具/照明/清掃用品 > 清掃用品 > 掃除用具 > フローリングワイパー > フローリングワイパー用スタンド
用途修正	20 (40%)	商品名: LED 電球 MIRROR(ミラー) ザ・バルブ 移動前: 厨房機器・キッチン/店舗用品 > 店舗什器・備品 > 店舗什器 > 什器用アクセサリ > 店舗・棚下照明 移動後: オフィス家具/照明/清掃用品 > 照明 > 電球 > LED 電球 > その他 LED 電球
詳細化	5 (10%)	商品名: UAG iPad Air(第 5 世代)OUTBACK Case(ブラック) 移動前: オフィスサプライ > PC > スマートフォンアクセサリ > カバー・ケース・ポーチ > タブレットケース > その他機種タブレットケース 移動後: オフィスサプライ > PC > スマホ/タブレット関連アクセサリ > カバー・ケース・ポーチ > タブレットケース > iPad ケース
一般化	4 (8%)	商品名: ポリイミドフィルム PIF075 移動前: 物流/保管/梱包用品/テープ > テープ > 工業用フィルム > PI フィルム 移動後: 物流/保管/梱包用品/テープ > テープ > 工業用フィルム
その他	5 (10%)	商品名: バケツ入り カーシャンプー 移動前: 自動車用品 > 洗車・清掃 > ボディー用 > ワックス in シャンプー 移動後: 自動車用品 > 洗車・清掃 > ボディー用 > カーシャンプー > 固形カーシャンプー

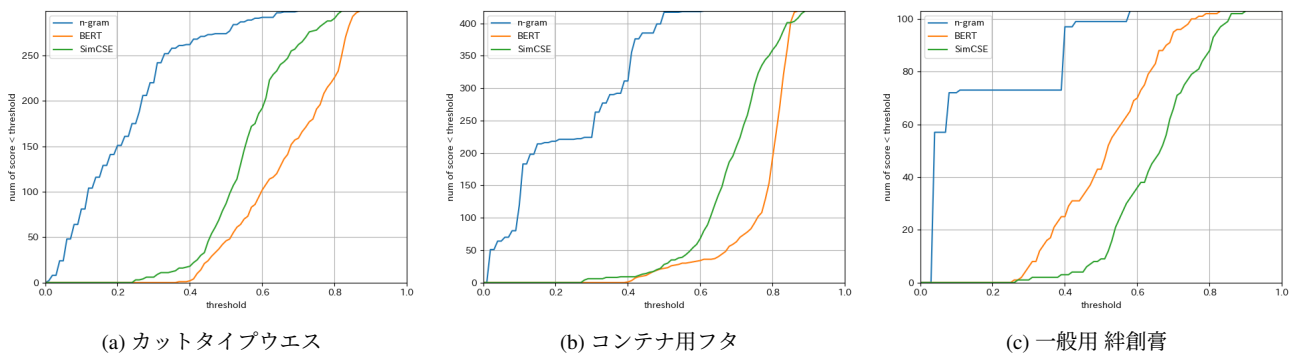


図 1: 各カテゴリにおけるしきい値以下の合致度をもつ商品の割合

base Japanese v3²⁾を利用した。また SimCSE の学習データとして、2022/12/01 から 2023/11/30 の間のカテゴリ検索のログを集計し、カテゴリが合致する正例を 38,555 件、合致しない Hard Negative の候補を 249,272 件取得した。なお、エポック数を 3、バッチサイズを 128、学習率を $5e-5$ 、最大入力トークン数を 512 に設定し、エポック毎に正例と負例が 1:1 になるようにダウンサンプリングした。

図 1 に各カテゴリと各手法ごとの合致度と商品数の分布を示す。図 1 より、N-gram は合致度が 0.2 未満の商品が半数以上を占め、0.6 を超えるものがほとんど無いことがわかる。これは N-gram が文字の表層一致をもとに合致度を計算しているためであ

り、少数の関連性の強い商品を見つける役には立つが、不適格な商品を見つけることには向いていないといえる。また BERT と SimCSE は、N-gram よりも合致度がより広く分布しており、商品に応じてカテゴリの合致度に差が出ていることが読み取れるが、これらの手法の間の明確な差は確認できなかった。

BERT と SimCSE の差を確認するために、それぞれのモデルで合致度合いが低い事例を目視で分析した。各カテゴリごとの商品-カテゴリ間のスコア下位 10%、すなわち各手法が相対的に誤りが高いと判断した上位 10%に含まれる商品に対して、著者が「誤り」「合致」「その他」の 3 択で判定した。なお「誤り」は明らかにカテゴリ付与誤りである事例、「合致」はカテゴリ名と商品が明らかに合致する事例、「その他」は状況によって「誤り」と「合致」の

2) <https://huggingface.co/cl-tohoku/bert-base-japanese-v3>

表 2: 合致度が低い商品-カテゴリに対する目視評価

対象カテゴリ	誤り		合致		その他	
	BERT	SimCSE	BERT	SimCSE	BERT	SimCSE
a. カットタイプウエス	1	10	23	10	6	10
b. コンテナ用フタ	0	3	42	23	0	16
c. 一般用絆創膏	0	2	6	9	5	0

判断が揺れる事例を指す。

表 2 に評価結果を示す。表 2 より「カットタイプウエス」のカテゴリでは、SimCSE の合致度が低い商品には「誤り」の商品が 30 商品中 10 個含まれていたが、BERT の合致度が低い商品には「誤り」は 30 商品中 1 個しか発見できなかった。次に「コンテナ用フタ」カテゴリでは、SimCSE で「誤り」と判定できた事例は少なかったが、一方で BERT では「合致」に相当する事例のみが存在していた。また「一般用絆創膏」のカテゴリでは、SimCSE で「誤り」と判断できる事例を発見できたが、BERT では「その他」に相当する事例のみが存在していた。以上より、「誤り」の検出は SimCSE が BERT よりも優れていたが、状況によって判断が分かれる「その他」は、カテゴリによっては SimCSE が BERT よりも優れていると判断できなかった。

4.3 低合致度の商品-カテゴリの実例分析

まず対象カテゴリの商品として適切だが、SimCSE において合致度の低かった商品を抽出し、低合致度の要因とみられる特徴を分析した。

第 1 に、テキスト情報の不足によって低い合致度を与えている事例が存在した。たとえば、「折りたたみコンテナフタ³⁾」という商品は「コンテナ用フタ」カテゴリに属するが、この商品には説明文が記載されていなかった。本研究で SimCSE の正例として利用した商品は、説明文などの商品情報が十分に記載されており、説明文の無い「折りたたみコンテナフタ」は商品情報のテキストを上手くベクトル表現に埋め込めなかったと考えられる。商品のテキスト情報が不足している事例の改善には、商品画像の類似性やキーワード検索におけるクリック履歴をはじめとするユーザー行動などのテキスト以外の情報の活用が必要である。

第 2 に、商品のカテゴリ付与合致度合いを判定するために十分なテキスト情報を含むが、明示的にカテゴリを示す表現を含まないために、低い合致度を付与している事例も存在した。たとえば、「オイル

ガードジェットスパン⁴⁾」は「紙ウエス」に分類される商品だが、明示的に「紙ウエス」を表すテキストが商品に含まれていなかった。商品情報には、液体の吸収や原料にパルプを使用していることなど「紙ウエス」と推論できる情報が記載されているため、商品に含まれるテキストからその商品が何であるかを推定する問題を解くことで改善が期待できる。

また表 2 の「一般用絆創膏」カテゴリで観察された、BERT での「その他」事例について分析した。「エレバンプレスタットミニ⁵⁾」という商品は、「一般用絆創膏」カテゴリに属しているが、兄弟カテゴリである「穿刺保護用絆創膏」としての機能も持つことがわかる。このように明確な線引ができない商品が存在する兄弟カテゴリが存在するため、SimCSE など教師あり学習の手法を使用する場合は、学習データの選定方法をより厳密に評価する必要があるといえる。

5 おわりに

本研究では、誤ったカテゴリを割り当てた商品を発見し、適切なカテゴリを設定する作業の支援を目指し、商品の持つテキスト情報とカテゴリ名の間合致度を計算するモデルを作成し、その出力について詳細な分析を実施した。作成したモデルの中で最も有用と判断された SimCSE による合致度が低い事例を観察したところ、正しくカテゴリ付与が間違っている事例を抽出できていたが、その一方で商品情報として明示的に記述されていない場合や、そもそもテキスト情報が不足している場合で、カテゴリ付与誤りでない事例に低い合致度を付与していることも確認できた。今後の課題として、明示的にカテゴリを示す表現が欠けた商品の種別予測や画像やユーザー行動ログなどのテキスト以外の情報を活用した合致度計算モデルの改善が挙げられる。また、商品-カテゴリ間の合致度を推定するモデルを実際の E コマースの商品情報整備作業に利用し、作業者の負荷を軽減できているかを評価する予定である。

3) <https://www.monotaro.com/g/04216787/>

4) <https://www.monotaro.com/g/01610720/>

5) <https://www.monotaro.com/g/01299502/>

参考文献

- [1] Yiu-Chang Lin, Pradipto Das, and Ankur Datta. Overview of the sigir 2018 ecom rakuten data challenge. In **eCOM@SIGIR**, 2018.
- [2] Zornitsa Kozareva. Everyone likes shopping! multi-class product categorization for e-commerce. In **Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: human language technologies**, pp. 1329–1333, 2015.
- [3] Lei Chen and Hirokazu Miyake. Label-guided learning for item categorization in e-commerce. In **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers**, pp. 296–303, 2021.
- [4] Alexander Brinkmann and Christian Bizer. Improving hierarchical product classification using domain-specific language modelling. **IEEE Data Eng. Bull.**, Vol. 44, No. 2, pp. 14–25, 2021.
- [5] 星野雄毅, 内海祥雅, 中田和秀. Contrastive learning を利用した類似特許検索. 言語処理学会 第 29 回年次大会, 2023.
- [6] Tianyu Gao, Xingcheng Yao, and Danqi Chen. SimCSE: Simple contrastive learning of sentence embeddings. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 6894–6910, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [7] Jan-Christoph Klie, Bonnie Webber, and Iryna Gurevych. Annotation error detection: Analyzing the past and present for a more coherent future, 2022.
- [8] Leon Weber and Barbara Plank. Activeaed: A human in the loop improves annotation error detection. In **Findings of the Association for Computational Linguistics: ACL 2023**, pp. 8834–8845. Association for Computational Linguistics, 2023.