

作中役割を考慮した物語文の類似文検索による 創作者の支援と感性の解析

Search for Similar Story Sentences based on Role of Characters
in order to Support and Analyze Contents Creator's Ideas

葛井 健文 上野 未貴 井佐原 均
Takefumi Katsui Miki Ueno Hitoshi Isahara

豊橋技術科学大学
Toyohashi University of Technology

A process of creating stories has been studied from various aspects. Recently, it has become important topic for the field of artificial intelligence. In the field, such studies are mainly divided into two groups, automatic generation of stories and assistance for human's creative activity. From the view of assistance for human's creative activity, we propose creative support method. This paper describes two systems; Creation support system and System of Search for Similar Story Sentences based on Role of Characters. In order to confirm the effectiveness of these proposed method, three types of experiments were carried out. These results suggested that the proposed system is useful for creating story and analyzing human's idea of the perspective of similar feeling.

1. はじめに

近年、文章解析として分散表現を用いた文のベクトル化の手法が数多く提案されており、機械翻訳や対話システムの応答文選択の精度が大幅に向上している。一方、物語文章の解析については作者固有の文体や作品の固有表現、セリフや回想文などが存在し、また評価指標も多岐にわたるため、文章の物語としての類似度の推定は困難である。そのため、物語の類似度を推定するための取り組みはとても重要であり、工学的な物語の自動生成と創作支援の両面において貢献できると考えられる。

本研究では、従来研究で提案された2種類のテンプレートの連携により物語のプロットを作成する創作支援システムと、新たに提案する類似した物語のあらすじを検索するシステムの連携により、より効率的に物語を作成できる創作支援手法を提案する。

2. 従来手法

2.1 創作支援システム

本研究で特に重要な創作支援システムに関する従来研究を示す。従来研究として、川越らは2種類のテンプレートを連携した創作支援システムを提案[kawagoe 15]し、著者らはこの創作支援手法の2種類のテンプレート間の連携を強化する想定感情線機能を追加した手法を提案した[葛井 17]。本稿では、上記の手法に基づく創作支援システムを従来手法と呼ぶ。

2.2 創作支援システムの概要

創作支援システムは3つの機能を持ち、ストーリーテンプレートを用いて物語のあらすじを作成するストーリー作成部、キャラクターテンプレートを用いて登場人物に関する設定を作成するキャラクター作成部、実際に作成されたプロットを用いて本文を入力するための本文入力部によって構成される。ユーザはこのシステムを用いることで物語のプロットを完成させることが出来る。

2.3 ストーリーテンプレート

物語のプロットを作成するための質問集合がストーリーテンプレートである。また、ストーリーテンプレート中の質問項目

には三幕構成[安藤 09]が用いられている。これにより、作成したプロットの物語中の時系列を整理出来る。

2.4 キャラクターテンプレート

登場人物を作成するための質問集合がキャラクターテンプレートである。キャラクターテンプレート内で作成した登場人物をグレイスの行為者モデル[樋口 89]に従って4つの役割のうちいずれか一つに当てはめる。グレイスの行為者モデルは物語の登場人物を6つの役割に整理したモデルであり、従来手法では、このうち主体、対象、援助者、敵対者を重要な役割であると考え、ユーザは作成したキャラクターを4つのうちどれか一つの役割に当てはめる。

2.5 想定感情線グラフ

2種類のテンプレート間の連携を強化するため、想定感情線グラフを導入した。想定感情線グラフは以下の機能を持つ。

- 想定感情線の手動入力
- キャラクター登場場面の表示
- 各期間の感情値の自動算出

2.6 従来研究の実験結果と提起

従来研究について得られた意見、課題点を以下に示す。

- 創作の参考になるアイデアをシステムから提示してほしい
- 両テンプレートともにバリエーションが少ない
- 物語の方向性が固定されやすい

本研究ではこれらの問題点を解決するため、保存機能の追加と、システム側から類似度に基づき参考となる作品を提示することで、作成できる物語のバリエーションの増加を目指す。

3. 提案手法

本章では、2.6で示した従来手法の問題点を解決し、より有用な創作支援手法となるように創作支援システムをあらすじ文章の内容から似た内容のあらすじを検索するシステム（以下、あらすじ検索システムとする）と連携して利用する手法を提案する。類似した物語を提示することは、創作支援システムのユーザにとって自身の参考になり、物語のバリエーションや方向性も新たな発想が得られると考えたためである。

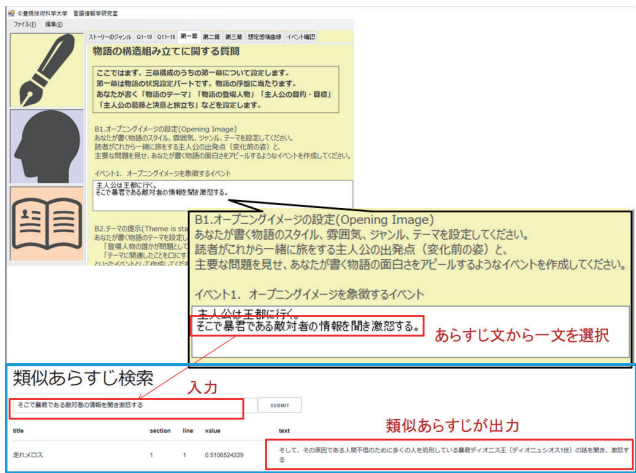


図 1: 創作支援システムに基づくあらすじ検索システムの利用

表 1: 実験に用いた PV-DM モデルの学習条件

文書サイズ	次元数	窓サイズ	エポック数
1MB	300	5	1000

3.1 あらすじ検索システムの概要

あらすじ検索システムは、物語のあらすじについて記した一文を入力として受け取り、類似したあらすじ文を出力するシステムである。図 1 に概要図を示す。

3.2 文章のベクトル化

文章の類似度などを計算する方法として文章のベクトル化が挙げられる。中でもニューラルネットワークによる学習による分散表現獲得手法として Paragraph Vector [Le 14] が存在する。Paragraph Vector は Word2Vec [Mikolov 13] から派生した手法で、本実験では PV-DM(Distributed Memory Model of Paragraph Vectors) モデルを用いる。表 1 にパラメータを示す。本稿では次に述べるコーパスを学習して PV-DM モデルをあらすじ文章モデルと呼ぶ。

3.3 コーパス作成と前処理

用いたコーパスは、類似していると判断された作品をユーザが読むことで創作への参考にする必要があるために、青空文庫収録作品のうち 2018 年 10 月 4 日時点で日本語 Wikipedia [Wikipedia] に記事があり、その中に「あらすじ」節を持つ 464 作品を対象とした。また、抽出の際に注釈やリンクの情報は消去している。本研究では、Wikipedia から抽出した各記事の文章から、「あらすじ」節に含まれる文章のみを利用し、文章コーパスとする。コーパス内の各文章には「どの作品のあらすじ文章か」を判別するためのラベル付けを行った。また、文書解析の精度向上および実際に創作に役立てるための情報に用いるため、コーパスに以下の 3 つの前処理をした。

3.3.1 物語に関係しない文章の削除

取得したあらすじ節の文章の中には、その物語の中の出来事の情報ではない、作品そのものへの解説や世界観の注釈、といった文章が含まれていることがある。本研究では創作支援システムに入力された物語のあらすじ文章と Wikipedia のあらすじ文章を比較し、二つの文章の類似度から物語の類似度を推定することが目的であるため、以下に示す内容を含む文章は今回のコーパス作成に不向きであるとして以下の操作をした。以下に操作内容を示す。

表 2: 三幕構成への分割を行ったあらすじ文章の内訳

内訳	第一幕	第二幕	第三幕
各幕を持つ作品数	462	427	357
各幕に含まれる文数	2542	4175	2354
各幕に含まれる単語数	56850	103612	50900

- 物語の歴史的背景や設定を補足する文章
- 複数言語での表記や、表現の違いを補足する文章
- 記事内で対象としている版以外の物語の展開へ言及している文章
- 作品の解説や制作事情など、物語の出来事以外に言及している文章

以上の操作の結果、あらすじ文が全て削除された作品が 2 作品あったため、対象作品の数は実質的に 462 作品となった。

3.3.2 物語の時系列の整理

Paragraph Vector を用いた文章の分散表現を用いて文書間の類似度を計算する際、「その文章はあらすじ文章中のどの位置にあった文章か、物語中のどのタイミングの出来事を表した文章か」という情報は考慮されない。しかし、物語においては「物語中で出来事が起きたタイミング」は非常に重要であり、表記の上では同じような文章でも、そのタイミングによって大きく意味合いが異なり、結果的に物語的な類似度は小さくなるのが予想される。そのため、コーパス中の各文章に対し、「どの作品のあらすじか」だけでなく、その文章は「創作支援システムのストーリーテンプレートで定義された三幕構成のうち、作品中で第一幕から第三幕のどのタイミングの出来事に当たる文章か」を判別するためのラベル付けをした。ラベル付けの方法として、あらすじ文章を文書量等を基準として均等に三分割するのではなく、各作品のあらすじ文章ごとにどのように三幕構成に分割できるかを人手で判断し、分類を行った。表 2 に各幕ごとの内訳を示す。

3.3.3 登場人物名の役割名への置換

あらすじ文章には、作品ごとに固有の登場人物の名前が含まれており、Paragraph Vector を用いて文章をベクトル化する際、固有表現はそのベクトルに強く作用する。一方で物語においては、「誰」が何をしたかは非常に重要であり、例えば戦いに勝利したのが主人公であるのか敵役であるのかは意味合いが大きく変わる。そこで、物語的に類似した文章を検索するために、登場人物名に関する固有表現と、その人物を表す代名詞を、創作支援システムのキャラクターテンプレートで定義された 4 つの役割名：主人公、対象、援助者、敵対者に置換する。登場した人物の各役割への当てはめは人手で行った。詳しいルールを以下に示す。

- 人物名と代名詞を単純に置き換える
- 各役割に人数の制限はない
- 役割名が連続するときは重複を削除する
- 「物語としての登場人物でない」人名に関しては置き換えない

3.4 あらすじ検索システムの実装

前節に示した挙動をする機能を実装した。入力文は創作支援システムにおける第一幕から第三幕の入力内容のうち一文を用い、その入力を分散表現としてあらすじ文章モデルの構築に用いた文章から類似度の高い文章を上位 n 文取り出し表示する。以降、 $n = 15$ とする。

表 3: 実験 1 の質問のアンケート結果 (抜粋)

質問内容	平均	各評価の頻度				
		1	2	3	4	5
第一幕で出力文と入力文章はどれくらい似ているか	4.1	0	1	2	4	5
第一幕で出力文が含まれる作品の第三幕のあらすじと選択した作品の第三幕のあらすじは似ていたか	3.8	2	0	2	2	6

3.5 創作支援システムとあらすじ検索システムの連携

あらすじ検索システムは、入力文として時系列とキャラクター役割を考慮したあらすじ文章を想定している。これは創作支援システムで作成されたあらすじ文章が満たしているものであり、創作支援システムの利用者はプロットの作成中に好きなタイミングであらすじ検索システムを利用して作成中のあらすじ文章と物語的に類似したあらすじ文章を検索することが出来る。創作中に参考になる作品の情報を提示することは有用であると考えられるため、この 2 つのシステムを使った創作支援によって効率的に人間の創作活動を支援できると考えられる。

4. 実験

研究目的に照らして、提案システムの有用性を確かめるため、以下の 3 種類の実験をした。また、実験は被験者に 2 種類のシステムと実験手順書、実験協力の同意書を送付し、被験者が各自で実験を行うという形式とした。実験にあたって、出力される可能性のあるあらすじを三幕構成に分割したものを被験者に提示している。

4.1 実験 1: 既存の物語に対するあらすじ文章モデルの評価実験 A

あらすじ検索システムを用いて、あらすじ文書モデルにより類似あらすじが算出される頻度およびユーザが類似すると感じる観点を解析する。

4.1.1 実験条件

成人男女 12 名を対象に、以下の手順により実験を行い、アンケートにより結果を確認した。

4.1.2 実験手順

被験者は、はじめに対象作品から一つ選択し、その作品を青空文庫から確認し創作支援システムを用いてあらすじを作成する。その後、作成したあらすじを一文ずつあらすじ検索システムに入力し、出力結果を確認した後アンケートに回答する。

4.1.3 対象作品

この実験で被験者があらすじを作成するために対象とした作品は「走れメロス」「雪の女王」「ジャックと豆の木」「赤ずきん」「ヘンゼルとグレーテル」の 5 作品である。

4.2 実験 1 結果

表 3 にアンケート中の 5 段階評価の各評価の回答数を示す。アンケートは 1-5 の 5 段階評価で、5 が最も高いとした。第一幕から第三幕まで、全ての入力に対して被験者が選択した作品と同一の作品のあらすじを出力した結果が多くみられる。

同一の作品が出力されなかった場合において、あらすじ文の類似度を判断した理由として「起きた出来事が似ていた」という回数が多く得られていることから、ユーザは別の作品のあらすじ文を比較したとき、出来事や状況を表す文章の類似度を似ていると感じる尺度に用いていると考えられる。入力作品と同一

表 4: 実験 2 のアンケート結果 (抜粋)

質問内容	平均	各評価の頻度				
		1	2	3	4	5
第一幕で出力文と入力文章はどれくらい似ているか	3.4	0	0	3	2	0
第一幕で出力文が含まれる作品の第三幕のあらすじと選択した作品の第三幕のあらすじは似ていたか	1.4	3	2	0	0	0

一の作品が出力されない場合、原因は大きく分けて二つに分類できる。

- あらすじ文章の粒度が異なるパターン
- 入力した出来事がコーパスに存在しないパターン

4.3 実験 2: 既存の物語に対するあらすじ文章モデルの評価実験 B

あらすじ文章モデルの構築に用いた文章内にあらすじ文が存在しない既存の物語のあらすじをあらすじ検索システムに入力し、出力結果についてのアンケートをすることであらすじ文書モデルにより類似あらすじが出力される頻度およびユーザが類似すると感じる観点を解析する。とくに、実験 1 に比べ明確な正解が存在しないためにユーザがどの観点から類似性を見出すかに重点を置く。

4.3.1 実験条件

成人男女 5 を対象に、以下の手順により実験を行い、アンケートにより結果を確認した。

4.3.2 実験手順

実験手順は、対象の作品が任意の創作物である以外は実験 1 と同様である。

4.4 実験 2 結果

表 4 にアンケート中の 5 段階評価の各評価の回答数を示す。入力した文と似たあらすじ文自体は出力されているが、あらすじ文章のうち一文が似ているからといってその作品が想定した作品と似ているとは限らないということが分かる。

一方で、物語中の部分的な展開や目的が似ている場合があり、その例を以下に示す。

実験 2 の入出力考察 1

入力「主人公は、対象が惨殺されたショックから禁酒に失敗し、バーで泥酔して失態を犯してしまう」
出力「しかし電話で連絡してきた柳吉に、お前は来ん方がええ、来たら都合が悪いと言われ、蝶子は発作的にガス自殺を図った」
(出典: フリー百科事典「Wikipedia:『夫婦善哉 (小説)』」)

以上の例の入出力に対して被験者からは「物語の全体的な展開は大きく違うが、入力文の該当場面と出力文の該当場面で主人公がどん底にいる描写は一致していた」という意見が得られている。この入力に対しては、他にも登場人物が失態を犯す描写を表す文章が出力されており、登場人物の状態を示す文として、またその文が含まれる物語の一場面は類似している部分があると考えられる。以上の事から、類似あらすじ文の検索によって提示された 2 つの物語間には、物語として目的、三幕構成における各幕、その幕内の一部分の展開、登場人物の行動といった要素的に似ている部分が見られると考えられる。

表 5: 実験 3 のアンケート結果

質問内容	被験者 1	被験者 2	被験者 3
第一幕で出力文と入力文章はどれくらい似ているか	4	4	4
第一幕で出力文が含まれる作品あらすじと選択した作品のあらすじは似ていたか	2	4	2

4.5 実験 3：システムを用いたプロット作成実験

創作支援システムを用いた物語プロット作成中に、あらすじ検索システムを用いてその作品のプロットと似たあらすじを持つ作品を提示することで創作に与える影響について調べる。

4.6 実験条件と手順

成人男性 3 名を対象に、物語のあらすじ作成中に類似した物語を提示することで自分の書きたかった物語の作成に役立ったかどうかを調べる。実験手順を以下に示す。被験者ははじめに創作支援システムを用いて物語のあらすじを作成する。作成中の以下のタイミングであらすじ検索システムに作成中のあらすじを入力し、出力結果を確認する。

- 第一幕に該当する回答が全て入力された時点、第一幕への回答から一文ずつ選択
- 第二幕に該当する回答が全て入力された時点、第二幕への回答から一文ずつ選択
- 第三幕に該当する回答が全て入力された時点、第三幕への回答から一文ずつ選択
- ストーリーテンプレートに被験者が考える必要な部分が全て入力された時点、すべての回答から一文を選択

各結果を確認した後、アンケートに回答する。

4.7 実験 3 結果

表 5 に 5 段階評価の各被験者の回答を示す。実験 2 と同様、あらすじ文章コーパスに用いた作品が含まれないため、入力文と似たあらすじ文は出力されているが、そのあらすじ文をもつ作品はあまり似ていないという結果が得られた。

出力されたあらすじを持つ作品は創作中の作品とあまり似ていないという結果が得られた一方で、多くの箇所です自身の創作に役立てられたという回答が得られた。

実験 3 の入出力考察 1

入力「自分には新しいメニューは無理だという主人公は夢の中でラーメン仙人に出会う」

出力「大和国に住む木こりの髪長彦は、ある日森の中で神に出会う」

(出典：フリー百科事典「Wikipedia：『犬と笛』」)

以上の例の入出力に対して被験者からは出力文を含む作品と自身の作成している物語は似ていないという解答があった一方で、「展開の被りが避けられた」「今後の展開の参考になった」という解答がされている。第一幕という物語の導入部分が類似していたが、そこから展開が異なっていくことから、創作中にユーザのイメージしていた展開以外のアイデアを提示できていたと考えられる。

5. まとめ

本研究では、創作活動支援のためのシステムとして 2 種類のシステムを組み合わせた手法を提案した。提案手法につい

て被験者実験をすることで、時系列と、キャラクター役割を考慮したあらすじ文章コーパスを用いて学習した Paragraph Vector モデルにより共通した要素を持つあらすじ文を出力できること、創作活動において、物語の創作中にその物語に共通した要素を持つ作品情報を提示することの有用性を確認した。今後の課題は以下の通りである。

- あらすじ文章コーパスの規模拡大、各文章の整理
- 創作支援システムの拡張によるユーザの類似あらすじ検索システムへの入力の手間の削減

謝辞

本研究の被験者実験に協力してくださった 12 名の方々、システムに様々な助言を下された作家 山口昌志様に心より感謝します。本研究は、一部 JSPS 科研費 (グラント番号:JP17K17809) および JST, ACT-I (グラント番号:JPMJPR17U4) の支援によります。

参考文献

- [kawagoe 15] Seiya Kawagoe and Miki Ueno and Hitoshi Isahara: A Study on the Efficiency of Creating Stories by the use of Templates, The 2015 International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA 2015), 978-1-4673-8142-0(2015)
- [葛井 17] 葛井 健文, 上野 未貴, 井佐原 均: 質問集合とグラフに基づく物語全体の流れを管理可能な創作支援システムの提案, 人工知能学会 2017 年度全国大会 4F1-3in2(2017)
- [安藤 09] 安藤 隼平, 加藤 正人 (訳), シド・フィールド (著): 映画を書くためにあなたがしなくてはならないこと シド・フィールドの脚本術, フィルムアート社 (2009(原著 1979))
- [樋口 89] 樋口 淳: 怖がることを習いに出かけた若者の話—A.J. グレマスの意味論をめぐる, 専修人文論集 (44), p93-129(1989)
- [小林 05] 小林 のぞみ, 乾 健太郎, 松本 裕治, 立石 健二, 福島 俊一: 意見抽出のための評価表現の収集, 自然言語処理, Vol.12, No.3, pp.203-222(2005)
- [Wikipedia] フリー百科事典 ウィキペディア日本語版 <https://ja.wikipedia.org>
- [Le 14] Quoc Le, Tomas Mikolov: Distributed Representations of Sentences and Documents, Proceedings of The 31st International Conference on Machine Learning (ICML 2014), pp. 1188 - 1196(2014)
- [Mikolov 13] Tomas Mikolov, Kai Chen, Greg S. Corrado, Jeffrey Dean: Efficient Estimation of Word Representations in Vector Space, conferencePoster-iclr2013-workshop(2013)