

ハンドメイド作品を対象としたECサイトにおける 単語の出現頻度を用いた稀観品の検出

野村 孔命^{1,a)} 三宅 悠介^{1,b)}

概要 :

ハンドメイド作品を扱う CtoC の EC サイトの流通額が増加している。ハンドメイド作品は、その希少性や独創性が価値の一端を担っていることから、運営者はこれらを満たす稀観品の発見に取り組む。しかし、出品者から大量の作品が登録されるにつれ、運営者が目視によってこれらの作品を発見することは困難になる。そのため、希少性や独創性を表現する尺度と基準を設け、自動的かつ継続的に多くの稀観品を検出する必要がある。また、多様な特性のハンドメイド作品を扱う EC サイトにおいては、特性に依存しない汎用的な尺度と基準であることが望ましい。本報告では、稀観品の検出を、ある集合における外れ値の検出とみなしてこれを解く。提案手法では、作品に常に設定される属性情報の内、作品の特徴を端的に表現する作品タイトルに含まれる単語の出現頻度の合計を尺度として用いる。そして、特定の作品カテゴリ内で相対的に尺度の値が低い作品を外れ値として検出する。評価では、実際のハンドメイド作品を扱う EC サイトの作品に対して提案手法を適用し、複数の外れ値の基準や作品カテゴリに対する、推薦や所属カテゴリの誤り検知への応用を考察する。

Detection of Rare Works Using Term Frequency on Electronic Commerce Site for Trading Handmade Works

KOMEI NOMURA^{1,a)} YUSUKE MIYAKE^{1,b)}

Abstract: The market size of Electronic Commerce (EC) site dealing with handmade works is increasing. The value of handmade works include the rarity and originality, and operators try to discover rare works that satisfy these value. However, it becomes difficult for operators to manually discover these works as exhibitors register a large number of works. Thus, it is necessary to determine measure and criteria that express rarity and originality to detect many rare items automatically and continuously. Furthermore, the measure and criteria should be independent of the characteristics of the work in EC sites dealing with handmade works with various characteristics. In this paper, we regard the detection of rare items as the detection of outliers in a certain set. In our proposed method, we use the frequency of word occurrences contained in the title of a handmade work among the attribute information of the handmade works for measurement. The proposed method detect works with relatively low value on this measure within a category of works as outliers. We apply the proposed method to the works of real EC sites that deal with handmade works and consider multiple outlier criteria. Furthermore, we consider using the proposed method for recommendation and category error detection.

1. はじめに

国内のクラフトホビー分野の市場規模は 2016 年時点で

8,900 億円に達している。その中でも、ハンドメイド作品は、これを取り扱う消費者-消費者間（以下 CtoC）の消費者向け電子商取引（以下 EC）の発展に伴って、インターネットでの流通総額が増加している [1]。ハンドメイド作品は、作家によってデザインや制作が手作業で行われたものであるため、その希少性や独創性が価値の一端を担って

¹ GMO ペパボ株式会社 ペパボ研究所, Pepabo Research and Development Institute, GMO Pepabo, Inc.

a) komei.nomura@pepabo.com

b) miyakey@pepabo.com

いる。そこで、ハンドメイド作品を扱う CtoC の EC サイトの運営者は、希少性や独創性を満たす珍しい作品、すなわち、稀観品の発見に取り組む。

ハンドメイド作品を扱う CtoC の EC サイトにとって、稀観品の用途は大きく二つある。一つ目は販売の促進である。希少性や独創性が価値の一端であると認知しているハンドメイド作品を求める利用者に対して、これを満たすような稀観品を紹介することで、購買率の向上が期待される。二つ目は検索性の向上である。稀観品には、ある作品集合において希少性や独創性の度合いが相対的に高すぎるものが含まれる。このような稀観品は所属する作品集合が誤っている可能性が高いと考えられる。EC サイトでは、作品の検索性を向上させるために、作品に共通する特性毎に分類したカテゴリが出品者によって作品に設定される。しかし、カテゴリ設定が適切でない場合があるため、運営者はこれらの作品の発見に取り組む。そのため、稀観品を検出し用途に応じた分類を効率的に行うことは、ハンドメイド作品を扱う CtoC の EC サイトにとって有用である。

運営者が目視によって手動で稀観品を発見する場合、出品者から大量に作品が登録されるにつれ、多くの作品を確認することが困難になる。効率よく多くの稀観品を発見するためには、希少性や独創性を表現する尺度と基準を設定し、自動的かつ継続的に稀観品を検出して、用途に応じて分類する必要がある。

稀観品を発見するためには、共通する特性を持つ作品集合から相対的に珍しい特性を持つ作品を探す必要がある。しかし、作品集合毎に共通する特性は異なるため、作品集合における一般的な作品も異なる。このことから、取り扱う作品の種類が多様かつ継続的に変化する EC サイトにおいては、作品集合に共通する特性を手動で個別に求めることなく稀観品を検出できることが望ましい。作品集合に共通する特性を個別に求めることなく実行可能な稀観品の検出方法として、稀観品と推薦における意外性の共通点に着目することで、利用者の評価から意外性を見出す推薦手法 [2-5] を用いることができると考えられる。しかしながら、運営者は多くの稀観品を検出することで販売の促進や検索性の向上につながることから、利用者の評価がない作品からも稀観品を検出できることが望ましい。利用者の評価を必要としない稀観品の検出方法として、投稿型の料理レシピサイトにおいて、珍しい食材を含むレシピを検出する手法 [6] がある。しかし、この手法では検出したレシピの珍しさに対する分類が行えない。ハンドメイド作品を扱う EC サイトでは、稀観品のうち販売促進に利用する作品と検索性の向上に利用する作品を分類する基準が必要となる。

本報告では、これらの課題を解決するために、カテゴリにおける作品タイトルに含まれる単語の出現頻度を尺度として、この尺度が相対的に低い作品を外れ値として検出す

る手法を提案する。提案手法では、カテゴリ内における作品タイトルに含まれる単語の出現頻度を表現する値として、TF-IDF [7] における IDF の値を利用する。単語の出現頻度のみから作品集合における一般性を求めるため、作品集合に共通する特性を手動で個別に把握する必要がない。また、利用者の評価のような追加情報ではなく、作品に常に設定される属性情報を用いることで、検出対象を広げることができる。検出した稀観品は、作品タイトルに含まれる単語の出現頻度の標準偏差の傾向に違いがあることを利用して用途別に分類する。

実験では、提案手法によって稀観品を検出できることを示すために、実環境から取得した作品集合に対して提案手法を適用し、作品毎に単語の IDF の値から作品の評価値を求め、作品の評価値の上位に稀観品が位置付けられることを確認した。また、稀観品を用途別に分類できることを確認した。

本稿の構成を述べる。2 章では、稀観品の検出とその課題をまとめる。3 章では、提案手法の概要を示したあと、稀観品の評価値の算出と用途別の分類方法について述べる。4 章では、提案手法によって求めた作品の評価値の上位に稀観品が位置付けられることと、単語の出現頻度の標準偏差によって稀観品を用途別に分類できることを実験によって確かめる。5 章でまとめとする。

2. 稀観品の検出とその課題

本章では、ハンドメイド作品を扱う CtoC の EC サイトにおける稀観品の検出とその課題について述べる。稀観品を発見するためには、共通する特性を持つ作品集合から相対的に珍しい特性を持つ作品を探す必要がある。しかし、作品集合毎に共通する特性は異なるため、作品集合における一般的な作品も異なる。このことから、取り扱う作品の種類が多様かつ継続的に変化する EC サイトにおいては、作品集合に共通する特性を手動で個別に求めることなく稀観品を検出できることが望ましい。

作品集合に共通する特性を個別に求めることなく実行可能な稀観品の検出方法として、利用者の購買履歴などの利用者の評価を利用する意外性を重視した推薦手法 [2-5] を応用できると考えられる。推薦の分野では、利用者に未知の情報や意外な情報を提供し驚きを与えることが、利用者の満足度の向上につながるとして、意外性を重視した推薦手法が研究されており、意外性の構成要素として新規性がある [8-10]。新規性がある作品は利用者が見たことのない作品のことである。稀観品は作品集合内の一般的な作品に比べ数が少なく、多くの人が目にしていない可能性が高い作品なので、この新規性の要素を含んでいる。そのため、意外性を重視した推薦手法は、稀観品の検出にも利用可能であると考えられる。しかし、ハンドメイド作品を扱う CtoC の EC サイトにおいては、日々新しい作品が追加されるた

め、利用者の評価が存在しない作品が多く存在し、意外性を重視した推薦手法を稀観品の検出に応用できたとしても、これらの作品を検出の候補から除外してしまう。運営者は多くの稀観品を検出することで販売の促進や検索性の向上につながるため、利用者の評価がない作品からも稀観品を検出できることが望ましい。利用者の評価に依存せず稀観品を検出するためには、作品に設定される属性情報を利用する必要がある。

属性情報のみを利用する意外性を重視した推薦手法として、動画投稿サイトにおいて多くの人は知らないが視聴した際に利用者が興味を持つような動画を検出する手法が研究されている [11]。この手法が対象としている動画投稿サイトでは、利用者が動画の特徴を表す言葉をタグとして自由に設定できる。このタグの中に動画が良質であるがあまり知られていない動画に付与されているタグがあり、このタグが付与されている動画を学習の正解データとして利用する。しかし、ハンドメイド作品を扱う EC サイトにおいて、稀観品を表す属性情報が存在しない場合、学習に用いる正解データを定めることができないため、この手法は利用できない。また、利用者が稀観品を表す属性情報を設定できたとしても、日々追加される新しい作品には稀観品を表す属性情報が付与されていない。そのため、稀観品の検出に利用する属性情報は作品に常に設定されるものである必要がある。

また、常に設定される属性情報を利用する意外性を重視した推薦手法として、投稿型の料理のレシピサイトから珍しいレシピを抽出する手法が研究されている [6]。この手法では、レシピの説明欄に TF-IDF を応用した方法を適用することで、ある料理においては利用されることが少ないが、その他の料理では頻繁に利用される食材を抽出する。そして、抽出した食材を含むレシピを意外性があるレシピとして検出する。この手法では、その他の料理で頻繁に利用される食材を食材の一般性としてレシピの評価に反映させることで、入手困難な食材を含まないレシピを抽出している。しかし、この手法は検出したレシピの珍しさに対する分類が行えない。ハンドメイド作品を扱う EC サイトでは、運営者の用途に合わせて、販売促進に利用する作品と検索性の向上に利用する作品を分類する基準が必要となる。

3. 提案手法

2章で述べた課題を解決するために、特定の作品集合における作品タイトルに含まれる単語の出現頻度を尺度として、この尺度が相対的に低い作品を外れ値として検出する手法を提案する。提案手法の概要を図 1 に示す。特定の作品集合における単語の出現頻度を求めるために、まず、作品集合内の全作品のタイトルに対して形態素解析を適用し、作品タイトルを単語に分割する。次に、単語の出現頻度の低さを求めるために、抽出した全単語に対して IDF の値

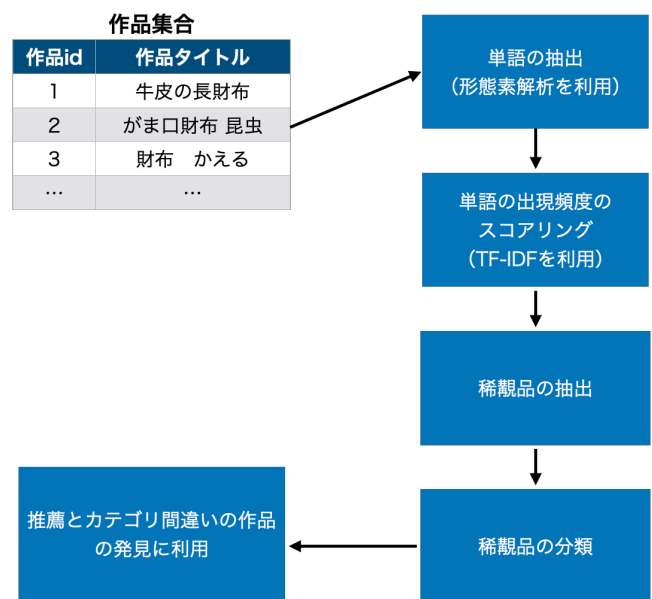


図 1 提案手法の概要

Fig. 1 Summary of the proposed method

を計算し、この IDF の値を単語のスコアとする。次に、作品毎に作品タイトルに含まれる単語のスコアが高い上位 N 件を合計し作品の評価値を計算する。作品の評価値は単語の出現頻度が低い単語が多く含まれているほど高くなり、この作品の評価値が高い作品を稀観品として検出する。最後に、検出した稀観品を用途別に分類するために、作品タイトルに含まれる単語のスコアの標準偏差の大きさを利用してこれらの分類を行う。

提案手法では、単語の出現頻度のみから作品集合における一般性を求めるため、作品集合に共通する特性を手動で個別に把握する必要がなく、作品に常に設定される属性情報を用いることで、検出対象を広げることができる。以降本章では、図 1 における提案手法の各処理について説明する。

3.1 作品タイトルに含まれる単語の抽出

提案手法では、常に作品に設定される属性情報のうち、作品の特徴を端的に表現していると考えられる作品タイトルに含まれる単語の出現頻度を尺度として稀観品の検出に用いる。単語の出現頻度を求める前処理として、作品タイトルを単語に分割し、その作品の特徴を表す単語を抽出する必要がある。単語の分割には、形態素解析エンジンとして MeCab [12] を採用し、それぞれの作品タイトルに対して形態素解析を行うことで、作品タイトルを単語に分割する。形態素解析によって得られた作品それぞれの単語集合から、品詞が名詞であるものを抽出する。ただし、単語の文字数が 1 である漢字以外の単語は、その 1 文字で単語として意味を持たないと考えられるため、品詞が名詞であっても除外する。例えば、アルファベット 1 文字の「a」などの文字は単語としての意味を持たないとみなして除外した。ま

た, MeCab の標準のシステム辞書では正しく分割できない固有表現や新語などが存在するため, Web 上の言語資源から得た新語が登録されている mecab-ipadic-NEologd [13] を辞書として利用した.

3.2 稀観品の検出

3.1 節で抽出した作品タイトルに含まれる単語集合を用いて, 単語の出現頻度の低さを計算する. 単語の出現頻度の低さの計算には, 文書内の単語の重要度を求める際に利用される TF-IDF における IDF を利用した. IDF は, 全文書のうち, ある単語が含まれている文書の割合の逆数である. ある文書集合における単語 t の $IDF(t)$ は式 1 で求められる.

$$IDF(t) = \log \frac{N_{doc}}{df(t)} \quad (1)$$

ここで, N_{doc} は全文書数であり, $df(t)$ は単語 t が出現する文書数である. IDF は, 文書集合における単語の出現頻度の低さを表した値であり, ある文書で現れる単語がその他の文書で現れないほど大きい値となる. 提案手法では, 作品タイトルを文書として単語それぞれの IDF を計算し, この IDF の値を単語のスコアとして稀観品の検出に利用する.

計算した単語のスコアを用いて, 稀観品の検出に用いる作品の評価値を求める. 作品の評価値は, 各作品のタイトルに含まれる単語のスコアが高い上位 N 件を合計することで求める. 作品 w の評価値 $E(w)$ は以下の式 2 で計算する.

$$E(w) = \sum_{t_i \in TOP(T_w, N)} x_{t_i} \quad (2)$$

ここで, $TOP(w, N)$ は作品 w が持つ単語集合 T_w において単語のスコアが大きい上位 N 件の単語, x_{t_i} は単語 t_i のスコアである. 式 2 では, 作品に単語のスコアが大きい単語が多く含まれるほど作品の評価値 $E(w)$ は大きくなる.

提案手法では, 作品のタイトルに含まれる単語のスコアが高い上位 N 件の合計を作品の評価値として計算するため, N 単語未満の作品は検出の対象外とする. 作品の評価値を作品タイトルに含まれる全単語のスコアの合計して求める場合, 作品タイトルに含まれる単語数は作品によって異なるため, 単語数が多いほど高い評価値となる. 本報告における稀観品は, 作品が持つ単語数に関係なく, 出現頻度が低い単語を含んでいる作品を検出する必要がある. そのため, 作品の評価値は, 作品のタイトルに含まれる単語のスコアが高い上位 N 件の合計値として求める. また, N の値は小さい方が提案手法が検出対象とする作品数は増加するが, 作品の評価値が同値となる場合が増加するため, N の値は検出対象となる作品数と作品の評価値のばらつきを考慮して設定する必要がある.

3.3 稀観品の分類

3.2 節で求めた評価値の高い作品を稀観品の候補とする. 提案手法ではこれらの作品を用途ごとに分類するため, 作品タイトルに含まれる単語のスコアの傾向の違いを利用する. 着目した傾向は以下である.

- 推薦による販売促進の効果が高いと考えられる作品は, 作品タイトルにカテゴリ内において出現頻度が低い単語を含みつつ, 出現頻度が高い単語も含んでいる. つまり, 作品タイトルに含まれる単語のスコアの標準偏差が大きい.
- カテゴリ間違いの作品は, 作品タイトルに含まれる殆どの単語がカテゴリ内において出現頻度が低い単語で構成されている. つまり, 作品タイトルに含まれる単語のスコアの標準偏差が小さい.

この傾向を利用して, 提案手法では, 作品タイトルに含まれる単語のスコアの標準偏差を作品毎に計算し, 標準偏差が小さい作品をカテゴリ間違いの作品として除外する. 作品 w のタイトルに含まれる単語 t_i の標準偏差 σ は, 式 3 で表される.

$$\sigma = \sqrt{\frac{1}{|T_w|} \sum_{t_i \in T_w} (x_{t_i} - \mu_w)^2} \quad (3)$$

ここで, T_w は作品 w に含まれる単語の集合, x_{t_i} は単語 t_i のスコア, μ_w は作品 w の単語のスコアの平均値である.

4. 実験

4.1 稀観品と単語の出現頻度の関係性

提案手法は 3.2 節で述べた作品の評価値を稀観品の検出に利用しており, 作品の評価値の上位に稀観品が位置付けられることを確認するために, GMO ペパボ株式会社が運営するハンドメイドの EC サイト minne [14] から作品を抽出し実験を行った. また, 作品の評価値の上位と下位に位置する作品を目視で確認し, 上位と下位それぞれに含まれる稀観品の割合を調査した.

検出対象の作品集合として, 財布カテゴリの作品 10,000 件を minne から取得した. 提案手法は 3.2 節で述べたように, 作品タイトルに含まれる単語の出現頻度が高い単語 N 件を利用するため, N 単語未満の作品を取り除く必要がある. 本実験では, 作品の評価値にばらつきがあり, かつ提案手法を適用できる作品数が多い N の値として 5 を設定し, 単語数が 5 未満の作品を取り除いた 6,904 作品を用いる. また, 稀観品のうちカテゴリ間違いの作品の例としてバッグカテゴリの作品 5 件, 推薦による販売促進の効果が高いと考えられる作品の例として minne の運営者がピックアップしている minne's セレクトの作品のうち財布カテゴリに属する作品 5 件を用いる.

また, minne では, 一人の出品者が同一のデザインや材料を用いた作品を出品することが多く, これらの作品の特

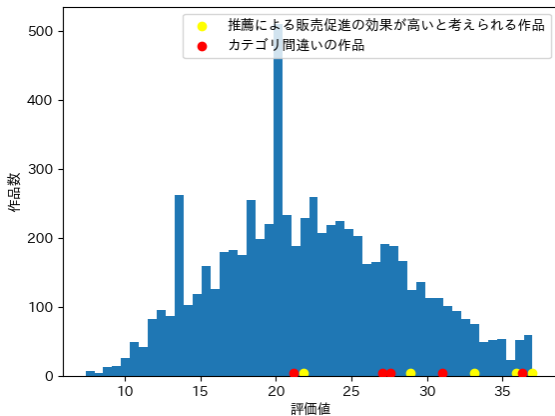


図 2 作品の評価値のヒストグラムと稀観品の例の分布

Fig. 2 Histogram of work evaluation and distribution of rare works

徴は類似しているため作品タイトルに同じ単語が頻繁に使われる。このことから、ある出品者のみが利用する特徴的な単語であっても、作品毎の単語の出現頻度は高くなるため、実験では、出品者毎に単語の出現頻度を計算することで単語スコアとして求めた。

4.1.1 評価値における稀観品の例の分布

提案手法で算出した作品の評価値の上位に稀観品の例が位置付けられることを確認するために、データセットに提案手法を適用し全作品の評価値を求め、作品の評価値における稀観品の例の分布を調査した。作品の評価値のヒストグラムと稀観品の例の分布を図 2 に示す。

図 2 から、稀観品の例が評価値の平均値である 22.32 付近から最大値にかけて分布していることが分かる。稀観品の例の一部は評価値の最大値付近に分布していることから、カテゴリにおける単語の出現頻度の低さを尺度として利用することで、稀観品の一部を上位に位置付けられることが確認された。しかし、評価値の平均値付近にも稀観品の例が分布しており、この尺度によって高く評価されない稀観品が存在することが確認された。

最大値付近と平均値付近に分布している稀観品の例の違いを説明するために、それぞれ作品のタイトルを表 1 に示す。表 1 に示す最大値付近の作品には、「Midsummer」や「北半球」などの財布カテゴリにおいて出現頻度が低い単語が多く利用されていた。一方で、表 1 に示す平均値付近の作品には、「ハチワレ」のように出現頻度が低い単語も含まれるが、その他の単語は、「コイン」や「カード」などの財布カテゴリにおいて頻繁に利用される単語であった。そのため、この作品の評価値は他の稀観品の例に比べ小さくなっている。

4.1.2 評価値の上位と下位に含まれる作品の比較

提案手法によって求めた作品の評価値の上位に、稀観品が多く分布することを確認するために、評価値の上位と下

表 1 評価値の最大値と平均値付近の稀観品の例

Table 1 Example of rare works near maximum and average evaluation

	作品タイトル
最大値付近の作品	がま口 story02/Midsummer festivities. -北半球二十四節気第 10 のはなし
平均値付近の作品	コインもカードも入るケース ハチワレねこ (コインケース・カードケース)

表 2 評価値の上位と下位に含まれる稀観品の割合

Table 2 Rate of rare works included in the high and low of the evaluation

	上位の作品	下位の作品
推薦による販売促進の効果が高いと考えられる作品	28(56%)	3(6%)
カテゴリ間違いの作品	10(20%)	0(0%)
稀観品 (小計)	38(76%)	3(6%)



図 3 評価値が上位の推薦による販売促進の効果が高いと考えられる作品の例

Fig. 3 Example of works that can be recommended with high evaluation

位に含まれる稀観品の割合を調査した。データセットとして定義した 6,909 件の作品に対して提案手法を適用することで、作品の評価値を計算し、評価値が高い順に作品を並び替えた。並び替えた作品の上位と下位に位置する作品を 50 件ずつ評価者が目視で確認を行い、デザインに独創性があり品質が高いと考えられる作品、もしくは別カテゴリの方が適していると考えられる作品を稀観品として数えた。評価値の上位と下位に含まれる稀観品の割合を表 2 に示す。

表 2 から、評価値が上位の作品には稀観品と判断できる作品が 76% と多く存在していた。評価値の上位に位置付けられていた推薦による販売促進の効果が高いと考えられる作品の例を図 3 に示す。一方で、評価値が下位の作品にも稀観品が含まれていたが、評価値が上位の作品に比べて割

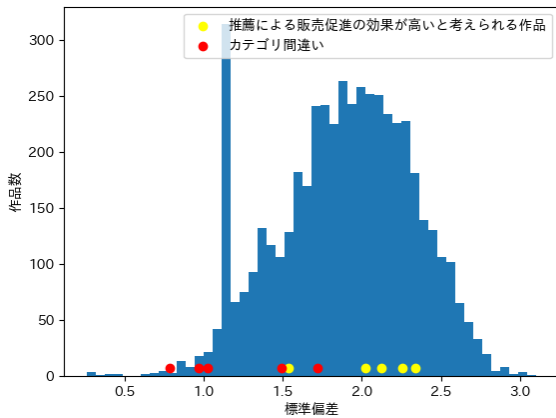


図 4 作品の標準偏差のヒストグラムと稀観品の例の分布

Fig. 4 Histogram of standard deviation of works and distribution of rare works

合は小さい。

この結果から、提案手法によって稀観品を評価値の上位に位置付けることができるため、運営者は評価値が上位となる作品を確認することで稀観品を効率よく発見することができると考えられる。しかし、作品の評価値が上位の作品には、推薦による販売促進の効果が高いと考えられる作品とカテゴリ間違いの作品が混在している。そのため、稀観品から推薦による販売促進の効果が高いと考えられる作品とカテゴリ間違いの作品を分類する指標が必要となる。この指標として、3.3 節で述べた作品タイトルに含まれる単語のスコアの標準偏差を利用する。この指標の効果の確認は 4.2 節で述べる。

4.2 推薦による販売促進の効果が高いと考えられる作品の検出

稀観品から推薦による販売促進の効果が高いと考えられる作品とカテゴリ間違いの作品を分類するために、3.3 節で述べたように、作品タイトルに含まれる単語のスコアの標準偏差に着目した。作品タイトルに含まれる単語のスコアの標準偏差によって稀観品のうち推薦による販売促進の効果が高いと考えられる作品とカテゴリ間違いの作品を分類できることを確認するために、データセットに対して、提案手法を適用し単語のスコアを求めた。その後、作品毎に作品タイトルに含まれる単語のスコアの標準偏差を計算した。標準偏差のヒストグラムと稀観品の例の分布を図 4 に示す。

図 4 から、推薦による販売促進の効果が高いと考えられる作品の殆どは、カテゴリ間違いの作品に比べ、標準偏差が高くなっており、推薦による販売促進の効果が高いと考えられる作品が分布する領域とカテゴリ間違いの作品が分布する領域が分離していることがわかる。このことから、作品タイトルに含まれる単語のスコアの標準偏差が高い作

品は、稀観品のうち推薦による販売促進の効果が高いと考えられる作品であり、これらの作品を検出することで、ハンドメイド作品を扱う CtoC の EC サイトの運営者はサイト上でこれらの作品を紹介し販売の促進を行うことができる。

一方で、標準偏差 1.5 付近には、カテゴリ間違いの作品に混じって推薦による販売促進の効果が高いと考えられる作品が 1 件分布している。この作品のタイトルは「コインもカードも入るケース ハチワレねこ (コインケース・カードケース)」であった。この作品タイトルでは、「ハチワレ」という単語のスコアのみが高くなっており、その他の単語のスコアは低くなっている。そのため、標準偏差はその他の推薦による販売促進の効果が高いと考えられる作品に比べ低くなり、カテゴリ間違いの作品と同程度の標準偏差となっている。このような作品は、作品タイトルに含まれる単語の標準偏差を利用しても分布する領域を分離できない。

5. まとめ

本報告では、稀観品を検出するために、特定のカテゴリにおいて作品タイトルに含まれる単語の出現頻度を尺度として、この尺度が相対的に低い作品を稀観品として検出する手法を提案した。実験から、提案手法によって求めた作品の評価値の上位に稀観品が位置付けられる傾向があることを確認した。また、作品タイトルに含まれる単語の出現頻度の標準偏差を求めることで、稀観品のうち推薦による販売促進の効果が高いと考えられる作品とカテゴリ間違いの作品が分布する領域が分離する傾向があることを確認した。今後の課題としては、作品の評価値の計算に標準偏差を組み込むことによって、稀観品のうち推薦による販売促進の効果が高いと考えられる作品の検出の精度改善が挙げられる。また、別カテゴリにおける稀観品の検出の実験や提案手法によって求めた稀観品のうち推薦による販売促進の効果が高いと考えられる作品を推薦に利用した場合の実環境での評価が挙げられる。

参考文献

- [1] 一般社団法人日本ホビー協会: ホビー白書 2017 年版 (2017).
- [2] 村上知子, 森紘一郎, 折原良平: 推薦の意外性向上のための手法とその評価, 人工知能学会論文誌, Vol. 24, No. 5, pp. 428–436 (2009).
- [3] Onuma, K., Tong, H. and Faloutsos, C.: TANGENT: a novel, 'Surprise me', recommendation algorithm, *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 657–666 (2009).
- [4] Zhang, Y. C., Séaghdha, D. Ó., Quercia, D. and Jambor, T.: Auralist: introducing serendipity into music recommendation, *Proceedings of the fifth ACM international conference on Web search and data mining*, pp. 13–22

- (2012).
- [5] Akiyama, T., Obara, K. and Tanizaki, M.: Proposal and Evaluation of Serendipitous Recommendation Method Using General Unexpectedness., *PRSAT RecSys*, pp. 3–10 (2010).
 - [6] 池尻恭介, 清雄一, 中川博之, 田原康之, 大須賀昭彦: 意外性のあるレシピを推薦するエージェントの提案, 電子情報通信学会論文誌 D, Vol. 98, No. 6, pp. 971–981 (2015).
 - [7] Jones, K. S.: A statistical interpretation of term specificity and its application in retrieval, *Journal of documentation* (1972).
 - [8] 村上知子, 森紘一郎, 折原良平: 推薦結果の意外性を評価する指標の提案, 人工知能学会全国大会論文集第 21 回全国大会 (2007), 一般社団法人人工知能学会, pp. 2C52–2C52 (2007).
 - [9] Kotkov, D., Wang, S. and Veijalainen, J.: A survey of serendipity in recommender systems, *Knowledge-Based Systems*, Vol. 111, pp. 180–192 (2016).
 - [10] 神 島 敏 弘 : 推 薦 シ ス テ ム の ア ル ゴ リ ズ ム , Toshihiro Kamishima (オンライン), 入手先 <http://www.kamishima.net/archive/recsysdoc.pdf> (参照 2020-04-07).
 - [11] 平澤真大, 小川祐樹, 諏訪博彦, 太田敏澄ほか: ニコニコ動画のログデータに基づくソーシャルノベルティのある動画の発見手法の提案, 情報処理学会論文誌, Vol. 54, No. 1, pp. 214–222 (2013).
 - [12] Taku, K.: MeCab: Yet Another Part-of-Speech and Morphological Analyzer, MeCab (online), available from <https://taku910.github.io/mecab/> (accessed 2020-04-07).
 - [13] Sato, T.: mecab-ipadic-NEologd: Neologism dictionary for MeCab, neologd (online), available from <https://github.com/neologd/mecab-ipadic-neologd> (accessed 2020-04-07).
 - [14] GMO ベバボ株式会社: minne, minne (online), available from <https://minne.com/> (accessed 2020-04-07).