自然言語処理モデル T5 を用いたオンライン小説タイトル生成手法に関する研究

2019321 藤井 一志

令和 4 年度提出

目次

1章	はじめに	2
2章	関連研究	4
	2.1 タイトルに関する研究と深層学習を利用したタイトル生成	4
	2.2 オンライン小説に関する研究	4
3章	小説家になろうと T5 について	5
	3.1 オンライン小説投稿, 閲覧サイト「小説家になろう」について	5
	3.2 小説家になろう特有のタイトルについて	5
	3.3 T5(Text-To-Text Transfer Transformer)について	6
4章	実験1 (タイトル生成システムの制作と ROUGE 評価実験)	8
	4.1 実験1の目的	8
	4.2 実験1の方法	8
	4.3 実験1の結果	10
5章	実験2 (作品にふさわしいタイトルの判断基準の調査実験)	11
	5.1 実験2の目的	11
	5.2 実験2の方法	11
	5.3 実験2の結果	13
6章	実験3 (自動生成されたタイトルの印象評価実験)	14
	6.1 実験3の目的	14
	6.2 実験3の方法	14
	6.3 実験3の結果	17
7章	まとめ	21
	参考文献	22

1章 はじめに

近年, オンライン小説投稿, 閲覧サイトが人気である. サイト内で人気になった小説はアニメ化や書籍化といった様々なメディアミックス展開がされるため, 「小説家になろう」を始めとした小説投稿, 閲覧サイトでは多くの利用者が小説を読み, 投稿するようになっている.

ユーザが「小説家になろう」サイト内で読む作品を探す際, ユーザは作品のタイトルから読む作品を決める. これはサイト内で作品が掲載される際, あらすじよりもタイトルの表示が優先されるためである. 例えば, 下記のような作品のタイトルがある.

「小説家になろう」に含まれるタイトルの例

「婚約者様にはもうすぐ運命のヒロインが現れますが、せっかくなので暫定婚約者を漫喫します！ところでその呪い、嫌われ悪女の私が解いちゃダメですか？」

この例からもわかるように, 小説家になろうの作品タイトルには, 「ユーザがタイトルを読むだけで作品の世界観やジャンル, 起承転結, 主人公の立場などがある程度わかるものが望ましい」という特徴がある.

12月8日現在で, 「小説家になろう」内の掲載小説数は995,741作品もあり, ランキング内の一握りの作品以外の多くの作品が埋もれてしまっているのが現状である. そのため, 一見するだけで作品の内容がある程度理解することが可能であり, より多くのユーザの目を引くことのできるタイトルを生成する手法が作品を投稿するユーザによって求められている. ただ, このようなタイトルは小説家になろうに親しくない小説家に対してなじみの浅いものであり, 考える際に時間がかかってしまう. 一般的な小説家であれば, 編集者などと相談しながらタイトルを決めることが出来るが, サイトの大多数をアマチュアの小説家が占める小説家になろうにおいては, 編集者がいない小説家が多い. そのため, システムを利用してあらすじから自動でタイトルを生成し, アイデアの一助とすることが出来れば, 編集者のいない小説家にとって非常に役に立つことになる. このような, 機械を利用して, あらすじからタイトルなどを自動生成する技術として自然言語処理の研究がある. 本研究では, 自然言語処理モデルの一つである T5(Text-To-Text Transfer Transformer)を利用して, 編集者がおらず, ユーザの目を引きやすいタイトルを制作することに困難を感じている小説家になろうユーザの一助となるような, あらすじからタイトルを自動生成するシステムを制作することを目指した.

システム制作の際に利用した言語処理モデルの T5(Text-To-Text Transfer Transformer)とは, 様々な自然言語処理タスクの入出力がともにテキストになるよう問題形式を再定義することにより, 一つの事前学習済みモデルを多様なタスク用に転移学習させることがで

きる高い柔軟性を持ち、かつ、性能も優れている深層ニューラルネットワークである。この自然言語処理モデルを利用し、あらすじとタイトルのパターンを転移学習させることで、あらすじ文からタイトルを自動で生成するシステムを作成することができる。

本研究では、「小説家になろう」の作品を学習データとし、T5 に転移学習させることで、ユーザが入力したあらすじから、ユーザの目を引きやすいタイトルを自動生成できることを明らかにする。システム制作の際には、小説家になろう内で評価の高い作品データを学習させることで、前述した、一見するだけで作品の大まかなあらすじが分かり、ユーザの目を引くことが出来るタイトルを生成することが可能になることを目指した。また、人がタイトルをあらすじにふさわしい、と判断する基準をアンケート調査し、その結果で得られた判断基準を元に、人間が作成した元タイトルとシステム生成のタイトルを比較して、印象評価の実験を行った。本研究では、システム生成されたタイトルが人手のタイトルより高い評価を得ることで、本システムのタイトル生成手法の有効性を示すことを目的とする。

以下、2章ではタイトルに関する関連研究と、深層学習を用いたタイトル生成に関する関連研究について述べる。また本実験と同じく、オンライン小説に焦点を当てた関連研究についても述べる。3章では、オンライン小説投稿、閲覧サイト「小説家になろう」の概要とサイト内に投稿されている作品が持つ、タイトルの特異性について述べる。また、自然言語処理とはどのようなものかと、実験で使用した自然言語処理モデル T5(Text-To-Text Transfer Transformer)の概要についても述べる。4章では小説家になろう内で高い評価を得た作品データをもとに T5 を転移学習させ、あらすじを入れるだけでタイトルを自動生成するシステムを制作し、自動生成したタイトルと元のタイトルの類似度を ROUGE で機械評価したことについて述べ、5章では、人がタイトルをあらすじにふさわしいと判断する際の判断基準についてアンケート調査を行い、その結果について述べ、6章では、5章の結果をもとに人手のタイトルとシステム生成のタイトルの印象の違いの評価実験を行い、本システムの有効性について述べる。最後に7章では全体をまとめる。

2章 関連研究

2.1 タイトルに関する研究と深層学習を利用したタイトル生成

オンライン小説に限らず、タイトルに焦点を当てた研究はさまざまな分野で行われてきた。福田らは、29,500 件のゲームソフトのタイトルについて文字数、文字種の時系列分析と、タグ付けによる部類と比較分析を行い、ゲームタイトルの経時的な長文化、またゲームタイトルは企業やターゲットごとに戦略的な特徴が生じていることを証明した[1]。山中らは、国会図書館サーチを利用して収集した約 2400 冊の一般書とベストセラーの日本語の書名を対象に、メインタイトルとサブタイトルを分け、全長やモダリティの分布、それぞれの字種の比率などから、書籍のジャンルごとに比較し、またベストセラーの書籍ではタイトルに年代による違いがあるのかを研究した[2]。また、深層学習を用いてタイトルをあらすじ文から自動生成するための研究では、大部らが論文執筆に慣れていない人に対して、的確な論文タイトルをアブストラクトから抽出した単語をもとに生成する手法を提案した[3]。

2.2 オンライン小説に関する研究

オンライン小説に関する研究はこれまでいくつかの試みがなされてきた。小川らは、オンライン小説投稿、閲覧サイトでは読者はその時に流行しているジャンルの小説を閲覧しがちであり、また作者も読まれやすい流行の小説を書きたがる傾向があるとし、読まれやすい流行のジャンルを知りたがるユーザに対して、既存小説群から過去の流行を分析し、将来の流行を予測することを目的とし、流行分析のために類似小説の抽出を行った[4]。堺らは、小説家になろうのメタデータを取得、分析を行い、ユーザが入力した単語の流行度を測定するツールを提案した[5]。また、小坂らはオンライン小説における問題点として、その作品数の多さから読者が好みの作品を探すことが困難であり、また人気の作品や最新の作品など限られた少数の作品に大多数の読者が集中して、良作が埋もれてしまうといった点をあげており、オンライン小説における隠れた良作を推薦するため、作品の類似度を示す S スコアと、作品の質を示す Q スコアの二つの指標を設定し、この二つの指標を作品自体の情報から予測することで推薦を行うシステムを提案した。また、小坂らは実際に二つの指標を設定し、機械学習モデルによる予測実験を行うことで、一定の精度で予測が可能だということを証明した[6]。

3章 小説家になろうと T5 について

3.1 オンライン小説投稿, 閲覧サイト「小説家になろう」について

「小説家になろう」は株式会社ヒナプロジェクトが提供するオンライン小説投稿, 閲覧サイトである。当該サイトはユーザがコンテンツを提供し, またユーザがそのコンテンツを利用する CGM(Consumer Generated Media)であり, ユーザが双方向に小説の投稿, 閲覧を行うことができる。12/9 現在で小説掲載数は 995,957 作品, 登録ユーザ数は 2,359,908 人であり, 日本最大級の投稿型小説サイトとなっている。^{*1} ユーザが当該サイトを開くと, まず総合ジャンルにおける日間ランキング 1 位~10 位の作品のタイトル, 作者名, ジャンルが表示される。また, 総合ジャンルの横に並ぶ「恋愛」, 「ファンタジー」などのそれぞれのジャンルの作品が同じく表示されるページに移動することができる。このように, 小説家になろうにおいては膨大な数の作品の中でも日間ランキング上位 10 作品というごく限られた作品しかトップページには掲載されず, またその限られた作品もタイトルや作者名などの情報しか知ることができないため, 作品の内容などが一目で分からないということとなる。これにより, 作品を投稿するユーザはユーザが一目で内容がわかるような, あらすじの役割を担ったタイトルを制作することが求められる。

3.2 小説家になろう特有のタイトルについて

小説家になろうには, 他の一般書籍に比べ, タイトル構造に特異性がある。以下に小説家になろう特有のタイトルを実在の作品タイトルを例としてその特異的な構造を説明する。

「小説家になろう」に含まれるタイトルの例

「婚約者様にはもうすぐ運命のヒロインが現れますが, せっかくなので暫定婚約者を漫喫します! ところでその呪い, 嫌われ悪女の私が解いちゃダメですか?」

(2022/11/15 時点で小説家になろう総合ランキング TOP10 より抜粋)

*1 小説家になろうトップページ <https://syosetu.com>

このタイトルの構造は以下のように分解できる.

「小説家になろう」のタイトルの構造の例

(1)婚約者様にはもうすぐ運命のヒロインが現れますが、せっかくなので暫定婚約者を漫喫します！

(2)ところでその呪い、

(3)嫌われ悪女の私が解いちゃダメですか？

(1)世界観、時代背景、大まかなストーリーを説明

(2)起承転結の「転」にあたる面白さのキモ

(3)立場や役職など主人公の説明

この構造により、ユーザはタイトルを見るだけで作品の大まかな内容を知ることができる。つまり、小説家になろうにおいて作品タイトルは長文で、内容を紹介するあらすじの役割を担っているため、このような特異性があるといえる。

3.3 T5(Text-To-Text Transfer Transformer)について

自然言語処理とは、AI(人工知能)に、人間の使用する言葉を学習させ、さまざまな分野に応用させる技術のことを言う。近年の目覚ましい科学技術の発展により、AIスピーカーや、機械翻訳などを初めとし、自然言語処理は私たちにとっても身近なものになり始めて久しい。

2021年9月に行われた、理系的発想力を問う文学賞、日経「星新一賞」(主催:日本経済新聞社)では、応募総数2,603編のうち114編がAIによって制作された作品であり、AIを一部利用した作品が大会初の入賞を果たした[7]。

また、2022年8月、米コロラド州で行われたファインアートコンテストで、画像生成AI「Midjourney」が生成した絵が1位をとったことでもAI技術の発展の目覚ましさが感じられる[8]。

このように、かつて人間だけが可能だと思われていた芸術の分野でさえも自然言語処理を活用した AI の発展により、我々の考える想定よりもはるかに高度なことが可能になっている。本論文では、そんな自然言語処理を自然言語処理モデル「T5(Text-To-Text Transfer Transformer)」を活用して実験を行った。

T5(Text-To-Text Transfer Transformer)とは、様々な自然言語処理タスクを入出力をテキストに固定し、転移学習を行わせることで可能にする自然言語処理モデルである。1つのモデルに対し、翻訳や要約など、様々なタスクを目的とした転移学習を行うことができるため、幅広い目的で使用するこのできる柔軟性を持っていることが特徴である。今回は、sonoisa 氏によって作られた日本語事前学習済みモデル*2を使用し、転移学習を行うことでタイトル生成を行うシステムを制作した。

*2 sonoisa 氏によるコード (実験 1 のシステム作成の参考コード)

<https://github.com/sonoisa/qiita-title-generation>

4 章 実験 1 (タイトル生成システムの制作と ROUGE 評価実験)

4.1 実験 1 の目的

実験 1 の目的は 3 章の 1-2 節で説明したような「小説家になろうユーザの目を引きやすいタイトル」を自然言語処理モデル T5 に転移学習を行うことで機械生成できることを明らかにすることである。また、機械生成されたタイトルの評価を調べるため、ROUGE による機械評価を行った。

4.2 実験 1 の方法

まず、「なろう小説 API」を利用した小説家になろう内作品データの収集を行った。なろう小説 API*³は「小説家になろう」に掲載されている小説メタデータを取得できる API である。この API の出力として小説名やあらすじなど、計 40 項目のデータを取得することが可能である。本研究では 2004 年 4 月 20 日から 2022 年 7 月 3 日までの期間に投稿された小説の中で、global_score(総合評価ポイント)の高いもの上位 10,000 件のメタデータを取得し、転移学習に用いた。取得したデータの構成を以下の表 4.1 に掲載する。

表 4.1 (なろう小説 API から取得したデータの構成)

パラメータ	内容
title	小説名
ncode	N コード
story	小説のあらすじ
gemre	ジャンル
general_firstup	初回掲載日 (YYYY-MM-DD)
general_lastup	最終掲載日 (YYYY-MM-DD)
global_point	総合評価ポイント
title_len	タイトル文字数

*³ なろう小説 API トップページ <https://dev.syosetu.com/man/api/>

これらメタデータを日本語事前学習済み T5 に転移学習させる。転移学習の際には作品データのあらすじとタイトルを使用し、このようなあらすじの時はこのようなタイトルというようなパターンを T5 に学習させる。この転移学習により、あらすじを入力すればタイトルを出力する AI が完成する。そのモデルを使い 10,000 件のあらすじからタイトルを生成する。T5 の転移学習には、2004 年 4 月 20 日から 2022 年 7 月 3 日までの期間に小説家になろうサイトに投稿された小説の中で、global_score(総合評価ポイント)の高いもの上位 10,000 件のメタデータを転移学習に用いた。エポック数は 10 に設定し、転移学習を行った。

次に、生成したタイトルを ROUGE で評価する。ROUGE とはテキスト要約の評価の際、一般的に用いられる機械評価の指標である。これは大まかに、人間の作成した要約と、システムが作成した要約との一致度を測る指標である。今回実験で使用したものはそれぞれ「ROUGE-1」, 「ROUGE-2」, 「ROUGE-L」と呼ばれるものである。

ROUGE-1 は文章を単語に分け、互いの文中にどれだけその単語が一致しているかで類似度を評価する指標である。ROUGE-2 は文章を単語に分け、隣り合った単語 2 つをひとかたまりに捉え、その単語の固まりの一致度から類似度を測定する指標である。ROUGE-2 では隣り合った単語をひと固まりとして評価するため、例えば、「おいしそうなおにぎり」と「おにぎりはおいしそう」の一致度は低くなってしまふ。この場合、ROUGE-1 では単語一つ一つの類似度が高いため高いスコアが出るが、「私は彼に傘を貸した」という文と「彼は私に傘を貸した」という文の場合では、意味がそれぞれ逆の文になるにも関わらず、単語が一致しているため、ROUGE-1 のスコアが高くなってしまふ。ROUGE-L は、単語が連続しているもののみ、その一致度を測る指標であるため、前者の ROUGE-1, ROUGE-2 よりも文脈を考慮した評価ができるという指標である。2 つの文章間で一致する最長の長さの連なりを評価すると言ひ換えられる。ただ、最長の単語の連なり以外は評価の考慮に入らない、という欠点がある。3 章で説明したとおり、小説家になろうにおけるタイトルは、一般書籍と少し異なり、あらすじ文を短く要約したものだととらえられる。そのため、作者である人間が作成したタイトルと、入力を元のあらすじにし、機械生成されたタイトルの一致度を測る目的で今回この ROUGE を使用し、評価実験を行った。

実験の例として、例えば、人手のタイトルが「スキル『屋台』で異世界生活食料生成スキルを手に入れたので、商会を立ち上げようと思います」であり、システム生成のタイトルが「食料生成スキルを手に入れたので、異世界で商会を立ち上げようと思います」の場合、

上記 2 タイトルを ROUGE-1, ROUGE-2, ROUGE-I で評価した場合, それぞれの類似度は 0.9268, 0.7692, 0.7804 となる.(小数点以下第五位切り捨て)

このように人手のタイトルとシステム生成のタイトルの類似度評価を, 学習データとして用いた 10,000 件のタイトルに対して行った. 表 4.2 に 10,000 件分の類似度評価の平均値を記載する.

表 4.2 (実験 2 で使用したタイトルの ROUGE 評価平均値, 小数点以下第五位切り捨て)

	ROUGE-1	ROUGE-2	ROUGE-I
平均値	0.3053	0.1694	0.2839

4.3 実験 1 の結果

実験の結果, 自然言語処理モデル T5 をベースになろう小説 API から取得したデータを転移学習させることで, あらすじを入力すると「小説家になろうらしい」タイトルを出力するシステムが完成した. これにより, 実験 1 の目的であった「小説家になろうにおいて評価の高い作品を学習データとすることで, ユーザが入力したあらすじからユーザの目を引きやすいタイトルを自動生成できることを明らかにする」が達成された.

また, 転移学習に使用した 10,000 件の作品のあらすじを入力とし, 出力されたタイトルと, 元のタイトルとの類似度を測定するため, ROUGE による評価を行い, その平均値を出した. その結果, 表 4.2 より, ROUGE-1 が 0.3053, ROUGE-2 が 0.1694, ROUGE-I が 0.2839 となった. 値の最高値が 1 であり, いずれの指標の数値も高すぎず低すぎずと言った結果となった.

5章 実験2（作品にふさわしいタイトルの判断基準の調査実験）

5.1 実験2の目的

実験1では小説家になろうでユーザ評価の高い作品を学習したシステムを制作し、それにより生成されたタイトルを ROUGE による機械評価を行った。実験2, 実験3では、「自動生成されたタイトルと人間（作者）が作成したタイトルを比較して、印象評価の実験をすることで、タイトル生成手法の有効性を示す」ことを目的として実験を行った。実験2では、まず、機械が生成したタイトルと人間が作成したタイトルの間に持つ印象の違いを明らかにするため、20代の被験者7名に対する評価調査を行った。

5.2 実験2の方法

20代の被験者7名に対し、作品のあらすじを見せ、そのあと、機械生成されたタイトルと元々の人手のタイトルを並べて見せる。2つのタイトルの内、どちらのタイトルが作品にふさわしいかを回答させ、その判断基準について記述させた。また、本調査は Google アンケートを利用して行った。一例として、被験者のうち一人の回答とアンケート文を下記に記す。

また、実験に使用した作品のあらすじ、人手のタイトル(元のタイトル)、システムによって機械生成されたタイトルを表5.1にまとめる。また、それぞれのタイトルの ROUGE での評価値を表5.2にまとめる。

表 5.1（実験2の調査で使用したタイトル10件）

	人手のタイトル	システム生成のタイトル
1	クラス「無職」の英雄譚～公爵家を追放されたが、実は殴っただけでスキルを獲得できるとわかり、大陸一の英雄に上り詰める～	クラス「無職」になってしまった公爵家を追放されたが、実は殴っただけでスキルを獲得できるので大陸一の英雄に上り詰める
2	戦国小町苦勞譚	戦国時代にタイムスリップした少女

3	ふよめし！	魔法大国の公爵令嬢は冒険者ギルド所属の密偵の弟子になる
4	エリザベス・ナザントという令嬢（連載版）	エリザベス・ナザントという令嬢の連載版
5	妾子王子の試練	妾の子
6	政略結婚の行方	悪役令嬢の政略結婚
7	世界再生 GAME	デスゲーム
8	小さな殿下と私	【書籍化&コミカライズ】悪役令嬢は王太子に溺愛される
9	虐げていた娘が精霊の愛し子でした～みんなしんじゃえ～	精霊の愛し子
10	生け贄のお嬢さんが、生き埋めのドラゴンさんに会いましたとさ	悪しきドラゴンの生け贄にされました。

表 5.2（実験 2 で使用したタイトルの ROUGE 評価値. 小数点以下第五位切り捨て）

	ROUGE-1	ROUGE-2	ROUGE-1
1	0.8070	0.7272	0.8070
2	0.1666	0.0000	0.1666
3	0.0000	0.0000	0.0000
4	0.9230	0.7272	0.9230
5	0.6666	0.0000	0.6666
6	0.6666	0.2857	0.4444
7	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0000
9	0.3529	0.2666	0.3529
10	0.6000	0.0000	0.3000

5.3 実験 2 の結果

実験の結果, 得られた回答データ 70 件の中から一部を抜粋し, 回答の内容別に以下の表 5.3 にまとめる.

表 5.3 (実験 2 で回答させたタイトルがあらすじにふさわしいかを判断する基準の抜粋)

1. 文法の正しさに関する回答
<u>日本語がおかしいから。</u>
B は <u>不自然</u> だから
タイトル A の <u>日本語がおかしい</u> から。
2. あらすじの一致度に関する回答
<u>ストーリーの序盤や主人公が持つ特徴的な能力が分かりやすく記述</u> されているため。
<u>タイトルで内容がわかる</u> から
<u>「誰が」、「何をする」のかが明確にわかる</u> から。
あらすじに悪役令嬢モノであるという情報と王太子に溺愛されるという情報がないため、 <u>タイトルとあらすじが乖離している</u> など感じたため
3. タイトルのひねり, 独自性に関する回答
A は、 <u>キャッチー</u> じゃないから
<u>テンポ</u> がいい。
A のほうが <u>先が気になるようなタイトルの付け方が</u> されていると思ったから
どんな GAME が展開されるのだろう、と <u>期待を持つことができる</u> から。
精霊の愛し子が報復するという <u>内容が伝わりやすい</u> かなと考えたため
正直どっちもいいが、文のテンポ良さや <u>遊び心</u> が A の方に感じられたから。
上記の理由と同様に、B よりも A の方が <u>遊びがあつていい</u> など感じたため

表 5.3 から, 人間が人手のタイトルと機械生成されたタイトルを見分ける際, 主に「文法の正しさ」, 「あらすじの一致度」, 「独自性(文章の単調さ)」が主な判断基準となっていることが明らかになった.

6章 実験3（自動生成されたタイトルの印象評価実験）

6.1 実験の目的

実験2で人手による評価を調べるため、20代の被験者7名に対し、作品のあらすじと、人手のタイトル、機械のタイトルを2つ表示し、どちらのタイトルが作品にふさわしいかを回答させる評価調査を行った。その結果、人間が小説のタイトルが作品に合うか否かを判断する主な基準として「文法の正しさ」、「あらすじとの一致度」、「独自性」の3つが挙げられることが分かった。そこで、本実験ではさらに詳細に人手のタイトルと機械生成のタイトルの印象の違いを明らかにするため、前者の3つの基準で比較評価実験を行った。

6.2 実験の方法

20代の被験者12名に対し、作品のあらすじを見せ、そのあと、機械生成されたタイトルと元々の人手のタイトルを並べて見せる。2つのタイトルを、「文法の正しさ」、「あらすじとの一致度」、「独自性」の3つの基準で「かなり良い」、「やや良い」、「やや悪い」、「かなり悪い」の4段階でそれぞれ評価させた。また、今回実験で使用したタイトルは10,000件の中からランダムに抽出したものを使用した。タイトルの偏りを避けるため、被験者を1グループ6名で2つグループ分けし、グループ別にそれぞれ抽出した別のタイトルを使用した。本調査はGoogleアンケートを利用して行った。一例として、アンケートの内容と回答を一つ取り上げ、下記に記す。

あらすじ：当代最高の勇者すらノーダメージ撃破の最強魔王——俺は身分を隠して勇者学校へと入学する。そこで、「理力ゼロ？ カスだな！」エリート貴族に笑われてしまう。勇者の適性ゼロ。問題ないね。美貌の副官で墮天使のルシフェルとともに圧倒的な基礎ステータスで無双無双無双！やがて俺は成り上がり、英雄と讃えられる。だけど……女生徒からは「魔王を倒すために弟子にしてほしい！」王さまからは「君は国の希望。打倒魔王、頼んだぞ！」なんて勘違いが！……俺が魔王なんだが？

タイトル A： オレ魔王、勇者学校で適性ゼロ判定されたけど最強ステータスで無双してたら成り上がった ～お前強いから魔王倒せと言われても困るんだけど～

タイトル B： 俺が魔王なんだが？

上記のように、提示されたあらすじに対し人手のタイトルとシステム生成のタイトルを表示し、それぞれに対し、実験 2 で判明した 3 つの基準で 4 段階評価をさせた。ちなみに上記の場合だと、タイトル A が人手のタイトルで、タイトル B がシステム生成のタイトルだが、被験者にはどちらがそのことは伝えず評価実験を行った。

また、実験に使用した作品のあらすじ、人手のタイトル(元のタイトル)、システムによって機械生成されたタイトル、それぞれの ROUGE 評価値を表 6.1 にまとめる。なお、実験ではグループ 1 に表中の 1~10 のタイトルを、グループ 2 に 11~20 のタイトルを使用した。

表 6.1 (実験 3 で使用したタイトル 20 件)

	人手のタイトル	システム生成のタイトル
1	婚約破棄されて退場します。	悪役令嬢の婚約者
2	転生者は裏切り者を許さない	悪役令嬢に転生した私
3	逃げ出した子爵令嬢はブルーに捕らわれる	公爵令嬢は婚約者様を解放してあげる
4	最果てのパラディン	古風な英雄譚
5	スマホのようなユニークスキルで異世界を生きる～便利なスマホ能力で最強に～	スマホのようなユニークスキルで異世界を生きていく
6	【旧】聖弾の射手	加護板に『』と『』があるのは何故だ？
7	黒龍の強くてニューゲーム	黒龍と黒龍
8	神皇勇者の英雄譚《りょこうたん》	異世界転生チートで異世界を楽しめ

9	私は殿下が本命を見つけるまでの繋ぎの婚約者候補なのに、本命を探している素振りがなくて困っています	【短編】 婚約者候補に絞られた公爵令嬢は、婚約者候補を辞退する
10	果てない恋心の行く末は	狼に求婚された俺は、妹に求恋される
11	異世界娼館の支配人 ～夜嘶百花～	異世界娼館の支配人
12	悪役に転生したので男装することにした	悪役令嬢は男装することにした
13	星見の聖女は死にたがり	星見の聖女は死にたくない
14	オレ魔王、勇者学校で適性ゼロ判定されたけど最強ステータスで無双してたら成り上がった ～お前強いから魔王倒せと言われても困るんだけど～	俺が魔王なんだが？
15	婚約の条件	婚約破棄でも継続でもどちらでも構いません
16	揺るぎない王子 ～あなたについているものが視えるんです～	婚約破棄ですか？
17	エルフ公爵は呪われ令嬢をイヤイヤ娶る	大の猫好き公爵の苦悩
18	ハズレ勇者の奮闘記	勇者サマにボコられたので、異世界から抜け出してみる
19	リユーリ・ベルとめんどくさい人々	【連載版】 そんなことよりランチしてた
20	子どもが書いた夏休みの作文が、ファンタジーすぎる件	夏休みの作文

表 6.2 (実験 3 で使用したタイトル 20 件の ROUGE 評価値. 小数点以下第五位切り捨て)

	ROUGE-1	ROUGE-2	ROUGE-1
1	0.1666	0.0000	0.1666
2	0.1666	0.0000	0.1666
3	0.2352	0.1333	0.2352

4	0.0000	0.0000	0.0000
5	0.5925	0.5600	0.5925
6	0.1333	0.0000	0.1333
7	0.3333	0.2000	0.3333
8	0.0000	0.0000	0.0000
9	0.1904	0.0500	0.1904
10	0.1052	0.0000	0.1052
11	0.7692	0.7272	0.7692
12	0.6315	0.4705	0.6315
13	0.7142	0.6666	0.7142
14	0.1463	0.0512	0.1463
15	0.1428	0.0000	0.1428
16	0.1333	0.0000	0.1333
17	0.1250	0.0000	0.1250
18	0.1176	0.0000	0.1176
19	0.0000	0.0000	0.0000
20	0.4285	0.3333	0.4285

6.3 実験の結果

実験の結果, 得られたデータを以下の表にまとめる. 表 6.3, 表 6.4 がグループ 1 の結果, 表 6.5, 表 6.6 がグループ 2 の結果となっている.

表 6.3 (グループ 1, 人間の作成したタイトル 10 件を対象とした評価)

	かなり良い	やや良い	やや悪い	悪い
文法の正しさ	30	12	7	1
あらすじ一致度	22	20	7	1
独自性	16	21	13	0
総合	68	53	27	2

表 6.4 (グループ 1, システムの作成したタイトル 10 件を対象とした評価)

	かなり良い	やや良い	やや悪い	悪い
文法の正しさ	20	15	14	1
あらすじ一致度	5	13	14	15
独自性	4	21	20	6
総合	29	17	48	22

表 6.3, 表 6.4 より, グループ 1 ではすべての基準で人間の作成したタイトルの評価値が高くなった. 特に, 独自性の基準では, 人手のタイトルの「かなり良い」の評価値が 16 件, システム生成のタイトルが 4 件とより顕著な差が生まれた. ただ, 文法の正しさの基準では, ポジティブな評価 (かなり良いとやや良いの合計値) が人手のタイトルが 42 件, システム生成のタイトルが 35 件と, 他の基準よりも比較的近い値となった.

表 6.5 (グループ 2, 人間の作成したタイトル 10 件を対象とした評価)

	かなり良い	やや良い	やや悪い	悪い
文法の正しさ	10	43	6	1
あらすじ一致度	24	25	10	1
独自性	14	28	16	2
総合	48	96	32	4

表 6.6 (グループ 2, システムの作成したタイトル 10 件を対象とした評価)

	かなり良い	やや良い	やや悪い	悪い
文法の正しさ	13	41	6	0
あらすじ一致度	17	27	14	2
独自性	10	27	20	3
総合	40	95	40	5

表 6.5, 表 6.6 より, 文法の正しさの基準では, 人手のタイトルのポジティブな評価 (かなり良いとやや良いの合計値) が「53」, システム生成のタイトルが「54」となり, あらすじの一致度の基準だと, 人手のタイトルが「49」, 機械生成のタイトルが「44」, 独自性の基準では人手のタイトルが「42」, 機械生成のタイトルが「37」となった. 評価値の総合の内, ポジティブな評価が, 人手のタイトルが「144」, 機械生成のタイトルが「135」となり, 人手のタイトルの評価が若干高いものの, 両方のタイトルにおいてとても近い評価値となった.

また, グループ 1, グループ 2 で回答させたタイトルのうち, ポジティブな評価 (かなり良いとやや良いを足した数値) が人手のタイトルより高くなったシステム生成のタイトルは 20 タイトル中以下の表 6.7 の 5 タイトルとなった. また, 表 6.8 より, タイトル 14, 15, 17 に関しては ROUGE 評価値が 0~0.17 程度と類似度が低く, 元の人手のタイトルとは似通っていないタイトルにもかかわらず, 高い評価を得た. この結果から, 本システムを活用したタイトル生成の有効性が明らかになり, 実験 2, 実験 3 の目的であった「自動生成されたタイトルと人間が作成したタイトルを比較して, 印象評価の実験をすることで, タイトル生成手法の有効性を示す」を達成した.

表 6.7 (人手のタイトルより評価が高かったシステム生成のタイトル 5 件)

	人手のタイトル	システム生成のタイトル
3	逃げ出した子爵令嬢はブルーに捕らわれる	公爵令嬢は婚約者様を解放してあげる
12	悪役に転生したので男装することにした	悪役令嬢は男装することにした
14	オレ魔王、勇者学校で適性ゼロ判定されたけど最強ステータスで無双してたら成り上がった ~お前強いから魔王倒せと言われても困るんだけど~	俺が魔王なんだが?
15	婚約の条件	婚約破棄でも継続でもどちらでも構いません

17	エルフ公爵は呪われ令嬢をイヤイヤ娶る	大の猫好き公爵の苦悩
----	--------------------	------------

表 6.8 (人手のタイトルより評価が高かったシステム生成のタイトルの ROUGE 評価値.
小数点以下第五位切り捨て)

	ROUGE-1	ROUGE-2	ROUGE-1
3	0.2352	0.1333	0.2352
12	0.6315	0.4705	0.6315
14	0.1463	0.0512	0.1463
15	0.1428	0.0000	0.1428
17	0.1250	0.0000	0.1250

7章 まとめ

本研究では、近年のオンライン小説投稿、閲覧サービスの台頭により生じた「より簡単に、人気が出やすいオンライン小説のタイトルを作りたい」という需要を、自然言語処理モデル T5 を転移学習することで、あらすじを入力するだけでタイトルを生成するシステムを作ることで満たすことができることを明らかにすることを目的として研究を行った。システム制作後は、被験者に対し、あらすじにふさわしいタイトルを判断する基準をアンケート調査で調べ、その結果をもとに、人手のタイトルと機械生成のタイトルの比較評価実験を行い、両者のタイトルの印象の違いを調べた。

実験 1 では、自然言語処理モデル T5 に対し、なろう小説 API から取得したデータをもとに転移学習を行い、あらすじを入力するだけで、タイトルを機械生成するシステムを制作した。この結果により、本研究の目的であった、「オンライン小説において評価の高い作品を学習データとすることで、ユーザが入力したあらすじからユーザの目を引きやすいタイトルを自動生成できることを明らかにする」という目的を達成した。

実験 2, 実験 3 では、「自動生成されたタイトルと人間が作成したタイトルを比較して、印象評価の実験をすることで、タイトル生成手法の有効性を示す」ことを目的とした実験を行った。実験 2 では、人間がタイトルを「作品にふさわしい」と感じる基準を調べるために、被験者 7 名に対し、提示されたあらすじと人手とシステム生成の 2 つのタイトルを見せ、どちらのタイトルがふさわしいか、またその判断基準に対して記述させる実験を行った。その結果、人間がタイトルをあらすじにふさわしいと判断する大きな基準として、「文法の正しさ(日本語としての違和感のなさ)」、「作品のあらすじとの一致度」、「独自性(ひねりの有無)」の 3 つの基準があることが分かった。

この結果を踏まえ、実験 3 では、被験者 12 名を 2 グループに分け、提示されたあらすじに対して、人手のタイトルと機械生成のタイトル 2 つのタイトルを見せ、それぞれを 3 つの基準で 4 段階評価させた。その結果、グループ 1 では、人手のタイトルの評価が全体的に高かった。だが、文法の正しさの基準では、ポジティブな評価が、人手のタイトルが「42」、システム生成のタイトルが「35」と両者の差が小さくなった。グループ 2 では、総合評価のうち、ポジティブな評価(かなり良い+やや良い)では、人手のタイトルが「144」、システム生成のタイトルが「135」となり、人手のタイトルとシステム生成のタイトルの間の評価にはほぼ違いがないという結果となった。また、実験 3 を通して、人手のタイトルより評価が高かった機械生成のタイトルが 20 タイトル中 5 タイトルとなり、本タイトル生成システムの有効性が証明された。この実験の結果から、「自動生成されたタイトルと人間が作成したタイトルを比較して、印象評価の実験をすることで、タイトル生成手法の有効性を示す」という目的を達成した。

参考文献

- [1] 福田一史. (2017). 家庭用ゲームソフトのタイトルに関する研究: DB を活用した文字数・文字種の観点からみたマクロ的分析. アート・リサーチ, 17, 29-44.
- [2] 山中信彦. (2015). 書名の言語学—国立国会図書館サーチに基づいて—. 計量国語学, 30(1), 14-31.
- [3] 大部達也. (2018). 深層学習に基づく論文タイトル生成手法とその応用. 第 80 回全国大会講演論文集, 2018(1), 561 - 562.
- [4] 小川明水. (2021). オンライン小説の傾向分析のための創作の連鎖関係の分析. 火の国情報シンポジウム論文集, 1-4.
- [5] 堺雄之助. (2020). オンライン小説の流行語抽出. 第 82 回全国大会講演論文集, 2020(1), pp.481-482.
- [6] 小坂直輝. (2020). 隠れた良作の発掘を助ける Web 小説推薦システムの構成と評価. 第 34 回全国大会(2020), 1-4.
- [7] 日経「星新一賞」グランプリに「リンネウス」など 3 作 (2022/2/18)
<https://prtimes.jp/main/html/rd/p/000000421.000011115.html>
(最終閲覧日 2022 年 12 月 28 日)
- [8] 画像生成 AI「Midjourney」の絵が米国の美術品評会で 1 位に 優勝者「物議を醸すことは分かっていた」 (2022/9/1)
<https://www.itmedia.co.jp/news/spv/2209/01/news148.html>
(最終閲覧日 2022 年 12 月 28 日)