

ジャンルを横断した類似作品ネットワークを用いた セレンディピティ指向推薦システムの提案

A Proposal for Serendipity-oriented Recommendation System using a Network of Similar Works across Multiple Genres

福本 虎太郎¹⁾ 伊藤 淳子¹⁾ 吉野 孝¹⁾

Kotaro Fukumoto Junko Itou Takashi Yoshino

1 はじめに

求める商品やサービスを探す際、協調フィルタリングというユーザの行動履歴を利用する手法を用いた推薦手法が多く利用されている。しかし、小説や映画等を対象とする作品推薦においては、ユーザにとって自明な作品が推薦される等の問題が存在する [1]。これらの問題を解決するため、レビューや Wikipedia 等から作品の内容を抽出し、推薦に利用する研究が行われている [1][2][3]。しかし、これらの既存研究はいずれも単一のジャンルのみを対象としている。推薦システムの利用目的には複数ジャンルへの興味の拡大が含まれる [4] ことから、複数ジャンルを対象とする推薦システムが求められる。興味の拡大という観点においては、ユーザが予期していなかった未知の有益な情報という、意外性や偶発性などの複数の要素が関連した指標であるセレンディピティ [5] が着目されている [6][7]。

そこで、我々はセレンディピティ向上を目的とした複数ジャンルを横断する推薦システムを提案した [8]。推薦対象を複数ジャンルへと単純に拡大すると、推薦結果が特定のジャンルに偏るといった問題が生じる。そのため、我々は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「ジャンル依存度」という指標を定義した。この指標を用いた推薦手法は、ジャンルの偏りを軽減することができた。また、従来手法と比較して意外性やセレンディピティの項目が有意に向上した。

我々の提案した推薦システムは意外性を向上させつつ、入力した作品と類似する作品をユーザに提示できる。一方で、偶発性を提供する機能が不足していた。また、一方的に作品を推薦されるシステムの利用はユーザにとって受動的な体験であると考えられる。

そこで、本研究では偶発的な作品の発見と能動的な作品探索を支援するセレンディピティ指向推薦システムを提案する。具体的には、類似作品ネットワークを構築し、任意の2作品間の最短経路上に存在する作品をユーザに推薦する。提案手法によって推薦される作品は入力した2作品と間接的に類似する作品であることから、ユーザに偶発的な作品発見を促し、推薦結果の意外性を高めることが出来ると考えられる。また、推薦結果を予測して作品を入力するため、推薦結果に対する興味関心を促進し、ユーザの能動的な作品探索を実現できると考えられる。

2 関連研究

2.1 クロスドメイン推薦に関する研究

特性の異なる複数の領域の情報を利用する推薦はクロスドメイン推薦と呼ばれる。クロスドメイン推薦を行う目的としては、コールドスタート問題の解決や推薦精度の向上、新規アイテムへの対応などが挙げられる [9]。

また、映画や書籍、テレビ番組や音楽などを対象とするクロスドメイン推薦の研究も行われている [10][11]。

富士谷らは、テレビ番組をソースドメイン、書籍をターゲットドメインと定義したクロスドメイン推薦を行っている [10]。この研究では、単語の重要度を表す TF-IDF とトピックモデルの一つである LDA の併用により、有効な特徴量が番組ごとに異なる点を考慮した。一方で、テレビ番組から抽出した特徴量を用いて書籍を推薦するというこの研究は、複数ドメインの作品を相互に推薦するという点で本研究と異なっている。

中本らは「映画・漫画・アニメ・小説・楽曲・ドラマ・ゲーム」を対象としたクロスドメイン推薦を行っている [11]。その際、作品の内容を表すテキストデータは Wikipedia から取得し、Doc2Vec を用いて分散表現を獲得している。その分散表現を用いて作品間のコサイン類似度を算出し、類似作品を予測した。一方で、この手法では推薦される作品の新規性と妥当性は保証されているが、セレンディピティの観点で課題がある。

2.2 セレンディピティ指向推薦に関する研究

推薦システムに関する研究の多くでは評価指標として推薦精度が用いられている [12]。一方、推薦精度とは異なる評価指標の必要性が言及されている [13]。その一つがセレンディピティである [5]。ユーザとシステムのインタラクションや、情報の可視化などを用いて推薦のセレンディピティを向上させる研究が行われている。

奥らは「偶然と察知による思いがけない発見をする能力」というセレンディピティの原義に着目し、偶発性を促進するユーザインタフェースを持つフュージョンベース推薦システムを提案している [7]。任意の2つの書籍を選択し混ぜ合わせるにより、それぞれの書籍が持つ属性データの両方を考慮した書籍がユーザに推薦される。そして、ユーザはこの操作を繰り返し実行することができる。そのため、このシステムはユーザが探索的にセレンディピティな書籍を発見することを支援している。一方で、利用している属性データが書籍タイトル、著者名、カテゴリ名に限定されており、書籍の内容を十分に考慮できていない可能性がある。

Thudt らは情報の可視化によりセレンディピティな発見を支援する複数の機能を持つ書籍推薦システム「Bohemian Bookshelf」を提案している [14]。機能の一つである「キーワード・チェーン」は、キーワードに基づく書籍間の関係を可視化し、中央の書籍と関連するキーワードが含まれる書籍を円形状にユーザに提示する。これにより、特定の用語のみを用いた検索とは異なり、より興味の拡大を促進することが可能となる。一方で、対象ドメインの数や関連するキーワードを複合的に用いるという点で本研究と異なっている。

1) 和歌山大学

3 提案手法

3.1 方針

白石らの研究 [1] において指摘された問題点を考慮して、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。文書間の類似度計算には分散表現が用いられる場合が多いが、あくまで数字上の類似性であり、ユーザはその情報から推薦理由を推測することは困難である。したがって、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、本研究が対象とするジャンルは「小説・漫画・映画・ドラマ・ゲーム・アニメ」の 6 ジャンルとする。

3.2 特徴語の抽出

本研究では「作品データベース¹⁾」及び「レビューン²⁾」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。初めに、このテキストデータを対象に Sudachi³⁾ を用いて形態素解析を行う。分割モードは C を指定し、辞書には Sudachi が提供する辞書の Full を使用する。形態素解析の際、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の他、日本語ストップワードとして一般的に用いられている「Slothlib⁴⁾」に含まれる単語は除去する。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 を用いて各単語の重要度を算出する。そして、TF-IDF の値が大きい順に上位 50 語を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：4041 作品
- 漫画：7207 作品
- 映画：7713 作品
- ドラマ：2690 作品
- ゲーム：5231 作品
- アニメ：6597 作品

3.3 TF-IDF の概要

TF-IDF とは単語の重要度を表す指標の 1 つであり、TF(Term Frequency) と IDF(Inverse Document Frequency) の積により求める。TF-IDF を求める計算式を式 (1) に示す。TF-IDF の値は最後に L2 正規化を行う。

TF とは、単語の出現頻度を表す指標であり、 $tf(w, d)$ は文書 d における単語 w の出現頻度を表す。本研究における文書 d は、作品 a のレビュー全文のことを指す。したがって、本研究における TF は、作品 a のレビューにおいて、単語 w が出現する頻度を表す。以降、作品 a のレビュー全文を指す文書 d を $d(a)$ と表記する。文書 $d(a)$ における、単語 w の TF を求める計算式を式 (2) に示す。 $n_{w,d(a)}$ は文書 $d(a)$ において単語 w が出現する回数を表す。対数を用いた TF の定義は文章中に複数回

用いられる単語の影響度を下げる効果がある [15]。レビューにおいて頻出する単語として「良い」や「主人公」などが考えられる。これらの単語は作品の内容を表すのに不適切であると考えられるため、本研究では対数を用いた定義を採用した。

IDF とは、単語の珍しさを表す指標であり、 $idf(w)$ は単語 w が出現する文書の頻度の逆数を表す。単語 w の IDF を求める計算式を式 (3) に示す。 M は全ジャンルの全作品の数、 $df(w)$ はレビュー中に単語 w が出現する作品の数を表す。また、分母が 0 になることを防ぐため、分母と分子に 1 を加算し、全文書に登場する単語の IDF が 0 になることを防ぐため、最後に 1 を加算する。

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

$$tf(w, d(a)) = \log(n_{w,d(a)} + 1) \quad (2)$$

$$idf(w) = \log\left(\frac{M+1}{df(w)+1}\right) + 1 \quad (3)$$

3.4 提案手法の概要

我々が提案した複数ジャンルを横断する推薦手法 [8] は意外性やセレンディピティの向上に成功した。しかし、入力した作品と直接類似する作品のみをユーザに提示するという方法は意外性を低下させる要因になり得る [16] と指摘されている。また、偶発性を提供する機能も不足していた。偶発性の促進はセレンディピティの向上に寄与すること [7] が述べられている。加えて、一方的に作品を推薦されるシステムの利用はユーザにとって受動的な体験であると考えられる。ユーザの能動的な働きかけは意外な発見につながることを示唆されている [17]。

これらの問題を解決するため、本研究では複数ジャンルを横断する類似作品ネットワークを用いたセレンディピティ指向推薦システムを提案する。類似作品ネットワークとは、作品をノード、類似関係をエッジとして構築されたグラフ構造である。この提案手法では、ユーザが任意の 2 作品を入力すると、その間の最短経路がダイクストラ法によって算出される。そして、その最短経路上に存在する作品をユーザに提示する。

提案手法によって推薦される作品は、入力された 2 作品と間接的に関連する作品であると言える。したがって、この機能は偶発的な作品の発見を支援し、推薦結果の意外性を向上させると考えられる。また、ユーザはシステムを利用する際、推薦結果を予想して作品を入力するため、ユーザが自らの意思によって新たな作品を発見したというユーザ体験につながると考えられる。

一方で、間接的な関連による推薦というユーザにとって推薦された理由が不明瞭になるという課題が存在する。推薦理由の把握とユーザの推薦結果への満足度には関連がある事が指摘されている [18]。そこで、提案手法では各作品間において共起関係にある特徴語をユーザに提示する。これにより、各作品間における関連性がユーザに提示されるため、推薦理由の把握が可能になると考えられる。

また、セレンディピティ指向推薦システムにおいて、多様性や新規性など興味の拡散に効果的な指標と、関連

1) <https://sakuhindb.com/>(2022 年 3 月時点)

2) <https://reviewne.jp/>(2022 年 3 月時点)

3) <https://github.com/WorksApplications/Sudachi/>(2023 年 5 月 26 日確認)

4) <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt>(2023 年 5 月 26 日確認)

性や類似性などの推薦精度に当たる指標はトレードオフの関係であるという課題が存在する [19][20]. 提案手法は、間接的に関連する作品の提示による興味の拡散と、共起関係にある特徴語の提示による類似性の説明により、この課題の解決を図る。

3.5 類似作品ネットワークの構築方法

類似作品ネットワークの構築には、Python のライブラリである NetworkX⁵⁾を用いた。構築の手順は以下の通りである。

1. 作品 X のノードを生成する。作品 X とは、本研究が対象とする作品の 1 作目にあたるものと定義する。
2. 作品 X の類似作品 10 作品のノードを生成する。この時、ノードが既に存在する場合は生成しない。
3. 作品 X から類似作品へ有向エッジを生成する。
4. 作品 X に対する以上の処理を全作品分繰り返す。

エッジの重みには作品間の類似度の値を用いた。この場合、最短経路は最も重みの小さい経路となるため類似度の低い作品が優先されることとなる。しかし、本研究の目的はセレンディピティの高い推薦であることから、この重みを採用する。

3.6 類似作品ネットワークの詳細

本研究で構築した類似作品ネットワークの詳細について述べる。ノード数の合計は 33479、エッジ数の合計は 336854 である。ノード数は本研究が対象とする全作品数と同じ値である。エッジ数は、3.5 節で述べた構築方法からノード数の 10 倍となるはずであるが、実際にはそれよりも多い数のエッジが生成された。この理由は、一部作品において類似度の値が重複し類似作品が 10 作品以上出力されたためである。これによりノードから出ていくエッジの数に不均衡が生じるが、その影響は小さいと考えたため本研究ではこの点を考慮しない。

4 評価実験とその考察

4.1 評価実験の概要

提案手法が目的の実現に即した機能を有しているか評価するための予備的な実験を行う。実験の被験者は 20 代の男性 2 名である。被験者はスタート作品とゴール作品の 2 作品を指定し、その結果得られる推薦リストの作品に対して意外性などを評価する。その際、共通する特徴語も同時に提示する。推薦された作品を知らない場合は、Wikipedia 等を用いて作品の内容を把握した上で回答するよう指示する。

4.2 評価項目

本研究では、新規性、意外性、有用性、納得性 [21]、セレンディピティなどの指標を用いて提案手法を評価する。Ge ら [5] が述べるセレンディピティに関する重要な点から、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。この 3 つの要素はそれぞれ新規性、意外性、有用性に対応させる。アンケート項目は以下の通りである。

- 出力される各作品に対して回答する項目

- 項目 1 この作品を知っていますか？
項目 2 この作品が推薦されることを予測できた

- 項目 3 この作品を面白そうだと感じる
項目 4 この作品が推薦されたことに納得できた
項目 5 一つ前の作品と共通する特徴語は作品間の共通点として適切である
項目 6 この作品の内容とスタート作品の内容は類似していると感じる
項目 7 この作品の内容と一つ前の作品の内容は類似していると感じる
項目 8 この作品の内容とゴール作品の内容は類似していると感じる

- 推薦リスト全体に対して一度だけ回答する項目

- 項目 9 どのような意図でこの 2 作品を入力しましたか？
項目 10 推薦リスト全体に対して納得感がある
項目 11 推薦リスト内の作品をスタート作品やゴール作品に指定して別のつながりを探したい
項目 12 この推薦システムを自由に利用できるなら、どのように利用したいですか？

回答項目は 5 段階のリッカート尺度を用いた。項目 1 の回答項目のみ「名前も内容も知っている」「名前は知っているが内容は知らない」「名前も内容も知らない」の 3 段階である。一つ目に経由する作品に対する項目 7 への回答は項目 6 と同義であるため、項目 7 の回答数は他の項目より 2 つ少ない。項目 9 及び項目 12 は自由記述による回答形式である。

4.3 比較手法

比較手法として我々が提案した複数ジャンルを横断する推薦手法 [8] を採用する。比較手法では「ジャンル依存度」という指標を定義した。ジャンル依存度の計算方法を式 (4) に示す。この二つの指標はそれぞれ、計算する際の全文書の範囲を全ジャンルの全作品と定義して算出した IDF と各ジャンルの全作品と定義して算出した IDF である。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、大きい値を取る。そして、ジャンル依存度が上位 1% である単語を類似度計算の際に除外する。

$$\text{ジャンル依存度} = \frac{\text{全ジャンル IDF}}{\text{ジャンル別 IDF}} \quad (4)$$

比較手法を用いて行った評価実験の被験者は 20 代から 30 代の 15 名である。被験者が作品を入力すると類似作品 10 作品が出力されるため、それらの作品に対して評価する。入力された作品は述べ 52 作品であるため、評価対象の合計は 520 作品である。

比較手法における課題点は 3.4 節で述べた偶発性やユーザの能動的な働きかけの不足、意外性と類似性の両立が挙げられる。他にも、同一シリーズの推薦による意外性の低下や、作品の内容を反映しない抽象的な特徴語の影響で関連の薄い作品が推薦される問題などが挙げられる。評価実験においては、これらの課題が解決可能であるかについて考察する。ただし、被験者の数が大きく異なることから厳密な比較は困難であるため、あくまで傾向程度の分析であることに留意する必要がある。

5) <https://networkx.org/>

4.4 実験結果

被験者 A が指定した作品は「コードギアス 反逆のルルーシュ (アニメ)」「ハリー・ポッターと死の秘宝 PART2 (映画)」である。被験者 B が指定した作品は「鋼の錬金術師 (漫画)」「残穢-住んではいけない部屋-(映画)」である。推薦結果と各作品間で共通する特徴語をそれぞれ表 1 と表 2 に示す。推薦リストはスタート作品からゴール作品へ向かって上の作品から順番に經由したことを表している。また、評価アンケートの結果を図 1 に示す。

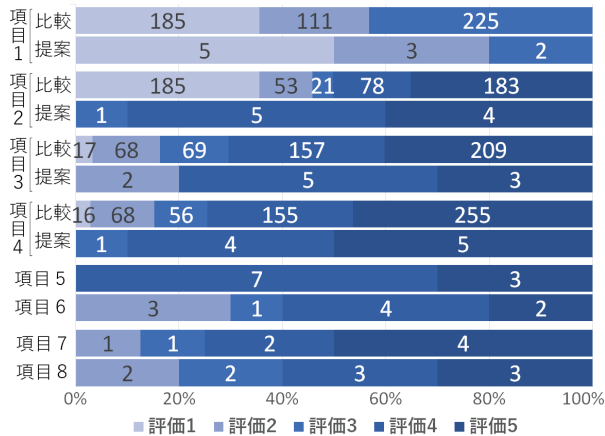


図 1 実験結果

4.5 推薦リストに関する考察

被験者 A の推薦リストについて考察する。スタート作品から「劇場版マジスティックプリンス-覚醒の遺伝子-(アニメ)」まではロボットが登場する作品という点で共通している。次に經由する「Infini-T Force ガッチャマンさらば友よ (アニメ)」との関連は曖昧であるが、いずれもテレビ番組として放送された作品と関連する映画作品という共通点がある。「劇場版 仮面ライダーウィザード インマジックランド (映画)」は魔法をテーマにした仮面ライダーシリーズの作品であり、この作品以降は魔法という共通点でつながっている。

被験者 B の推薦結果について考察する。初めにスタート作品と同一シリーズのゲームが推薦された。次に同じゲームジャンルという共通点により、ホラーゲームが推薦され、次に同一シリーズである「サイレントヒル (映画)」が推薦された。その後は、ホラー映画という共通点の「仄暗い水の底から (映画)」が推薦され、ゴール作品に到達する。

被験者 A の推薦リストはロボットが登場する作品群と魔法に関連する作品群に二分できる。また、被験者 B の推薦リストについても、鋼の錬金術師シリーズとホラー作品に二分される。

ジャンルの観点から推薦作品の変遷を見ると、被験者 A の推薦リストはアニメジャンルのスタート作品と映画ジャンルのゴール作品のつながりとしてアニメ映画を經由している。また、被験者 B の推薦リストは漫画ジャンルのスタート作品と映画ジャンルのゴール作品のつながりとして一度ゲームジャンルを經由している。これはホラー映画であるゴール作品への最短経路を算出する際に、メディアミックス作品であるサイレントヒルシリーズを經由したためである。

これらの傾向について、類似性を評価する項目 6,7,8 から詳しく検証する。項目 6 の評価分布と最頻値の値から一つ手前の作品との類似性は提供できたと考えられる。項目 7,8 はいずれも 1 の評価はないものの、2 から 5 の間に評価が分散している。これは、推薦リストの前半はスタート作品と類似しており、後半はゴール作品と類似していると評価されたためである。

つまり、作品間の最短経路は前半にスタート作品の要素を持ち、後半にゴール作品の要素を持つような経路が算出されやすい傾向がある。この特徴は推薦される作品がスタート作品とゴール作品に直接類似していることを示している。最短経路の中盤において間接的に関連する作品が推薦されることもあるが、推薦リスト内の作品全てが間接的な関連による推薦とは言えない。したがって、提案手法の機能は間接的に関連する作品の提示という目的の達成には不十分であった。今後は間接的に関連する作品を推薦できるように類似作品ネットワークのエッジの重みを工夫するなどの対策が考えられる。

4.6 共通する特徴語に関する考察

共通する特徴語の適切さを評価する項目 5 の最頻値は 4 である。1,2,3 の評価数が 0 個であることから、作品間の共通する特徴語は適切であったと考える。

一方で「テレビ」「視聴者」などアニメジャンルやドラマジャンルでのみ多く使われる単語や「ゲーム」「操作」などゲームジャンルでのみ多く使われる単語が存在する。特に「鋼の錬金術師～翔べない天使～(ゲーム)」と「サイレントヒル(ゲーム)」の間で共通する特徴語は 6 個中 4 個がゲームジャンルでのみ使われる、作品の内容を反映しない単語である。これらのような特定ジャンルでのみ多く使われる単語は我々が定義した「ジャンル依存度」により除去することが可能である。したがって、今後は類似作品ネットワークの重みにジャンル依存度を導入するという対策が考えられる。

4.7 偶発性に関する考察

提案手法の目的の一つである偶発的な作品の発見の支援という観点から考察する。この目的を評価する質問項目として、新規性を評価する項目 1 と意外性を評価する項目 2 が挙げられる。

提案手法における項目 1 の最頻値は「名前も内容も知らない」であり、比較手法における最頻値は「名前も内容も知っている」である。比較手法では、単一の作品の類似作品が推薦されるため入力によっては同一シリーズのみが推薦されるという問題が発生した。この問題が比較手法における「名前も内容も知っている」の評価数の増加につながった。提案手法では同一シリーズの推薦は含まれるものの推薦結果全てが同一シリーズにはならないため、比較手法よりも新規性が高い傾向になると考えられる。一方で、作品を既知であるかについては被験者の個人差による影響が大きく、被験者 B が「名前も内容も知らない」と評価した作品はなかった。

次に、項目 2 について考察する。提案手法における最頻値は 4 であり、1 と 2 への評価は 0 個だった。比較手法に対する評価は 1 と 5 に分散しているものの、最頻値は 1 である。項目 2 は 5 段階評価の値が小さいほど意外性の高い推薦であることを示すため、提案手法による推薦は意外性が低い可能性がある。この点について、被験者 B からは「有名なホラー作品が多く推薦されたため

表 1 被験者 A の推薦結果

推薦リスト	共通する特徴語
コードギアス 反逆のルルーシュ (アニメ) 革命機ヴァルヴレイヴ (アニメ) クロスアンジュ 天使と竜の輪舞 (アニメ) 劇場版 マジェスティックプリンス -覚醒の遺伝子- (アニメ) Infini-T Force ガッチャマン さらば友よ (アニメ) 劇場版 仮面ライダーウィザード インマジックランド (映画) ハリー・ポッターと炎のゴブレット (映画) ハリー・ポッターと死の秘宝 PART2 (映画)	ロボット, コード, サンライズ, ギアス, ロボアニメ 視聴者, ロボット, 機体, ギアス, サンライズ アンジュ, 機体 戦死, 振り返り, tv 世界, 劇場版, 科学, インフィニティー, テレビ ドラゴン, 魔法, ウィザード 騎士団, ロン, 魔法, 不死鳥, ハリーポッターシリーズ

表 2 被験者 B の推薦結果

推薦リスト	共通する特徴語
鋼の錬金術師 (漫画) 鋼の錬金術師 ~翔べない天使~ (ゲーム) サイレントヒル (ゲーム) サイレントヒル (映画) 仄暗い水の底から (映画) 残機 -住んではいけない部屋- (映画)	マスタング, 錬成, 大佐, 練成, 付録, アル, ハガレン ゲーム, 武器, クリア, エンディング, ムービー, 操作 ゲーム, 不気味, 怖い, 恐怖, モンスター, ダリア, ホラー, アレッサ 母親, 不気味, 和製, 怖い, 映画, 恐怖, ホラー マンション, 怪談, 怪異, 怖い, リング, 恐怖, ホラー

推薦結果を予測できた」という意見が挙げられた。推薦結果に知名度の高い作品が多く含まれる原因については、レビュー数が多い作品ほど多くの作品で使われる特徴語が抽出されやすくなるため、他作品の類似作品として出力されやすい、つまり類似作品ネットワーク上のハブとなっている可能性が挙げられる。そのため、類似作品ネットワークの分析を通して、知名度の高い作品のつながりやすさや推薦されやすい作品の傾向について調査を進める。

入力する作品によって推薦される作品の傾向が異なるという特徴は、提案手法が偶発的な作品の発見に寄与していると解釈することもできる。しかし、評価実験においてこの特徴は被験者ごとの評価の差につながる。したがって、今後は被験者の数を増やした上で、偶発性を適切に評価できる実験を設計する必要がある。

4.8 ユーザの能動的な働きかけに関する考察

提案手法の目的の一つである能動的な作品発見の支援という観点から考察する。この目的を評価する質問項目として、項目 9,10,11,12 が挙げられる。

項目 9 では「何度も見返すほど好きな作品を入力した」「ジャンルの異なる好きな作品を入力した」という回答が得られた。また、項目 12 では「好みの映画などを入力して作品を探してみたい」「自分好みのマイナーな作品を探すために利用したい」という回答が得られた。これらの意見から、提案手法を利用する際には好みの 2 作品を入力し、嗜好に合致する意外性の高い作品を探すというユースケースが多くなると考えられる。

推薦リスト全体の納得感を評価する項目 10 はいずれの被験者も 4 と評価した。したがって、入力した 2 作品から得られる推薦リストとしてユーザに納得感を提供できたと考える。一方で、項目 11 については、推薦リストの作品が全て既知であった被験者 B が 2 と評価し、ほとんどの作品が未知であった被験者 A が 4 と評価した。このことから、推薦結果の新規性が高いほど探索的に作品を探したくなる傾向があると考えられる。

これらの結果から、能動的に作品を探したいというユーザの意思が存在する一方で、現時点ではそれを支援

する機能が十分でない可能性が示唆された。提案手法が提供する機能について被験者からは「最短経路以外のいくつかの経路を出して欲しい」という意見が上がった。このことから、今後は推薦リスト中の作品を再入力できる機能の他、経路を複数提示する機能などの実装により能動的なユーザ体験を支援する必要がある。

4.9 興味の拡散と類似性の両立に関する考察

提案手法では興味の拡散を評価する指標と、類似性などの推薦精度を評価する指標はトレードオフの関係であるという課題 [19][20] の解決を図る。具体的には、間接的に関連する作品の提示により興味の拡散を促し、共通する特徴語の提示により類似性の説明を提供する。

間接的に関連する作品の提示に関しては 4.5 節で述べた通り、実現できなかった。一方で、納得性を評価する項目 4 と共通する特徴語を評価する項目 5 は高い評価であったことから、推薦された作品間の類似性の説明は提供できたと考えられる。また、有用性を評価する項目 3 についても比較的高く評価されたため、提案手法はユーザに有用な作品を推薦できたと考えられる。

以上の考察から、提案手法の機能は間接的に関連する作品の提示による興味の拡散の支援には不十分であったが、共通する特徴語を用いた類似性の説明は納得性や有用性の向上に有効であったと考えられる。

4.10 セレンディピティに関する考察

本研究におけるセレンディピティの定義は「未知の作品」かつ「予期できない作品」かつ「評価者が面白そうだと感じた作品」である。これらの条件と、各アンケート項目の評価値を以下のように対応付ける。

未知の作品 項目 1 において「内容も名前も知らない」「名前は知っているが内容は知らない」と回答された作品

予期できない作品 項目 2 において、評価値が 1,2 であった作品

評価者が面白そうだと感じた作品 項目 3 において、評価値が 4,5 であった作品

セレンディピティの定義を満たす作品の数とその割合を

表3 セレンディピティな作品の割合

手法	全体		意外性除く	
	作品数	割合	作品数	割合
比較手法	101/520	19.4%	160/520	30.8%
提案手法	0/10	0.0%	6/10	60.0%

表3に示す。提案手法によって推薦された作品の中で「予期できない作品」を満たす作品は存在しなかった。そのため、セレンディピティの条件を満たす作品の数も0個である。この点から、提案手法はセレンディピティの向上を実現できなかった。

一方で、意外性の条件を除いた場合該当する作品は10作品中6作品となる。したがって、セレンディピティを向上させるためには意外性を向上させる必要がある。提案手法は作品間の最短経路を算出するため同一シリーズが推薦リストに含まれる可能性が高い。また、同一シリーズが多く存在する作品群は類似作品ネットワークにおいてクラスターを形成しており、クラスター外への経路の選択肢が限られている。このような作品をスタート作品やゴール作品に指定した場合、最短経路の一部が固定されるという問題が発生する。これらの問題は意外性を低下させる要因である。そのため、今後は最短経路以外の経路の提示や作品間の経路の算出方法の工夫などの対策が必要である。他には、探索的にシステムを利用できる機能の追加により、ユーザが満足する推薦リストが得られるまで作品探索を続けられるようにするという対策も考えられる。

5 おわりに

本研究では、偶発的な作品の発見やユーザの能動的な働きかけを支援するセレンディピティ指向推薦システムを提案した。この提案システムは、間接的に関連する類似作品の提示により偶発的な作品の発見を支援する。また、ユーザは推薦される作品を予測しながら作品を入力するため、ユーザが能動的に作品を探索できると考えた。

評価実験の結果、提案手法は入力した作品と直接関連する作品を推薦しやすい傾向があることが明らかになった。そのため、意外性やセレンディピティは低く評価された。一方で、共通する特徴語を提示する機能は作品間の類似性の理解に効果的であった。今後は、提案手法の狙いである間接的に関連する作品の提示を実現するため、インタラクティブな機能の実装や類似作品ネットワークの重みの変更などが必要である。

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] 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).
- [6] Ziarani, R.J., and Ravanmehr, R.: Serendipity in recommender systems: a systematic literature review, Journal of Computer Science and Technology, Vol.36, pp.375-396 (2021).
- [7] 奥健太, 服部文夫: セレンディピティ指向情報推薦のためのフュージョンベース推薦システム. 知能と情報 (日本知能情報ファジィ学会誌), Vol.25, No.1, pp.524-539 (2013).
- [8] 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薦手法の提案, 情報処理学会論文誌, Vol.64, No.1, pp.55-66 (2023).
- [9] Cantador, I., Fernández-tobías, I., Berkovsky, S. and Cremonesi, P.: Cross-Domain Recommender Systems, Recommender Systems Handbook, pp.919-959, Springer (2015).
- [10] 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).
- [11] 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).
- [12] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J.: GroupLens: an Openarchitecture for Collaborative Filtering of Netnews, Proceedings of the 1994 ACM conference on Computer supported cooperative work (In CSCW 1994), pp.175-186 (1994).
- [13] Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T.: Evaluating Collaborative Filtering Recommender Systems, ACM Transactionson Information Systems (TOIS), Vol.22, No.1, pp.5-53 (2004).
- [14] Thudt, A., Hinrichs, U. and Carpendale, S.: The Bohemian bookshelf: Supporting serendipitous book discoveries through information visualization, Proceedings of CHI'12, pp.1461-1470 (2012).
- [15] Lan, M., Tan, C.L., Su, J. and Lu, Y.: Supervised andtraditional term weighting methods for automatic text categorization, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.31, No.4, pp.721-735 (2009).
- [16] 河合末夢, 能上慎也, 保坂忠明, 安藤晋: 商品間のネットワーク構造を用いた意外性を高める推薦モデルの提案. 情報処理学会研究報告, Vol.2016-MPS-107, No.15, pp.1-7 (2016).
- [17] 大坪五郎: 人間の知的活動支援へのアプローチ特集: 「人間」に情報を推薦することについて考える. ヒューマンインタフェース学会誌, Vol.9, No.4, pp.21-26 (2007).
- [18] 松島ひろむ, 森澤 竣, 石山 琢己, 山名 早人: 推薦システムにおける推薦理由の説明可能性に関するサーベイ, 第20回情報科学技術フォーラム (FIT2021), No.D-012, pp.171-178, (2021).
- [19] Hurley, N. and Zhang, M.: Novelty and Diversity in Top-N Recommendation - Analysis and Evaluation, ACM Transactions on Internet Technology, Vol.10, No.4, Article 14, pp.1-30 (2011).
- [20] Maccatrozzo, V., Terstall, M., Aroyo, L. and Schreiber, G.: SIRUP: Serendipity in recommendations via user perceptions, Proceedings of the 22nd International Conference on Intelligent User Interfaces, pp.35-44 (2017).
- [21] 中山田淳, 笹嶋宗彦, 來村徳信, 溝口理一郎: タスクに誘発される多様な広告推薦, 人工知能学会全国大会論文誌, Vol.24, pp.1D2-2 1-4 (2010).