

物語作品の短文要約によるストーリーの特徴分析

藤川 雄翔^{†,a} 松下 光範^{†,b} 山西 良典^{†,c}

[†] 関西大学大学院総合情報学研究科 ^{††} 関西大学総合情報学部

a) k920357@kansai-u.ac.jp b) m_mat@kansai-u.ac.jp c) ryama@kansai-u.ac.jp

概要 本研究の目的は、物語作品のコンセプトに着目した情報アクセス技術の実現である。物語作品のコンセプトとは「敵対する組織に属する男女二人が恋に落ちる話」のように、その物語の中核をなす内容を端的に言語化したものである。こうした研究を進める上の基盤となるコンセプトデータセットを構築するため、物語作品の読者にその物語を端的に要約した短文を生成してもらい収集した。本稿では、収集した短文の特徴を分析し、要約された短文がコンセプトとして利用可能であるかについて検討を行う。

キーワード 物語作品, コンセプト, 要約, 短文

1 はじめに

物語作品におけるストーリーには基本的なコンセプトがある。そのコンセプトにしたがってキャラクターが行動することにより、ストーリーは駆動していく。コンセプトとはストーリーの土台となるアイデアであり、物語を創作する際の起点としてだけでなく、他者に物語作品を推薦するための簡単な説明としてや、漠然としたストーリーの表現から物語作品を探索する際のクエリとして活用することができる。本研究では、“敵対する組織に属する男女二人が恋に落ちる話”のようにストーリーの内容を端的に要約した文章をコンセプト文と呼ぶ。

物語作品のストーリーのコンセプトを示すコンテンツとして要約文が考えられる。Wikipedia などの Web 百科事典や Amazon や楽天といった EC サイトに記載されるあらすじは物語作品の要約文である。これらの要約文は、物語作品全体の要約や 1 巻ごとの要約、その物語作品の“世界観”に言及した要約、その物語作品のメタコンテンツが付与されたもの [1] となっており、ストーリー進行に沿って要約文は長くなり、その物語作品の内容情報を理解することは難しい。特に、物語作品のメタ情報（売上や掲載誌など）が付与された要約文に関しては、メタ情報そのものは物語作品のストーリーの内容に直接関係がないため、ストーリーを理解するうえではノイズとなりえる。一方で、コンセプト文は短く端的にその物語作品のストーリーの内容を伝えるため、コンセプト文の作成者にとってストーリーを説明する上で重要な情報が記述されると推察される。

本研究では、ストーリーのコンセプトに基づいた物語作品への情報アクセスを可能にすることを目的とする。コンセプトに基づいた物語作品への情報アクセスは、従来の物語作品の検索に用いられるジャンルや著者などと

いった書誌情報よりもストーリーの内容を踏まえた検索が可能になると期待される。ストーリーのコンセプトに基づいた物語作品への情報アクセスを可能にするためには、個々の物語作品に対するコンセプト文が必要になる。しかし、ストーリーのコンセプトはストーリーの見どころを伝えるハイライトやストーリーの流れを示すダイジェストと異なり、コンセプト文そのものを読むことでのエンタテインメント性は高くない。そのため、コンセプトが記述されたリソースは少なく、物語作品についてのコンセプト文を収集することは容易ではない。コンセプト文を機械的に自動生成するアプローチも考えられるが、生成モデルの構築には物語作品とその物語作品に対応するコンセプト文のデータセットが必要となる。

こうした背景の下、本稿では人手により、コンセプト文のデータセットを構築する。コンセプト文の特徴を分析し、分析結果として得られる特徴を用いたコンセプト文の自動生成の可能性を検討する。

2 コンセプト文の収集

本研究ではコンセプト文を収集するために、関西大学総合情報学部所属する学部生を対象としたアンケートを行った。自身の好きな物語作品を他者に推薦するという状況設定のもと、287 件の短文を収集した。短文の記述に際し、(1) 140 字以内で要約すること、(2) 文の末尾を「～物語」で締めくくることが、(3) その物語作品の固有名詞 (e.g., キャラクター名, 組織名) を使わないこと、の 3 つの条件を課した。アンケート収集の際、短文に加えてその物語作品のタイトルとジャンルについても回答を課した。収集された 287 件の短文のうち、同一の物語作品について 3 件以上得られた物語作品は 19 作品あり、1 作品あたり最大 11 件の短文が得られた。本稿では、これらのうち物語作品に固有の名詞が含まれていたものを除いた 97 件を対象として特徴分析を行う。

3 分析

前章の方法で収集した短文は、その物語作品に対する捉え方が回答者によって異なる場合がある。そこで本稿では、収集した短文各々の文章ベクトルを算出し、同一物語作品の文章ベクトルの類似度から、物語作品が要約された短文が物語作品としてのコンセプトとして利用可能であるか検討する。

3.1 前処理

形態素解析器には MeCab (ver. 0.996)¹を採用し、形態素解析用辞書は ipadic-NEologd²を用いた。Slothlib³に含まれる単語と、1文字のひらがなとカタカナをストップワードとして設定した。また、どのコンセプト文も文末は「～物語」で締めくくるところから、「物語」の単語も除外することとした。形態素解析後、名詞と動詞（それぞれ非自立を除く）、形容詞、副詞のみを抽出した。

3.2 短文のベクトル化

本稿では、短文の文章ベクトルの算出に word2vec[2]を用いた。鈴木ら [3] が日本語版 Wikipedia の本文全文から作成した Wikipedia エンティティベクトルモデル (次元数は 200 次元)⁴を用いて、短文に含まれる単語の分散表現ベクトルを得た。このとき、「余儀ない」「謎解き」「ずば抜ける」の 3 単語は分散表現ベクトルを得ることができなかつたため、除外することとした。本稿では、1つの短文に含まれる全単語の分散表現ベクトルの平均を文章ベクトルとして算出した。

4 考察

図 1 は、3.2 節で得られた 200 次元の文章ベクトルを Scikit-learn (ver. 1.3.2) の TSNE を用いて 2 次元に圧縮し、散布図にプロットしたものである。図 1 より、同一物語作品の短文群にまとまりがあることが確認された。これらのコンセプト群の文章を確認したところ、同一もしくは類似した単語が複数含まれていた。ただし、同一物語作品の群から外れる短文もいくつか確認された。これらの内容を確認したところ、短文の長さが他の短文に比べて短い、他の短文には含まれていない単語が含まれている、という傾向が確認された。また、異なる物語作品の短文であるにも関わらず、散布図上で群を形成しているように見えるものが観察された (図 1 中 A)。これは同一物語作品群から外れていた物語作品同士の短文が、200 次元の文章ベクトルを 2 次元に圧縮したため

に近くにプロットされた可能性がある。

表 1 に、アンケートによって収集された短文の一部を示す。本稿では特に、作品単位で群としてのまとまりが見られた“ブルーロック” (図 1 中 B) と作品内で 2 つの群に分かれて分布した“ONE PIECE” (図 1 中 C) の 2 作品について考察する。“ブルーロック”では、どの短文にも“ストライカー”という単語が含まれることや、ストライカーを“目指す”、“誕生させる”などのその物語作品のキャラクターたちの目的が共通して記述されている。このことは、本作品を象徴する短文では、同様の単語が用いられて類似した短文によって、作品のある特定の 1 側面について記述されることが示唆される。“ONE PIECE”も同様に、“仲間”という単語や“海賊王を目指す”等のキャラクターの目的が共通して記述されている。しかし、キャラクターの目的を達成するための行動として、“冒険をする” (図 1 中 C-(1)) と“敵と戦う” (図 1 中 C-(2)) という 2 種類の異なる見解が記述されたことで、短文群が二分したと考えられる。本作品は複数の側面から作品が評価され、本作品を象徴する短文が記述されることを示唆する。

5 まとめ

本稿では、読者から生成された物語作品のコンセプトとなりうるコンセプト文を収集し、物語作品のコンセプトデータセットの作成を試みた。収集したコンセプト文から文章ベクトルを算出し、同一物語作品の文章ベクトルの類似度から要約されたコンセプト文が物語作品のコンセプトとして利用可能か検討を行った。同一物語作品同士で、文章ベクトルにある程度まとまりが見られることから、物語作品のコンセプト要約はコンセプトとして利用可能であることが示唆された。

今後は、それぞれのコンセプト文の文脈を意味的に考慮した文章ベクトルの算出を行い、要約文からのコンセプト要約の自動生成を試みる。

謝辞

本研究は、一部、科研費 22K12338 および 20K12130 の支援のもと行われた。記して謝意を表す。

参考文献

- [1] 山西良典, 西原陽子, 松下光範: 語彙の標本化と量子化によるあらすじの特性表現に関する基礎検討, 第 7 回コミック工学研究会, pp. 31–38 (2022).
- [2] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality, *NIPS2013*, pp. 3111–3119 (2013).
- [3] 鈴木正敏, 松田耕史, 関根 聡, 岡崎直観, 乾健太郎: Wikipedia 記事に対する拡張固有表現ラベルの多重付与, 言語処理学会第 22 回年次大会, pp. 797–800 (2016).

¹<https://taku910.github.io/mecab/> (2023/10/28 確認)

²<https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md> (2023/10/28 確認)

³<http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt> (2023/10/28 確認)

⁴<https://github.com/singletonue/WikiEntVec/releases> の `jawiki.all_vectors.200d.txt.bz2` (2019 年 5 月 20 日版)

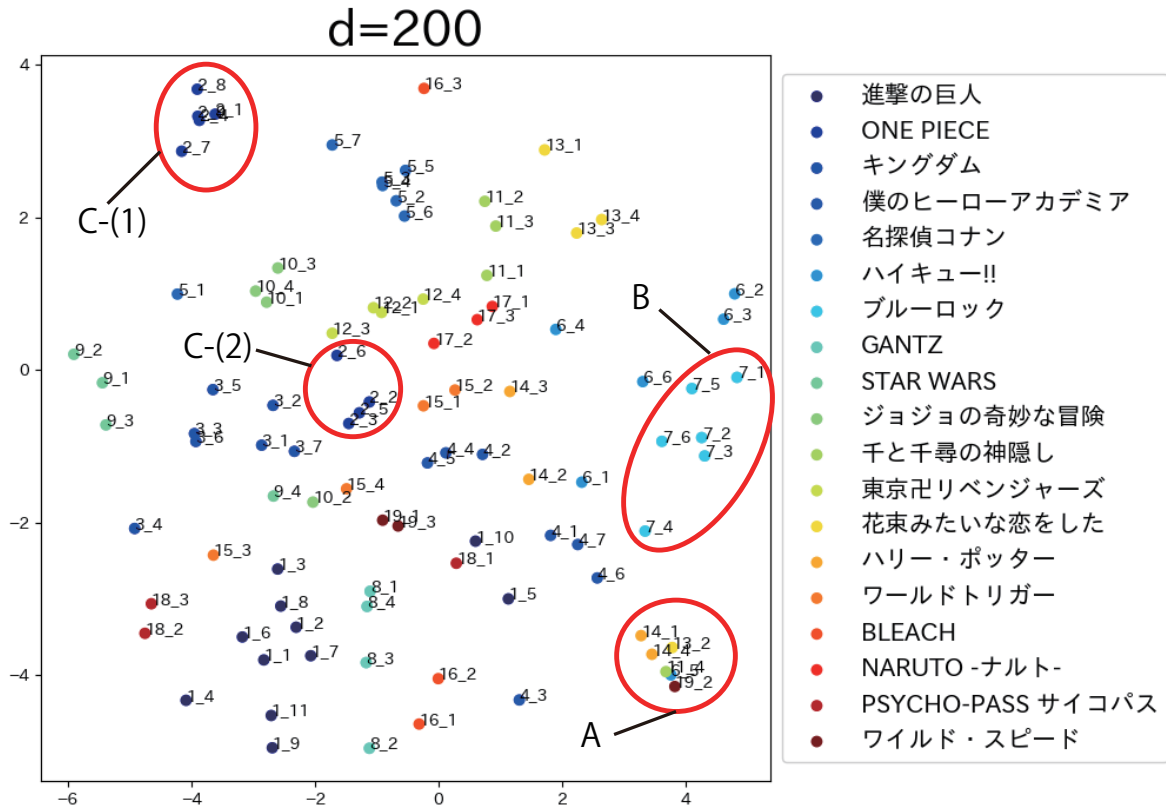


図 1 3.2 節で得られた 200 次元の文章ベクトルを 2 次元に圧縮し t-SNE を用いてプロットした散布図. それぞれのラベルは作品番号と文章 id を示している.

表 1 アンケートによって収集されたコンセプト文 (一部)

id	タイトル	コンセプト文
2.1	ONE PIECE	夢見る少年が仲間を集めながら海の王を目指す物語
2.2	ONE PIECE	主人公が航海に出て信頼できる仲間を集め、数々の敵と戦いながら、夢を追い求める姿や仲間との友情、強くなるための努力を描いた物語。
2.3	ONE PIECE	「海賊王になる」という夢を持ち、目指す主人公がその夢の実現のために航海し、その過程でそれぞれ「世界一の大剣豪」や「世界地図を描く」といった夢や野望を持った多種多様な仲間たちを集めながら航海していき、またその中で国や人を救うために悪者と戦いながら心身ともに成長していく物語
2.4	ONE PIECE	子供の頃に憧れた海賊を追いかけて仲間を集めながら冒険し、海賊王を目指す物語。
2.5	ONE PIECE	とある果実を口にしたことで不思議な能力を得た主人公が仲間とともに海賊王を目指す物語
2.6	ONE PIECE	自分の夢を叶えるため主人公が海に出て旅をしていく中で仲間を集め強敵や困難な出来事また悲しい出来事や自分の中での葛藤を乗り越え海賊王を目指す物語。
2.7	ONE PIECE	麦わら帽子がトレードマークの海賊である主人公が多くの海賊が蔓延る時代に、伝説の大秘宝を目指して仲間と共に航海に出て、冒険を繰り返す物語。
2.8	ONE PIECE	一人の青年が仲間と共に海賊王を目指す物語
2.9	ONE PIECE	麦わら帽子を被った少年が仲間を集めて海賊の王を目指す物語
7.1	ブルーロック	「日本一のストライカー」を目指し、強豪選手たちとの過酷なトライアウトに挑む物語
7.2	ブルーロック	主人公のエゴと才能を武器に、周りと競い合って世界一の点取り屋 (ストライカー) を目指す物語。
7.3	ブルーロック	集められた若きサッカー選手たちが、極限の環境の中自分のエゴを磨き、最強のストライカーを目指す物語。
7.4	ブルーロック	日本が W 杯優勝するために日本サッカーで欠けている絶対的エースストライカーを生み出す物語。
7.5	ブルーロック	無名の高校のサッカー部の主人公が全国から集められた高校生のフォワード達と勝負や切磋琢磨し、世界一のストライカーのストライカーを目指す物語
7.6	ブルーロック	日本中から集められた高校生のストライカーの中から、世界一のストライカーを誕生させる物語