

レビューに基づく複数ジャンルを対象とした類似作品の推薦と評価

Recommendation and evaluation of similar works in multiple Genres based on reviews

福本 虎太郎[†] 伊藤 淳子[‡] 吉野 孝[‡]
Kotaro Fukumoto Junko Itou Takashi Yoshino

1. はじめに

求める商品やサービス、情報などを探す際、推薦システムが広く利用されている。中でも、ユーザのプロファイルを利用する協調フィルタリングという手法が既存サービスでは多く用いられている。しかし、この協調フィルタリングという手法を小説や映画、ゲーム等を対象とする作品推薦に応用した場合、ユーザにとって自明な作品が推薦される等、いくつかの問題が発生する [1]。

これに対し、レビューや Wikipedia から作品の内容を抽出し、得られた特徴量を推薦に利用する研究が行われている [1][2][3]。これらの既存研究は単一ジャンルを対象としている。一方で、推薦システムの利用目的には、単一ジャンルだけでなく他ジャンルへの関心の拡大が含まれるとされている [4]。したがって、ジャンルを横断する類似作品の推薦を実現することにより、ユーザの満足度が向上することが考えられる。

そこで、我々はレビューから TF-IDF を用いて作品の特徴語を抽出し、その特徴語の共起関係を用いてジャンルを横断する作品推薦手法を提案した [5]。単語の珍しさを表す IDF に重みを付けた類似度計算手法を導入することにより、出力される作品のジャンルの多様性が向上した。一方で、IDF のみを考慮したため、特定のジャンルでのみ多く用いられる単語を適切に抽出できないという課題が明らかになった。この課題に対しては、単語が特定のジャンルで多く用いられる傾向の強さを表す「ジャンル依存度」という指標を定義し、類似度計算に導入して、特定のジャンルにおける影響の低減を目指した。この結果、従来手法と比較して、出力される類似作品のジャンルの偏りを減少させることができた。しかしながら、固有名詞の共起による同一シリーズ作品の出力や、多くの作品で用いられる一般的な単語の共起による内容の類似性が低い作品の出力が見られた。本稿では、ジャンル依存度を取り入れた類似度計算手法におけるこれらの課題について検討する。

2. 関連研究

2.1 概要と読者レビューを用いた特徴語に関する研究

既存の商用サイト等における漫画の推薦では、協調フィルタリングの手法が用いられている。これは、他ユーザの購入履歴や購入傾向等をもとに推薦を行う手法であり、ユーザにとって自明な作品が推薦されやすいなどの問題が存在する。

白石らはこの問題を解決するため、概要と読者レビューの両方から概要キーワードと評判キーワードの 2 種の特徴

語を抽出し、作品の特徴を直感的に理解可能とする手法を提案した [1]。しかし、Wikipedia は記事によって記述量が大きく異なることから、作品によっては適切に特徴語を抽出できない可能性がある。

2.2 レビュー間の類似性と作品間の類似度に関する研究

櫛見らは、レビューの持つ多様な情報に着目し、レビュー間の類似度を作品間の類似度とする手法を提案している [2]。視聴する映画を選択する際にユーザが参考にする情報を事前調査した上で、ユーザの好む映画と同様の特徴を持つ映画を推薦している。

櫛見らの手法では、全体的観点と部分的観点に分けて類似度を算出し、それぞれを全体類似度、部分類似度と定義した。そして、全体類似度と部分類似度の比率を調整しながら、ユーザの求める特徴を含んだ映画の提示を試みた。その結果、映画のジャンルや出演俳優、緊迫感、恋愛要素などに応じた映画が推薦された。しかし、対象は映画のみであり、複数のジャンルを対象とした場合に同様の手法で同程度の推薦が可能であるかは不明である。

2.3 特徴語の共起関係に関する研究

漫画は、イラストやキャラクターなどの非言語情報が含まれるコンテンツである。そのため、ユーザは漫画の内容を重視した情報アクセスが困難となる。この問題を解決するため、山下らはレビューから抽出した特徴語を用いることにより、漫画の内容を重視した情報アクセスを支援するシステムを提案している [3]。

この研究において、山下らは特徴語の共起関係を作品間の関連と定義した。この共起関係とは、2 つの作品間において特徴語が共通して存在するか否かの 2 値の関係である。特徴語の共起関係を作品間の関連と定義することにより、ユーザが作品間の共通点を把握しながら、探索的に漫画にアクセスすることを可能とした。一方で、この研究では、特徴語の共起関係を 2 値の関係として扱っているため、特徴語間の重要度の違いを反映できていない。

3. ジャンルの偏りを減少させる推薦手法

3.1 方針

白石らの研究 [1] において指摘された問題点から、本研究は Wikipedia ではなく、レビューから作品の特徴量を抽出する。作品間の類似度については、櫛見らの研究 [2] と同じくレビュー間の類似度を作品間の類似度として用いる。類似度の計算には、山下らの研究 [3] と同じく特徴語の共起関係を用いる。また、本研究の対象ジャンルは「小説・漫画・映画・ドラマ・ゲーム・アニメ」の 6 ジャンルとする。

[†] 和歌山大学大学院システム工学研究科, Graduate School of Systems Engineering, Wakayama University

[‡] 和歌山大学システム工学部, Faculty of Systems Engineering, Wakayama University

3.2 特徴語の抽出

本研究では、解析対象として「作品データベース¹⁾」に投稿されたレビューを利用する。この Web サイトから取得したレビューを、作品の内容を表すテキストデータとして使用する。そのテキストデータを対象に Janome²⁾を用いて形態素解析を行う。辞書には mecab-ipadic-NEologd³⁾を利用する。この際、名詞と形容詞は作品の特徴を表すと考え、それ以外の単語を除去する。また、名詞の中でも代名詞や人名は除去し、単一作品のレビューでのみ用いられる単語についても、他作品との関連に用いることができないため、除去する。この処理によって得られた、単語ごとに分割されたレビュー文章を対象に TF-IDF を用いて各作品の特徴語を抽出する。文書 $d(a)$ の単語 w の TF-IDF を求める計算式を式 (1) に示す。文書 $d(a)$ とは、作品 a のレビュー全文を指す。また、単語の出現頻度を表す TF の計算式を式 (2) に、単語の逆文書頻度を表す IDF を式 (3) に示す。

$$tfidf(w, d(a)) = tf(w, d(a))idf(w) \quad (1)$$

$$tf(w, d(a)) = \frac{\text{単語 } w \text{ の出現回数}}{\text{文書 } d(a) \text{ に含まれる全単語数}} \quad (2)$$

$$idf(w) = \log\left(\frac{\text{全ジャンルの全作品数} + 1}{\text{単語 } w \text{ が出現する作品数} + 1}\right) + 1 \quad (3)$$

求められた TF-IDF の値が大きい順に上位 50 単語をその作品の特徴語と定義する。特徴語が 50 語に満たなかった作品は、作品の特徴を表すための文章量に達していないと判断し、対象から除外した。本研究が対象とした作品数は各ジャンルごとに以下の通りである。

- 小説：3249 作品
- 漫画：6407 作品
- 映画：6364 作品
- ドラマ：2312 作品
- ゲーム：4939 作品
- アニメ：5802 作品

3.3 出力されるジャンル割合の計算方法

推薦システムにおいては、関連の度合いや推薦の精度だけでなく、推薦されたアイテムの多様性を向上させることが、ユーザの満足度の向上につながるとされている [6][7][8]。そこで、本研究では出力される類似作品のジャンルの多様性や偏りについて調査する。具体的な計算手順は次の通りである。手法 X には、以降に提案する各類似度計算手法が入るものとする。

全ジャンルの全作品を対象に、手法 X によって得られる類似度が高い順に、10 作品ずつ抽出する。この際、類似度の重複により 10 作品を超過する作品が一部発生するが、本稿では超過した作品数をそのまま用いる。抽出された類似作品群を、入力作品のジャンルごとに分ける。入力作品のジャンルごとに分けられた類似作品群のジャンルの割合を、出力される類似作品のジャンルの割合とする。

¹⁾<https://sakuhibdb.com/>(2022 年 7 月 14 日確認)

²⁾<https://mocobeta.github.io/janome/>(2022 年 7 月 14 日確認)

³⁾<https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md>(2022 年 7 月 14 日確認)

表 1: 従来手法による出力ジャンルの割合

出力	入力					
	小説	漫画	映画	ドラマ	ゲーム	アニメ
小説	39.1	4.86	4.87	5.29	0.452	3.02
漫画	15.9	58.8	8.04	11.6	1.80	11.4
映画	16.1	7.25	64.5	17.3	1.49	6.04
ドラマ	5.98	3.67	6.54	51.5	0.300	1.74
ゲーム	4.14	3.74	3.00	1.74	91.7	4.05
アニメ	18.8	21.7	13.0	12.6	4.31	73.7

※単位は割合 (%) である。

3.4 対象ジャンル数を考慮しない手法

3.4.1 手法の概要

複数ジャンルのコンテンツを対象としたクロスドメイン推薦に関する研究 [9][10] は行われているが、出力される類似作品のジャンルの多様性や偏りに着目した研究は行われていない。したがって、既存研究で用いられている手法を、複数ジャンルを横断する推薦システムに応用した際に得られる結果について検証する必要がある。そこで、既存研究で用いられている類似度計算手法を代表する手法として、共起関係にある特徴語の TF-IDF の合計を作品間の類似度とする手法を採用する。この類似度計算手法は、対象のジャンル数を考慮しない手法であるため、既存研究の手法を代表する類似度計算手法として適切であると考えられる。

入力した作品 A のレビュー全文を文書 $D(A)$ とし、共起関係にある特徴語を W とする。また、入力作品 A と出力作品 B の類似度を $sim_{A,B}$ とする。その際の、計算式を式 (4) に示す。TF-IDF を求める式については、式 (1) を参照する。式 (4) によって求められる値を作品間の類似度とする類似度計算手法を以降、従来手法と定義する。

$$sim_{A,B} = \sum tfidf(W, D(A)) \quad (4)$$

3.4.2 出力される類似作品のジャンルの割合

3.4.1 項で定義した従来手法によって出力された類似作品のジャンルの割合を表 1 に示す。計算方法については、3.3 節を参照する。有効数字は 3 桁とし、表の横軸が入力した作品のジャンル、縦軸が出力された作品のジャンルを示している。

出力されるジャンルの割合が最も高いのは、いずれのジャンルにおいても、入力ジャンルと同一のジャンルであることが、表 1 から分かる。このことから、従来手法によって出力される作品のジャンルは、入力した作品と同一のジャンルに偏ることが明らかになった。

3.4.3 課題点と改善点

3.4.2 項で明らかになった問題の原因は IDF の計算方法であると考えられる。式 (3) で定義した IDF の計算式は全ジャンルの作品における単語の珍しさを表している。したがって、特定のジャンルでのみ多く用いられる単語であっても、全ジャンルの作品の中では珍しい単語として、IDF が高く算出される。この性質は、特定のジャンルでのみ多く用いられる特徴語が共起関係にあるとき、作品間の類似度が高く算出されることを示している。この問題を解決す

表 2: ジャンル依存度の高い単語上位 5 個

小説	漫画	映画	ゲーム	アニメ
小説	漫画	映画	play	アニメ
文体	連載	俳優	ゲーム	声優
文章	作者	gyao	難易度	作画
ミステリ	画力	吹き替え	プレイヤー	op
著者	単行	監督	クリア	ed

るためには、特定のジャンルでのみ多く用いられる単語への対策が必要となる。

3.5 指標の提案

3.4 節で明らかになった問題を解決するために、各単語が特定のジャンルにおいて用いられる傾向の強さを表す指標「ジャンル依存度」を新たに定義する。ジャンル依存度とは、全ジャンルにおける単語の珍しさと特定のジャンルにおける単語の珍しさの差異を利用して算出する指標である。全ジャンルにおける単語の珍しさは式 (3) によって求められる。この指標を以降、全ジャンル IDF と定義する。また、特定のジャンルにおける単語の珍しさは特定のジャンルにおける単語の IDF の値によって求めることが出来る。ジャンル g における単語 w の珍しさを式 (5) に示す。この指標を以降、ジャンル別 IDF と定義する。そして、ジャンル g における単語 w のジャンル依存度の計算式を式 (6) に示す。

式 (6) に示す定義により、ジャンル依存度は特定のジャンルにおける IDF の値が低い一方で、全ジャンルにおける IDF の値が高いような単語の場合に大きい値をとる指標となる。したがって、特定のジャンルでのみ多く用いられる傾向が強い単語ほど、大きい値をとる指標である。

ジャンル別 IDF(w, g)

$$= \log\left(\frac{\text{ジャンル } g \text{ の全作品数} + 1}{\text{単語 } w \text{ が出現するジャンル } g \text{ の作品数} + 1}\right) + 1 \quad (5)$$

$$\text{ジャンル依存度}(w, g) = \frac{\text{全ジャンル IDF}(w)}{\text{ジャンル別 IDF}(w, g)} \quad (6)$$

一部のジャンルにおけるジャンル依存度の高い単語上位 5 個を表 2 に示す。小説ジャンルにおいては「小説」「文体」「文章」等の単語が、漫画ジャンルにおいては「漫画」「連載」「作者」等の単語が示されている。この傾向は、他のジャンルにおいても確認される。したがって、ジャンル依存度は各単語が特定のジャンルにおいて用いられる傾向の強さを表す指標として適切であると考えられる。

3.6 ジャンル依存度を用いた類似度計算手法

3.6.1 手法の概要

特定のジャンルでのみ多く用いられる単語への対策を行った類似度計算手法を提案する。この類似度計算手法では、各ジャンルにおいてジャンル依存度が上位 1% である単語を「ジャンル依存ワード」と定義し、類似度計算から除外する。具体的な計算式を式 (7) に示す。入力した作品 A のレビュー全文を文書 $D(A)$ 、共起関係にある特徴語を W 、ジャンル依存ワードを $Gword$ とする。式 (7) に示すように、共

表 3: 提案手法による出力ジャンルの割合

出力	入力					
	小説	漫画	映画	ドラマ	ゲーム	アニメ
小説	27.2	6.47	7.67	8.20	3.57	8.19
漫画	19.0	36.8	13.0	17.5	10.8	23.7
映画	19.1	11.1	47.9	22.1	10.4	15.1
ドラマ	7.21	5.39	8.38	27.8	1.86	4.87
ゲーム	5.01	5.79	4.86	3.18	51.9	12.4
アニメ	22.5	34.4	18.2	21.2	21.4	35.7

※単位は割合 (%) である。

起関係にある特徴語 W がジャンル依存ワードでない場合のみ類似度の加算を行う。

$$sim_{A,B} = \sum_{W \notin Gword} tfidf(W, D(A)) \quad (7)$$

3.6.2 出力される類似作品のジャンルの割合

提案手法によって出力された、類似作品のジャンルの割合を表 3 に示す。入力ジャンルと同一ジャンルの作品が出力される割合は、従来手法と比較して低下した。このことから、ジャンル依存ワードの除去は出力されるジャンルの多様性を向上させると考えられる。

4. 比較実験

4.1 実験概要

出力される作品のジャンルの多様性向上がユーザの評価に与える影響について検証するため比較実験を行う。この比較実験においては、従来手法と提案手法によって得られる類似作品 10 作品に対して、それぞれ被験者が評価する。実験の被験者は 20 代から 30 代の男女 15 名である。実験期間は 2021 年 12 月 17 日から 2022 年 1 月 5 日までの 18 日間である。全被験者が評価した作品数は、延べ 520 作品である。

比較実験には、実験者の作成した比較実験用システムを用いる。評価項目には、内容の類似性や意外性 [11]、セレンディピティ [12] などの指標を用いる。セレンディピティとは、Ge らが提唱した推薦システムの評価指標であり、本研究ではセレンディピティな作品の評価を「未知の作品」「予期できない作品」「評価者が面白そうだった作品」の 3 つの観点から行う。被験者はそれぞれの評価項目に対応するアンケートに回答する。各アンケート項目には 5 段階のリッカート尺度を用い「未知の作品」に対応するアンケート項目のみ「名前も内容も知らない」「名前は知っているが内容は知らない」「名前も内容も知っている」の 3 段階を用いた。

4.2 実験結果と考察

アンケート結果の一部を表 4 に示す。有意水準は 5% とし、有意差検定には Wilcoxon の符号順位和検定を用いた。意外性に対応するアンケート結果では、従来手法の最頻値が 5、提案手法の最頻値は 1 となり、有意差もみられた。この項目は、推薦されることが予期できた作品には 5、予期できなかった作品には 1 と評価するため、値が小さいほど

表 4: 比較実験の結果

項目	手法	評価値					中央値	最頻値
		1	2	3	4	5		
意外性	従来手法	139	48	38	107	188	4	5
	提案手法	185	53	21	78	183	4	1
類似性	従来手法	36	66	74	149	195	4	5
	提案手法	46	62	87	123	202	4	5

評価値 1: 全く当てはまらない, 2: やや当てはまる, 3: どちらでもない, 4: やや当てはまる, 5: 非常に当てはまる

表 5: セレンディピティな作品数

手法	作品数	百分率	有意差
従来手法	68/520	13.0%	p=0.00202
提案手法	101/520	19.4%	

意外性が高い作品となる。したがって、従来手法より提案手法の方が意外性の高い作品を推薦できたといえる。

内容の類似性に対応するアンケート結果では、従来手法と提案手法のいずれもが中央値 4, 最頻値 5 となった。また、二つの手法の間に有意差はみられなかった。したがって、従来手法と提案手法によって得られる類似作品の内容の類似性は同程度である可能性がある。

次に、セレンディピティについて考察する。本稿では、セレンディピティな作品の定義を以下のとおりとする。「未知の作品」に対応するアンケート項目の評価が「名前も内容も知らない」もしくは「名前は知っているが内容は知らない」であること、かつ「評価者が面白そうと思った作品」に対応するアンケート項目の評価が 4 もしくは 5 であること、かつ「予期できない作品」に対応するアンケート項目の評価が 1 もしくは 2 であること、という 3 つの条件を満たした作品とする。セレンディピティな作品の割合について表 5 に示す。有意水準は 5% とし、有意差検定にはカイ二乗検定を用いた。セレンディピティな作品の割合は従来手法よりも提案手法の方が高く、有意差もみられた。したがって、従来手法よりも提案手法の方が推薦システムとして高く評価できると考えられる。

5. 課題と対策

5.1 対策の提案

比較実験の結果、出力されるジャンルの多様性向上がユーザの満足度向上に影響を与えることが明らかになった。一方で、固有名詞や一般的に用いられる単語が特徴語として抽出される影響で、同一シリーズ作品の推薦や内容の類似性が低い作品の推薦がみられた。これらの問題への対策について検討する。

本稿では固有名詞の共起による作品間の繋がりも有用であると考えていたが、実際は同一シリーズ作品の推薦に影響を与える例が多くみられた。したがって、固有名詞を除去することにより、この問題の解決を図る。この対策を対策 A と定義する。なお、固有名詞の特定には 3.2 節で述べた NEologd 辞書を使用する。また、一般的に用いられる単語の影響で内容の類似性が低い作品が推薦される問題への対策には、TF-IDF の重みの改良が考えられる。TF-IDF には複数の定義が存在する [13]。このうち、TF の重みが下が

表 6: 各種対策後の特徴語

対策	特徴語
対策なし	キノ, エルメス, 小説, 青い空, あとがき, 旅人, 作品, 好き, 面白い, キノの旅
対策 A	小説, あとがき, 旅人, 作品, 好き, 面白い, 城壁, 短編, 人間, 良い
対策 A + 対策 B	旅人, 城壁, 全的, あとがき, アウトサイダー, コロシウム, 黒星, ティー, 国々, 短編

る定義を採用することにより、レビュー文中に複数回使用されている抽象的な単語の影響を抑えることができると考えられる。この TF の定義を用いた際の計算式を式 (8) に示す。この対策を対策 B と定義する。

$$tf(w, d(a)) = \log(N_{d(a)} + 1) \quad (8)$$

5.2 対策後の特徴語に関する考察

5.1 節で提案した各対策の効果について、小説「キノの旅」[14]における特徴語を例に考察する。対策なしの状態及び各対策後の特徴語上位 10 語を表 6 に示す。対策なしの状態では「キノ」「エルメス」などの固有名詞の他「作品」「面白い」などの一般的に用いられ、作品の内容が反映されていない特徴語が含まれている。対策 A を実施すると「キノ」「エルメス」などの固有名詞は除去された。このことから、固有名詞の除去という対策は有効に機能していると考えられる。

次に、対策 A と対策 B の両方の対策後における特徴語について考察する。これらの対策を実施することにより「作品」「面白い」などの一般的に用いられる特徴語の除去に成功している。したがって、TF-IDF の定義の改良は一般的に用いられる単語の除去に有効であると考えられる。一方で「黒星」「ティー」などの固有名詞がみられる。これは、NEologd 辞書に固有名詞として登録されていない単語であることから一般名詞と誤って認識されたために発生したと考えられる。この問題を解決するためには、固有名詞がより多く登録されている辞書の導入や固有名詞を機械的に判定する機能の実装などの対策が必要である。

5.3 形態素解析器と辞書の変更による特徴語の変化

5.2 節で述べた問題の解決を図るため、形態素解析器と辞書の変更が特徴語に与える変化について調査する。具体

表 7: Sudachi を用いた場合の特徴語

対策	特徴語
対策 A + 対策 B	キノ, 旅人, 城壁, あとがき, アウトサイダー, 青い, コロシウム, 黒星, 傍観者, ファルス

的には Sudachi[15] という形態素解析器と Sudachi 辞書を導入する。Sudachi の特徴として、表記ゆれの考慮や形態素の分割用パラメータを 3 種類から選択できる点などが挙げられる。特に、形態素の分割用パラメータを「C 単位」とすることにより固有表現の抽出が可能となる。また、Sudachi 辞書は UniDic¹ と NEologd 辞書を組み合わせた辞書である。Sudachi 辞書の作成時には、手動の調整が行われるため NEologd 辞書よりも高い精度で固有名詞を識別することが期待される。

Sudachi の実装には SudachiPy² を利用した。また、形態素の分割パラメータは「C 単位」とし辞書のタイプは「Full」とした。Sudachi と Sudachi 辞書への変更による影響を検証するため、5.2 節と同様に小説「キノの旅」における特徴語を例に考察する。5.2 節で述べた対策 A と対策 B を適用した状態において、Sudachi と Sudachi 辞書を用いた場合の特徴語上位 10 語を表 7 に示す。5.2 節で問題とした一般名詞と識別されてしまう固有名詞については「黒星」「ファルス」などが見られることから、問題は解決されていないと考えられる。また、5.2 節では除去されていた「キノ」という固有名詞が Sudachi を利用した場合では除去されていない。このことから、Sudachi 及び Sudachi 辞書を利用した際には、必ずしも固有名詞の識別能力が向上するとは限らないことが明らかになった。一方で、小説「キノの旅」を用いた検証は局所的な例であるため、対象データ全体でみれば固有名詞の識別能力が向上している可能性がある。したがって、今後は Sudachi 及び Sudachi 辞書の導入による対象データ全体への効果を検証する必要がある。

6. おわりに

本稿では、複数ジャンルを横断した類似作品の推薦において特定のジャンルの作品が多く推薦されるという問題を解決するため、出力結果が特定のジャンルに偏らない類似作品推薦手法を提案した。具体的には、単語が特定のジャンルで用いられる傾向の強さを表すジャンル依存度という指標を提案し、類似度計算手法に導入した。結果として、従来手法よりも出力される作品のジャンルの多様性を向上させることができた。また、出力される作品のジャンルの多様性向上がユーザの評価に与える影響について検証するため比較実験を行った。その結果、従来手法よりも提案手法の方が推薦システムとして高く評価できることが明らかになった。

一方で、固有名詞による同一シリーズ作品の推薦や、一般的に用いられる単語による類似性の低い作品の推薦がみられた。これらの問題に対して、固有名詞の除去と TF-IDF の定義の変更の二つの対策を実施した。結果として、この

¹<https://clrd.ninjal.ac.jp/unidic/>(2022 年 7 月 14 日時点)

²<https://github.com/WorksApplications/SudachiPy>(2022 年 7 月 14 日時点)

二つの対策は有効であったが誤って一般名詞と識別されてしまう固有名詞の重要度が高くなるという新たな問題が明らかになった。今後はこの問題に対する改善が必要である。

参考文献

- [1] 白石絵里奈, 田村亮介, 浅沼爽汰, 白井聡一, 藤田和成, 町田翔, 延澤志保: 概要と読者レビューに基づく漫画特徴抽出, 情報処理学会第 81 回全国大会講演論文集, Vol.2019, No.1, pp.443-444 (2019).
- [2] 櫛見圭司, 北山大輔: ユーザレビューを用いた全体的・部分的観点の類似に基づく映画推薦, 第 9 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2017), No.B5-4, pp.1-6 (2017).
- [3] 山下諒, 朴炳宣, 松下光範: コミックの内容情報に基づいた探索的な情報アクセスの支援, 人工知能学会論文誌, Vol.32, No.1, pp.WII-D 1-11 (2017).
- [4] Swearingen, K. and Sinha, R.: Beyond algorithms: An HCI perspective on recommender systems, ACM SIGIR Workshop on Recommender Systems (2001).
- [5] 福本虎太郎, 伊藤淳子: レビューに基づく複数ジャンルを対象とした類似作品推定手法の検討, 情報処理学会関西支部支部大会講演論文集, Vol.2022, No.G-32, pp.1-4 (2022).
- [6] Ziegler, C., McNee, S.M., Konstan, J.A. and Lausen, G.: Improving recommendation lists through topic diversification, Proceedings of the 14th international conference on World Wide Web, pp.22-32 (2005).
- [7] 内田 匠, 吉田 健一: Cold-Start 問題と多様性を考慮したレコメンデーションの総合評価, 情報処理学会論文誌, Vol.62, No.12, pp.2108-2118 (2021).
- [8] Zhang, M. and Hurley, N.: Avoiding Monotony: Improving the Diversity of Recommendation Lists, Proceedings of the 2nd ACM conference on Recommender systems (RecSys 2008), pp.123-130 (2008).
- [9] 富士谷康, 村尾和哉, 望月祐洋, 西尾信彦: コンテンツの多様性を考慮したクロスドメイン推薦, 情報処理学会論文誌, Vol.57, No.10, pp.2210-2221 (2016).
- [10] 中本昌吾, 宮治裕: 自然言語処理を用いたコンテンツ作品のクロスドメイン推薦, 情報処理学会第 81 回全国大会講演論文集, Vol.2019, No.1, pp.441-442 (2019).
- [11] Adamopoulos, P. and Tuzhilin, A.: On Unexpectedness in Recommender Systems: Or How to Expect the Unexpected, Proceedings of the Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011), at the 5th ACM International Conference on Recommender Systems (RecSys 2011), No.DiveRS, pp.11-18 (2011).
- [12] Ge, M., Delgado-Battenfeld, C. and Jannach, D.: Beyond accuracy: evaluating recommender systems by coverage and serendipity, Proceedings of the 4th ACM conference on Recommender systems (RecSys 2010), pp.257-260 (2010).
- [13] Manning, C., Raghavan, P. and Schütze, H.: Introduction to Information Retrieval, Cambridge, Cambridge University Press (2008).

- [14] 時雨沢恵一：キノの旅 -the Beautiful World-, 電撃文庫 (2000).
- [15] Takaoka, K., Hisamoto, S., Kawa-hara, N., Sakamoto, M., Uchida, Y., and Mat-sumoto, Y.: Sudachi: a Japanese Tokenizer for Business, Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan. European Language Resources Association (ELRA) (2018).