

オンライン小説における
効果的な検索タグ推薦手法に関する研究

筑波大学

人間総合科学学術院人間総合科学研究群

情報学学位プログラム

2023年3月

山崎 睦月

オンライン小説における
効果的な検索タグ推薦手法に関する研究
A Study on Effective-Tag Recommendation Methods
in Web Novel Retrieval

氏名：山崎 睦月
Mutsuki Yamazaki

オンライン小説のCGMサイトである小説投稿サイトでは、日々、小説が投稿されるとともに、多くのユーザに依って検索・閲覧がされている。日本で最大級の小説投稿サイトにおいては、登録小説数が100万件に達しようとしており、閲覧ユーザが自分の趣向に合致した小説を探すことが難しくなっている。多くの小説投稿サイトでは、投稿小説にタグを設定することができる。タグの設定によって、小説のタイトルやあらすじだけでは表すことが難しい、新たな情報を小説に付与することができている。そのため、小説に付与されたタグを検索に用いることは、小説の検索を効果的に行えると考えられる。

本研究では、1つの小説を読了後に、次に読みたくなる小説を検索するために効果的な検索タグを推薦する手法を提案する。提案手法では、あらかじめ小説本文と小説に付与されているタグを収集しておく。個々の小説の特徴量は、小説本文にDoc2Vecを適用し算出する。一方、タグの特徴量は、そのタグが付与されている1つあるいは複数の小説の特徴量を合成して算出する。本研究における検索に効果的なタグとは、入力された小説の内容と類似した特徴を有するタグ、推薦されたタグ同士でタグの表す情報に異なる特徴を有するタグ、複数のタグで絞り込みができるタグの3つの要件を満たすタグとした。1つ目は、入力された小説とコサイン類似度を算出することで、2つ目は、タグ同士で特徴量を比較し特徴量の差が大きいタグの組を求めることで実現する。3つ目の要件は、出現頻度を用いて情報量を算出し、絞り込みに適さない情報量を持つタグを除外することとした。

提案手法を実運用されている日本国内最大規模の小説投稿サイトのデータに適用・評価し、提案手法の有効性を検証した。その結果、限定された範囲の小説に関してではあるが、効果的な検索タグを推薦できることが確認できた。

主研究指導教員：佐藤 哲司
副研究指導教員：松村 敦

目次

第1章	はじめに	1
1.1	研究背景	1
1.2	研究目的	2
1.3	本論文の構成	2
第2章	関連研究	3
第3章	小説投稿サイト	5
3.1	なろうサイトのデータ構造	5
3.2	タグの特性	6
第4章	提案手法	20
4.1	概要	20
4.2	タグの情報量	21
4.3	タグの特徴量	22
4.4	タグ推薦	22
4.4.1	手法A	23
4.4.2	手法B	23
第5章	実験・評価	26
5.1	データセット	26
5.1.1	タグ	26
5.1.2	小説本文	27
5.2	ベクトル生成	27
5.2.1	本文のベクトル生成	27
5.2.2	タグのベクトル生成	29
5.3	タグの推薦実験	29
5.3.1	手法A	29
5.3.2	手法B	30
5.4	考察	35
第6章	おわりに	36
	謝辞	37
	参考文献	38

目次

3.1	タグの出現頻度分布 (全体)	8
3.2	タグの出現頻度分布 (分割拡大)	9
4.1	推薦タグ T_i の図解	25
5.1	1つの小説あたりに設定されたタグ数	28
5.2	「小説家になろう」内のタグの出現頻度分布	28
5.3	小説 N7805EI と手法 A パターン 1 で推薦されたタグの距離	31
5.4	小説 N7805EI と手法 A パターン 2 で推薦されたタグの距離	31
5.5	小説 N7805EI と手法 B パターン $1\alpha = 0.1$ で推薦されたタグの距離	32
5.6	小説 N7805EI と手法 B パターン $1\alpha = 0.5$ で推薦されたタグの距離	32
5.7	小説 N7805EI と手法 B パターン $1\alpha = 0.9$ で推薦されたタグの距離	33
5.8	小説 N7805EI と手法 B パターン $2\alpha = 0.1$ で推薦されたタグの距離	33
5.9	小説 N7805EI と手法 B パターン $2\alpha = 0.5$ で推薦されたタグの距離	34
5.10	小説 N7805EI と手法 B パターン $2\alpha = 0.9$ で推薦されたタグの距離	34

表目次

3.1	小説家になろうサイトの規模	6
3.2	公式キーワード一覧	7
3.3	おすすめキーワード一覧	7
3.4	登録必須キーワード一覧	7
3.5	なろう小説 API の概要	8
3.6	タグ間の編集距離: タグ長 3 文字	11
3.7	タグ間の編集距離: タグ長 4 文字	12
3.8	タグ間の編集距離: タグ長 5 文字	13
3.9	タグ間の編集距離: タグ長 6 文字	14
3.10	タグ間の編集距離: タグ長 7 文字	15
3.11	タグ間の編集距離: タグ長 8 文字	16
3.12	タグ間の編集距離: タグ長 9 文字	17
3.13	タグ間の編集距離: タグ長 10~11 文字	18
3.14	タグ間の編集距離: タグ長 12~21 文字	19
5.1	小説ベクトルと手法 A の推薦タグベクトルの Cos 類似度平均値	29
5.2	手法 A の推薦タグと著作者タグの比較	30
5.3	小説ベクトルと手法 B 推薦タグベクトルの Cos 類似度平均値	30
5.4	手法 B の推薦タグと著作者タグの比較	35

第1章 はじめに

1.1 研究背景

小説投稿サイトとは、オンライン小説を専門に扱う CGM(Consumer Generated Media) サイトである。小説投稿サイトの登場・発展によって、誰もが容易にインターネット上に自身の書いた作品を投稿できるようになった。投稿された小説はウェブサイトの利用が可能な状態であれば、いつでも読むことができ、多くの人々がサイトを利用している。ヒナプロジェクト社が運営している日本の小説投稿サイトの一つである「小説家になろう」¹にはおよそ8か月間の中に約29,000件の小説が投稿され、登録ユーザ数は約114,000人増加している(2022年4月8日では657,825件だった投稿数が同年12月6日には995,423件に増加し、登録ユーザは2,245,258人から2,358,933人へと増加している)。登録ユーザは小説の投稿、ブックマーク機能などのサイト利用が可能である。登録ユーザ以外でも、オンライン小説を読むことはできるため、より多くの人々がサイトを利用していると推測される。

小説投稿サイト上に投稿されたオンライン小説が増加することは、サイトのさらなる発展や利用ユーザの獲得などのメリットがあると考えられる。他方で、オンライン小説の増加に伴い、サイト上の小説を利用するユーザ、読者は非常に多くの小説の中から自分の趣味嗜好に適合した小説を探索する必要があるという困難な問題に直面している。

小説投稿サイトでは、読者が小説を探しやすくするために様々な工夫がされている。例えば、小説のジャンルをわかりやすく掲示する方法や、サイト内の小説全体やジャンル別でPV(閲覧数)や評価などをもとに、週間、月間、四半期など期間に分けてランキングを作成し掲示する方法などがあげられる。これらの方法は、ある程度読者のオンライン小説探索の支援になると考えられるが、読者の小説探索の支援には十分であるとは言えない。前者の方法では、ジャンルだけを示すだけではジャンル内での小説の絞り込みが更に必要であり、問題解決には至らない。後者のランキングから小説を探す方法では、ジャンル別で人気のあるもしくは評価されている小説を探すことには適している。しかし、投稿されて間もない小説や閲覧数や評価の少ない小説は当然探索の候補から外れてしまう。また、期間ごとにランキングは分けられているものの、同じ小説が上位にランクインしているケースも多くみられる。よって、読者が自分の趣味嗜好にあった小説を探す方法にはランキングの手法は十分ではないと考えられる。

小説投稿サイトの多くでは投稿した小説にタグを設定する機能がある。タグは、作品の持つテーマ、舞台、世界観などの特徴や作品に登場する人物や要素などについての情報を小説に明示的に付与するものである。タグには、サイト側が用意したタグ(システムタグ)と利用者(多くの場合は著作者)が自由に設定できるタグ(自由タグ)の2種類がある。小説投稿サイトで、キーワード検索をするときにその対象となるものは、主にタイトル、あらすじ、タグである。タイトルはやあらすじは、読者に向けて興味を引く内容にするための工夫や特に伝えたい部分や導入的な内容が多く、小説の特徴的な要素を示すことは難しい。タグを設定することは、小説のタイトルとあらすじで足りない情報を補完することができる

¹<https://syosetsu.com/>

る。そのため、タグは読者が作品を検索する際の大きな手掛かりとなる。特に自由タグは、著作者の手によって投稿小説の内容をより詳細に反映した記述が可能のため、自由タグを活用して小説の探索を行うことは、読者の嗜好に沿った探索に効果的であると考えられる。

実際に小説投稿サイトでも、このタグを活用した小説の検索は推奨されている。「小説家になろう」では、人気のキーワード（上記のタグと同等の性質を持つ）として過去 60 日のうちに投稿小説に設定されたキーワードを掲示し、そこから小説を探せるようになっている。キーワードをクリックすることでそのキーワードを含む小説の候補に飛び、ユーザが感覚的に小説を探索しやすいような作りとなっている。一方で、一部の表示されるタグには「ごまあ」や「もう遅い」などといったサイトのライトユーザでは理解が難しい、一般的な語ではないタグも散見される。

1.2 研究目的

本研究では、小説投稿サイトを利用する読者の小説検索支援を目的として、読者が読了した小説の本文をもとに、その小説と類似した要素を持った小説を検索するために効果的に機能するタグを推薦する手法について提案する。推薦するタグは、自由タグからのみ選出する。これは、システムタグは、サイト側が用意したタグであるため、非常に一般的な語であると考えられ、読者に推薦する必要性が少ないためである。

推薦の際に、検索に効果的なタグは、複数推薦する。これは、1つのタグでは、検索するときの絞り込みが困難である可能性があると考えられるためである。また、複数のタグを推薦することで、読者がその読了した小説の気に入った特徴に該当するタグのみを選出して活用するケースにも対応できる。

1.3 本論文の構成

本論文の構成について説明する。まず2章で、オンライン小説に関する研究やタグに関係した研究などの関連研究について説明し、本研究の位置づけを明確にする。次に3章では、本研究で取り扱う小説投稿サイトの構造や特性、タグの特性について説明を行う。続いて4章では、検索に効果的なタグを推薦する手法について概要を表し、各節でより詳細に説明する。5章では、4章で説明をした提案手法を用いて、タグの推薦実験を行い、実験の評価を行う。最後に6章で、本研究全体についてまとめる。

第2章 関連研究

小説投稿サイトは、ユーザが趣味で小説を投稿する場にとどまらず、投稿小説が一般に紙媒体や電子媒体で書籍化することも一般的になってきており、商業的価値も高くなってきている。多くの人々が利用する小説投稿サイトやオンライン小説について多くの研究が行われている。オンライン小説を推薦する研究の1つとして、高田 [1] らの研究がある。オンライン小説の投稿数が増加することによって発生する問題に対して、高田らの研究では、読者に小説を推薦するというアプローチで問題解決を図っている。オンライン小説には、評価数やブックマーク数が少ない小説が多く存在することや著作の少ない投稿者や新規投稿者が多いこと、一般的に表紙や大きさなどの形状的要因が存在しないことから、同著者の著作推薦、協調フィルタリングを用いた手法、表紙からの推薦など既存の手法を用いることが難しいことを指摘している。そこで、高田らは立ち読みという選択手法に注目している。立ち読みを行う時、人々が考慮する特徴量として文章の言い回しと語義の2つを合わせて、文体と定義した。そして、文体を新たな小説推薦の指標として、ジャンルと組み合わせた推薦手法を提案している。高田らは、投稿者の支援と読者の満足度の2つの観点から評価を行っている。投稿者の支援としては、数の少ない小説でも推薦可能なことを示した。読者満足度の観点からは、異なる著者でも文体類似度の高い小説の推薦は、同著者の著作を推薦されたときの満足度と同程度の高い満足度を獲得できることを示している。

推薦以外にも、オンライン小説を対象としたさまざまな研究が行われている。飯田 [2] らは、オンライン小説のあらすじを Doc2Vec を用いてベクトル化し、Cos 類似度の総和を求めることによって、投稿されたオンライン小説の多様性について定量的に評価を行った。そこで、月別新規小説投稿数が増加するにしたがって、Cos 類似度の総和の平均値が増加していることから、オンライン小説の多様性が減少していることが定量的に明らかにしている。

清水 [3] らの研究では、現状の小説投稿サイトが提供するランキングでは現在の人気作品しか示せないことを問題点にあげ、読者のつけたブックマークのリンク構造を利用して新たなランキング手法を提案し、将来ランキングに入る人気作品の予測を行っている。清水らの実験では、複数のジャンルで2か月後に人気ランキング上位になる小説の推定に成功している。

伊藤 [4] らの研究では、投稿されているオンラインの中には章や節で分けられていないものも多く、紙の書籍などに比べて、読者が一度に読む量を自身で決定することが難しいことを問題点にあげている。そこで、オンライン小説の本文を1つのテキストデータとして扱い、それを意味のある位置で区切る、段落分割の手法の提案を行っている。

実崎 [5] らの研究では、オンライン小説の人気度について、読者のつけたブックマークから人気度の推定を行おうとしている。

オンライン小説以外を対象としたタグの推薦には、井上 [6] らの研究や、Wang [7] らの研究がある。井上らは、SNSの投稿を対象に、ハッシュタグが付与された投稿をひとつの文書として扱い、文書の TF-IDF を求めクラスタリングし、推薦する投稿文書と TF-IDF ベクトルの比較を行い、ハッシュタグのクラスタの推薦を行っている。Wang らの研究では、SNS 上で投稿される画像の類似度だけでなく、投稿の人気度と投稿ユーザの人気度、タグの人気

度を考慮し、タグランキングを作成した。作成したタグランキングをもとに推薦を行うことで、従来のタグ推薦よりも閲覧数を伸ばす推薦が可能であるという結果を示している。

本研究は、ユーザが読了した小説投稿サイト上のオンライン小説の本文の特徴量とタグの特徴量を類似度比較することで、ユーザが読了したオンライン小説と類似した小説を検索するときに効果的なタグの推薦を行う。オンライン小説上に、上記の高田らのような小説を推薦する研究はあったが、投稿者にタグを推薦する研究は見当たらない。小説の特徴量とタグの特徴量の類似度を直接比較をして推薦するタグを選ぶだけでなく、タグ同士で類似度の比較を行うことで、検索に効果的なタグの組を推薦するというところに本研究の新規性がある。

第3章 小説投稿サイト

本章では、インターネット上で広く利用されている日本の小説投稿サイト「小説家になろう¹」を対象として、小説投稿サイトのデータ構造、特性について説明を行う。データ構造の説明は現在の小説家になろうをもとに行っている。「小説家になろう」はインターネット上での愛称として「なろう」と頻繁に表記される。また「小説家になろう」に投稿されている小説は「なろう小説」とも呼称されている。このことから本章では、「小説家になろう」を「なろうサイト」と以下表記する。

3.1 なろうサイトのデータ構造

なろうサイトは、日本の小説投稿サイトでも多くの人々が利用している。表3.1がなろうサイトの規模である。

なろうサイトでは、小説をNコードというアルファベットのNから始まるコードで小説を識別している。このNコードは小説を投稿した時に割り振られる。

なろうサイトでは、小説を5つの大ジャンルに区分している。それぞれ、恋愛、ファンタジー、文芸、SF、その他である。この大ジャンルの元、それぞれで2つ以上のジャンルに分けられている。ジャンルは合計で20に分けられている。これらの大ジャンル、ジャンルは著作者が小説を投稿するときに、それぞれ設定を行う。

なろうサイトでは、小説に何が書かれているのかを端的に表す語としてキーワードを設定することができる。キーワードは、後述するとおり様々な種類があるが、1つの小説に全種類の合算して15個まで設定できる。キーワードの設定の仕方は、なろうサイトが用意したものを選択して設定する方法と著作者による手動入力による設定する方法の2つがある。前者に分類されるものは、公式キーワードとおすすめキーワードである。公式キーワードはなろうサイトが利用頻度が高いと考え用意されたものである。作品傾向、登場キャラクター、舞台、時代設定、要素の5つに分けられている。公式キーワードの一覧を表3.2に示す。おすすめキーワードは、なろうサイトが読者の検索性向上と著作者へのジャンル区分の傾向を示すという2つの目的のために用意されたものである。おすすめキーワードの一覧を表3.3に示す。他には、二次創作やTRPG（テーブルトークロールプレイングゲーム）の公式シナリオを利用した作品には自動的に設定される固定キーワード（固定キーワード内でさらに二次創作キーワード、公式シナリオ用キーワード、リプレイ用キーワード、管理キーワードに区分される）となろうサイトで開催されているタイアップコンテストや公式企画に参加するための企画キーワードがある。後者に分類されるキーワードは、手動入力キーワードである。これは、著作者が10文字以内で自由な文字列を入力して設定できるキーワードである。この手動入力キーワードは、著作者の手により自由に設定できるという特徴からより各小説にマッチした内容をタグに反映できると考えられる。一方で、手動

¹<https://syosetu.com/> 本章での分析は2022年10月20日に収集したデータに基づいている。

表 3.1: 小説家になろうサイトの規模

登録ユーザ数	2,359,569 人
小説掲載数	995,762 作品
全掲載文字数	42,125,610,988 文字
利用形態	完全 (投稿, 閲覧とも) 無料

入力キーワードはその内容が、著作者に強く依存するため、記述に著作者ごとの偏りなどが生じる可能性がある。

上記のキーワードとは、異なり小説本文中にその要素が含まれている場合必ず小説に設定しなければならない登録必須キーワードというものも存在する。登録必須キーワードの一覧を表に 3.4 示す。これらは、上記のキーワードとは取り扱いが異なり、小説に設定できるキーワード 15 個の制限の外にある。しかし、後述するなろう API では、キーワードとして獲得できる。

なろうサイトでは、商用利用も認められている API (通称 なろう小説 API) が公開されている。技術マニュアル²も公開されている。なろう小説 API の概要³を表 3.5 に示す。このなろう小説 API では、小説の N コード、大ジャンル・ジャンル、キーワード、あらすじ、投稿日時、更新日時、評価ポイントなど様々なメタデータを入手することが可能である。

3.2 タグの特性

著作者によって個々の小説に付与された複数のタグは、異なる小説にも付与されている事がある。このようにして、小説とタグの間には二部グラフが形成されている。このことを明らかにするために、それぞれのタグが付与されている小説数を調査することとした。タグを出現頻度順に横軸方向に並べ、縦軸にそれぞれのタグの出現回数を描画した図が、図 3.1 である。この図は、横軸及び縦軸は対数とした両対数グラフであり、グラフからべき乗則に従って分布していることが分かる。

図 3.2 は、出現数が多いタグ領域と少ないタグ領域とに分けて拡大表示した図である。この図から 10 万件以上の小説に設定されているタグの存在や、タグが 1 つのみ設定された小説が 10 万件以上あることがわかる。

類似したタグの付与状況を明らかにするために、編集距離 d が以下の条件を満たすタグの組を調査することとした。

$$d \leq \lfloor \frac{\text{タグ長}}{3} \rfloor \cdots (d \neq 0) \quad (3.1)$$

ここで、タグ長とは、アルファベット等の 1 バイト文字、ひらがな・カタカナ・漢字などの 2 バイト文字に関わらず、タグを構成する文字数とし、タグ長の 3 分の 1 以内にある編集距離にあるタグを類似したタグとして調査の対象とした。

調査にあたっては計算量を削減する目的で、以下の 2 つの制限を設けた。

タグ長が 3 文字以上 3 文字に満たないタグは、略称等に使用されていることが想定されること、また、文字の組み合わせの自由度が高いことから、タグ付与者が意図して付けた

²<https://dev.syosetu.com/man/man/>

³<https://dev.syosetu.com/man/api/> から抜粋

表 3.2: 公式キーワード一覧

作品傾向	ギャグ, シリアス, ほのぼの, ダーク
登場キャラクター	男主人公, 女主人公, 人外, 魔王, 勇者
舞台	和風, 西洋, 中華, 学園
時代設定	戦国, 幕末, 明治/大正, 昭和, 平成 古代, 中世, 近世, 近代, 現代, 未来
要素	ロボット, アンドロイド, 職業もの ハーレム, 逆ハーレム, 群像劇, チート 内政, 魔法, 冒険, ミリタリー, 日常 ハッピーエンド, バッドエンド, グルメ 青春, ゲーム, 超能力, タイムトラベル ダンジョン, パラレルワールド, タイムリープ

表 3.3: おすすめキーワード一覧

該当大ジャンル	該当ジャンル	おすすめキーワード
恋愛	異世界	異類婚姻譚, 身分差, 年の差, 悲恋
	現実世界	ヒストリカル, 乙女ゲーム, 悪役令嬢 オフィスラブ, スクールラブ, 古典恋愛
ファンタジー	ハイファンタジー	オリジナル戦記
	ローファンタジー	伝記
文芸	ヒューマンドラマ	日常, 青春, ハードボイルド, 私小説, ホームドラマ
	歴史	IF 戦記, 史実, 時代小説, 逆行転生
	推理	ミステリー, サスペンス, 探偵小説
	ホラー	スプラッタ, 怪談, サイコホラー
	アクション	異能力バトル, ヒーロー, スパイ, 冒険
	コメディ	ラブコメ
SF	VR ゲーム	近未来, 人工知能, 電脳世界 VRMMO
	宇宙	スペースオペラ, エイリアン
	空想科学	サイバーパンク, スチームパンク, ディストピア, タイムマシン
	パニック	怪獣, 天災, バイオハザード, パンデミック

表 3.4: 登録必須キーワード一覧

R15
ボーイズラブ
ガールズラブ
残酷な描写あり
異世界転生
異世界転移

表 3.5: なろう小説 API の概要

出力形式	JSON 形式, YAML 形式など
文字コード	UTF-8
API の URL	https://api.syosetu.com/novelapi/api/
大ジャンル区分	1: 恋愛, 2: ファンタジー, 3: 文芸, 4: SF, 99: その他, 98: ノンジャンル
ジャンル区分	101: 異世界 [恋愛], 102: 現実世界 [恋愛] 201: ハイファンタジー [ファンタジー], 202: ローファンタジー [ファンタジー] 301: 純文学 [文芸], 302: ヒューマンドラマ [文芸], 303: 歴史 [文芸], 304: 推理 [文芸], 305: ホラー [文芸], 306: アクション [文芸], 307: コメディ [文芸] 401: VR ゲーム [SF], 402: 宇宙 [SF], 403: 空想科学 [SF], 404: パニック [SF] 9901: 童話 [その他], 9902: 詩 [その他], 9903: エッセイ [その他], 9904: リプレイ [その他], 9999: その他 [その他] 9801: ノンジャンル [ノンジャンル]

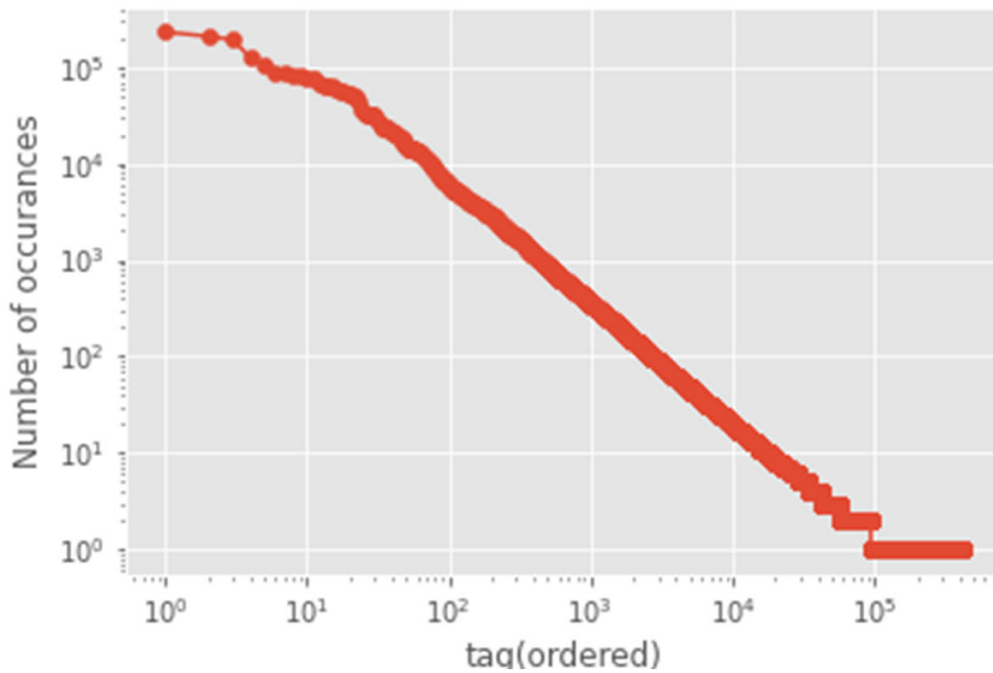
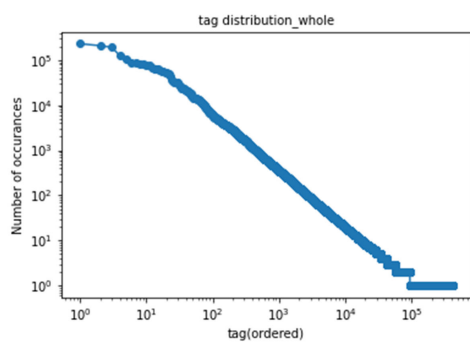
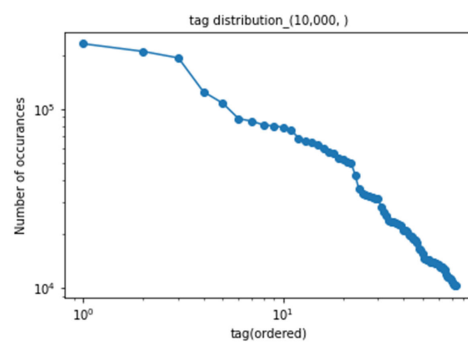


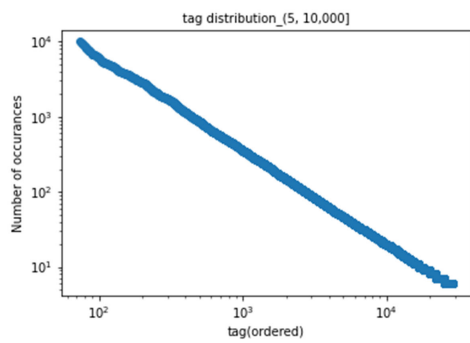
図 3.1: タグの出現頻度分布 (全体)



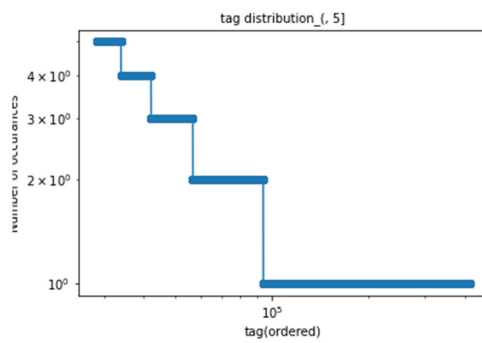
(a) 全体



(b) 先頭部



(c) 昼間部



(d) 末尾部

図 3.2: タグの出現頻度分布 (分割拡大)

のか、あるいは、誤って意味の無い文字列を付与したのかの判断が難しいと考えた。このため、最短のタグ長を3とすることにした。なお、タグ長の最長は21文字であった。

タグ長の差が高々1以内 比較する2個のタグ長が大きく異なった場合でも、編集距離の条件式3.1を満たす場合はあるが、数10万件あるタグ全ての組み合わせを評価することは計算量的に現実的では無い。このため、比較するタグは長さの違いが高々1以内、すなわち、自分自身の長さ ± 1 の範囲にあるタグとの比較に限定することとした。

なろうサイトで使用されているタグ間の編集距離を計測した結果を表に示す。表3.6はタグ長が3文字の例、表3.7はタグ長が4文字の例、表3.8はタグ長が5文字の例、表3.9はタグ長が6文字の例、表3.10はタグ長が7文字の例、表3.11はタグ長が8文字の例、表3.12はタグ長が9文字の例、表3.13はタグ長が10文字および11文字の例、表3.14はタグ長が12文字から21文字の例である。

それぞれ、第一タグと第二タグについて見ていくと、それぞれ異なる内容を示すタグがある一方で、タイプミスや表記ゆれなどによって発生したと推測される出現数が小さい類似したタグが散見される。また、これはタグ長が長くなるにつれて増加する傾向がみられる。

以上のことからタグ付与に関する顕著な傾向を読み取ることは難しいが、比較した2個のタグ間で付与されている小説数に大きな差異がある場合は、少ない方のタグがタイプミス等によって付与された可能性が示唆される。このようなタイプミス等によって生じたタグには、より多数の出現頻度を持つタグを推薦することで、誤ったタグの付与を防ぐことができる。

表 3.6: タグ間の編集距離: タグ長 3 文字

編集距離	第一タグ		第二タグ	
	タグ	使用小説数	タグ	使用小説数
タグ長 3 同士の比較 (一部)				
1	R15	231170	R18	66
1	R15	231170	R 15	30
1	R15	231170	R12	22
1	R15	231170	r15	16
1	R15	231170	R17	8
1	R15	231170	R10	4
1	異世界	52588	別世界	142
1	異世界	52588	新世界	86
1	異世界	52588	裏世界	56
1	異世界	52588	夢世界	48
1	異世界	52588	多世界	26
1	異世界	52588	銀世界	13
1	チート	35753	チープ	8
1	チート	35753	チー牛	8
1	チート	35753	シート	3
1	チート	35753	チーと	3
1	チート	35753	チーフ	2
1	チート	35753	チート	2
1	チート	35753	チート	2
1	チート	35753	二ート	2
1	チート	35753	チー鱈	1
1	チート	35753	チーク	1
1	チート	35753	ポート	1
1	チート	35753	ソート	1
1	チート	35753	チータ	1
1	チート	35753	チ○ト	1
1	チート	35753	メート	1
1	チート	35753	血ート	1
1	チート	35753	リート	1
1	チート	35753	チー虎	1

表 3.7: タグ間の編集距離: タグ長 4 文字

編集距離	第一タグ		第二タグ	
	タグ	使用小説数	タグ	使用小説数
タグ長 4 同士の比較 (一部)				
1	男主人公	107669	女主人公	68625
1	男主人公	107669	W 主人公	188
1	男主人公	107669	W主人公	33
1	男主人公	107669	猫主人公	32
1	男主人公	107669	強主人公	14
1	男主人公	107669	屑主人公	9
1	男主人公	107669	多主人公	7
1	男主人公	107669	爺主人公	6
1	男主人公	107669	犬主人公	5
1	男主人公	107669	w主人公	4
1	ほのぼの	79055	ぼのぼの	28
1	ほのぼの	79055	ほもぼの	8
1	ほのぼの	79055	ほのぼぼ	2
1	ほのぼの	79055	ほのぼの	2
1	ほのぼの	79055	ほのぼほ	1
1	ほのぼの	79055	ほのぼの	1
1	ほのぼの	79055	ほのもの	1
1	シリアス	76104	シリアル	173
1	シリアス	76104	シリウス	13
1	シリアス	76104	イリアス	4
1	女主人公	68625	W 主人公	188
1	女主人公	68625	W主人公	33
1	女主人公	68625	猫主人公	32
1	女主人公	68625	強主人公	14
1	女主人公	68625	屑主人公	9
1	ラブコメ	63225	ラヴコメ	11
1	ラブコメ	63225	ラブホメ	1
1	ラブコメ	63225	ラベコメ	1
1	ラブコメ	63225	ラバコメ	1
1	ラブコメ	63225	バブコメ	1
1	ラブコメ	63225	ラブコン	1
1	ハーレム	21114	ゴーレム	311
1	ハーレム	21114	チーレム	123
1	ハーレム	21114	ハーレー	8
1	ハーレム	21114	ハーレ無	5
1	ハーレム	21114	ハーレ夢	1
1	ハーレム	21114	パーレム	1
1	ハーレム	21114	〇ーレム	1
1	ハーレム	21114	ハーレム	1

表 3.8: タグ間の編集距離: タグ長 5 文字

編集距離	第一タグ		第二タグ	
	タグ	使用小説数	タグ	使用小説数
タグ長 5 同士の比較 (一部)				
1	異世界転移	65851	異世界転生	60395
1	異世界転移	65851	異世界転送	22
1	異世界転移	65851	異世界転位	18
1	異世界転移	65851	異世界転成	10
1	異世界転移	65851	異世界転性	3
1	異世界転移	65851	異世界転写	3
1	異世界転移	65851	異世界転死	3
1	異世界転移	65851	異世界転職	3
1	異世界転移	65851	自世界転移	2
1	異世界転移	65851	異世界転校	2
1	異世界転移	65851	異世界転々	2
1	ミステリー	23374	ミステリ?	23
1	ミステリー	23374	ミステリィ	11
1	ミステリー	23374	ヒステリー	10
1	ミステリー	23374	ミステリ風	9
1	ミステリー	23374	ミステリー	8
1	ミステリー	23374	微ステリー	3
1	主人公最強	13810	主人公最弱	295
1	主人公最強	13810	主人公最凶	30
1	主人公最強	13810	主人公最低	21
1	主人公最強	13810	主人公最狂	12
1	主人公最強	13810	主人公最恐	11
1	主人公最強	13810	主人公最高	8
1	コメディ	13093	コメディ?	351
1	コメディ	13093	コメディ風	20
1	コメディ	13093	コメディ調	18
1	コメディ	13093	コメディ	13
1	コメディ	13093	コメディ?	13
1	コメディ	13093	コメディー	7
1	乙女ゲーム	11577	乙女ゲー?	26
1	乙女ゲーム	11577	乙女ゲー風	5
1	乙女ゲーム	11577	漢女ゲーム	1
1	乙女ゲーム	11577	乙女ゲーム	1
1	乙女ゲーム	11577	乙女ゲーム	1
1	学校/学園	10330	学校/学園	42
1	学校/学園	10330	学校・学園	17
1	学校/学園	10330	学校/学院	4
1	学校/学園	10330	学院/学園	1
1	学校/学園	10330	学校/学生	1

表 3.9: タグ間の編集距離: タグ長 6 文字

編集距離	第一タグ		第二タグ	
	タグ	使用小説数	タグ	使用小説数
タグ長 6 同士の比較 (一部)				
1	ファンタジー	65494	ファンダジー	18
2	ファンタジー	65494	ファンレター	14
1	ファンタジー	65494	ファンタジ	13
1	ファンタジー	65494	ファンタジ	11
1	ファンタジー	65494	ファンタ字	1
2	ファンタジー	65494	ファンタジ	1
1	ファンタジー	65494	ファンタジ-	1
1	ファンタジー	65494	ファンタゾ	1
2	ファンタジー	65494	ファンスキー	1
1	ファンタジー	65494	腐ファンタジ	1
1	ファンタジー	65494	ファンタジ	1
1	異能力バトル	52796	超能力バトル	59
2	異能力バトル	52796	異世界バトル	50
2	異能力バトル	52796	能力者バトル	32
1	異能力バトル	52796	異能者バトル	8
2	異能力バトル	52796	能力バトル物	6
1	異能力バトル	52796	霊能力バトル	5
2	異能力バトル	52796	能力系バトル	4
2	異能力バトル	52796	異能バトル系	3
1	異能力バトル	52796	胃能力バトル	2
2	スクールラブ	33729	スクール水着	22
2	スクールラブ	33729	スクールバス	3
2	スクールラブ	33729	スクール戦記	1
2	ガールズラブ	22367	ボーイズラブ	16463
2	ガールズラブ	22367	ガールズバー	9
1	ガールズラブ	22367	ガールズラズ	1
2	ガールズラブ	22367	ボーイズラブ	1
1	ディストピア	6764	ディストピア	1
1	ディストピア	6764	ディストピア	1
1	ディストピア	6764	ディストニア	1
1	ESN 大賞 4	6609	ESN 大賞 3	2731
1	ESN 大賞 4	6609	ESN 大賞 2	2220
1	ESN 大賞 4	6609	ESN 大賞 4	808
1	ESN 大賞 4	6609	ESN 大賞 3	71
1	ESN 大賞 4	6609	ESN 大賞 2	64
2	ESN 大賞 4	6609	ENS 大賞 4	6

表 3.10: タグ間の編集距離: タグ長 7 文字

編集距離	第一タグ		第二タグ	
	タグ	使用小説数	タグ	使用小説数
タグ長 7 同士の比較 (一部)				
1	残酷な描写あり	209350	残酷な描写有り	1821
1	残酷な描写あり	209350	残念な描写あり	103
2	残酷な描写あり	209350	残酷な表現あり	53
2	残酷な描写あり	209350	残酷な描写なし	29
2	残酷な描写あり	209350	性的な描写あり	25
1	残酷な描写あり	209350	残虐な描写あり	18
2	残酷な描写あり	209350	不快な描写あり	8
2	残酷な描写あり	209350	下品な描写あり	6
1	残酷な描写あり	209350	残酷な模写あり	5
2	オリジナル戦記	56753	オリジナル小説	133
2	オリジナル戦記	56753	オリジナル作品	61
2	オリジナル戦記	56753	オリジナル設定	45
2	オリジナル戦記	56753	オリジナル兵器	33
2	オリジナル戦記	56753	オリジナル武器	22
2	オリジナル戦記	56753	オリジナル神話	18
2	オリジナル戦記	56753	オリジナル童話	18
2	オリジナル戦記	56753	オリジナル展開	16
2	オリジナル戦記	56753	オリジナル魔法	14
2	オリジナル戦記	56753	オリジナル世界	13
1	ハッピーエンド	50453	パッピーエンド	20
1	ハッピーエンド	50453	ハーピーエンド	3
1	ハッピーエンド	50453	ハッピーエンド	2
2	現代 (モダン)	26570	現代 (モダン)	487
1	現代 (モダン)	26570	現実 (モダン)	6
1	現代 (モダン)	26570	現代 (モダン)	3
1	現代 (モダン)	26570	現在 (モダン)	1
1	現代 (モダン)	26570	現代 (モダン)	1
1	ハードボイルド	9354	ハーフボイルド	9
1	ハードボイルド	9354	ハートボイルド	2
2	ハードボイルド	9354	ハード茄子ルド	1
1	ハードボイルド	9354	フードボイルド	1
1	魔法使い/魔女	3358	魔法使い/魔女	10
1	魔法使い/魔女	3358	魔法使い・魔女	4
1	魔法使い/魔女	3358	魔法使い、魔女	2
2	魔法使い/魔女	3358	魔法使い/賢者	1
2	魔法使い/魔女	3358	魔法使いの少女	1
2	魔法使い/魔女	3358	魔法使い/剣士	1
2	魔法使い/魔女	3358	魔法使い、巫女	1

表 3.11: タグ間の編集距離: タグ長 8 文字

編集距離	第一タグ		第二タグ	
	タグ	使用小説数	タグ	使用小説数
タグ長 8 同士の比較 (一部)				
2	ショートショート	11306	ショートシリーズ	4
1	ショートショート	11306	シュートショート	1
1	ショートショート	11306	ショートショーツ	1
1	ショートショート	11306	ショートシュート	1
1	ショートショート	11306	ショーでショート	1
1	ネット小説大賞十	3626	ネット小説大賞九	3289
1	ネット小説大賞十	3626	ネット小説大賞五	1806
1	ネット小説大賞十	3626	ネット小説大賞八	1527
1	ネット小説大賞十	3626	ネット小説大賞六	900
1	ネット小説大賞十	3626	ネット小説大賞七	848
1	ネット小説大賞十	3626	ネット小説大賞+	45
1	ネット小説大賞十	3626	ネット小説大賞+	7
1	ネット小説大賞十	3626	ネット小説大賞 6	6
1	ネット小説大賞十	3626	ネット小説大賞 5	5
2	ノンフィクション	2252	メタフィクション	214
1	ノンフィクション	2252	ノーフィクション	10
2	ノンフィクション	2252	完全フィクション	8
2	ノンフィクション	2252	一部フィクション	7
2	ノンフィクション	2252	半分フィクション	6
2	ノンフィクション	2252	セミフィクション	5
2	ノンフィクション	2252	歴史フィクション	4
2	ノンフィクション	2252	多分フィクション	4
2	ノンフィクション	2252	ほぼフィクション	3
2	ノンフィクション	2252	少しフィクション	2
2	ヒューマンドラマ	2071	ヒューマンホラー	9
2	ヒューマンドラマ	2071	ヒューマンエラー	3
2	ヒューマンドラマ	2071	ニヤーマンドラマ	1
1	ヒューマンドラマ	2071	ヒュウマンドラマ	1
2	ハイファンタジー	1853	現代ファンタジー	1765
2	ハイファンタジー	1853	ローファンタジー	1141
2	ハイファンタジー	1853	和風ファンタジー	733
2	ハイファンタジー	1853	恋愛ファンタジー	235
2	ハイファンタジー	1853	SF ファンタジー	226
2	ハイファンタジー	1853	学園ファンタジー	180
2	ハイファンタジー	1853	王道ファンタジー	176
2	ハイファンタジー	1853	冒険ファンタジー	86
2	ハイファンタジー	1853	歴史ファンタジー	76
2	ハイファンタジー	1853	中世ファンタジー	72
2	ハイファンタジー	1853	中華ファンタジー	71

表 3.12: タグ間の編集距離: タグ長 9 文字

編集距離	第一タグ		第二タグ	
	タグ	使用小説数	タグ	使用小説数
タグ長 9 同士の比較 (一部)				
3	OL/サラリーマン	4188	異世界サラリーマン	2
3	OL/サラリーマン	4188	普通のサラリーマン	2
2	OL/サラリーマン	4188	OL/サラリーマン	2
3	OL/サラリーマン	4188	OL/サラリーマン	2
3	OL/サラリーマン	4188	昭和のサラリーマン	1
3	OL/サラリーマン	4188	なろうサラリーマン	1
1	エンターテイメント	3948	エンターテイメント	8
1	エンターテイメント	3948	エンターテイメント	6
2	エンターテイメント	3948	エンターテイメント	1
1	エンターテイメント	3948	エンターテイメント	1
3	ダークファンタジー	2526	異世界ファンタジー	1258
3	ダークファンタジー	2526	バトルファンタジー	104
3	ダークファンタジー	2526	中華風ファンタジー	78
3	ダークファンタジー	2526	西洋風ファンタジー	73
3	ダークファンタジー	2526	ハイ・ファンタジー	50
3	ダークファンタジー	2526	中世風ファンタジー	44
3	異世界ファンタジー	1258	バトルファンタジー	104
3	異世界ファンタジー	1258	中華風ファンタジー	78
3	異世界ファンタジー	1258	西洋風ファンタジー	73
3	異世界ファンタジー	1258	ハイ・ファンタジー	50
2	異世界ファンタジー	1258	中世風ファンタジー	44
3	異世界ファンタジー	1258	現代風ファンタジー	42
3	異世界ファンタジー	1258	ロー・ファンタジー	28
3	異世界ファンタジー	1258	ライトファンタジー	26
2	ボーイミーツガール	1155	ガールミーツガール	45
3	ボーイミーツガール	1155	ボーイッシュガール	3
1	ボーイミーツガール	1155	ボーズミーツガール	2
1	ボーイミーツガール	1155	ボーイミーツガール	2
2	ボーイミーツガール	1155	ボーイミーツボーイ	1
1	夏のホラー 2020	1148	夏のホラー 2021	1006
1	夏のホラー 2020	1148	夏のホラー 2022	803
2	夏のホラー 2020	1148	夏のホラー 2019	543
2	夏のホラー 2020	1148	夏のホラー 2017	542
2	夏のホラー 2020	1148	夏のホラー 2018	512
1	夏のホラー 2020	1148	夏のホラー 2010	16
2	夏のホラー 2020	1148	夏のホラー 2009	4
3	夏のホラー 2020	1148	秋もホラー 2022	2
2	夏のホラー 2020	1148	秋のホラー 2021	2
3	夏のホラー 2020	1148	冬のホラー 2016	2

表 3.13: タグ間の編集距離: タグ長 10~11 文字

編集距離	第一タグ		第二タグ	
	タグ	使用小説数	タグ	使用小説数
タグ長 10 同士の比較例				
1	ネット小説大賞十感想	3876	ネット小説大賞九感想	2535
1	ネット小説大賞十感想	3876	ネット小説大賞六感想	2265
1	ネット小説大賞十感想	3876	ネット小説大賞七感想	2030
1	ネット小説大賞十感想	3876	ネット小説大賞八感想	1658
1	ネット小説大賞十感想	3876	ネット小説大賞+感想	30
1	ネット小説大賞十感想	3876	ネット小説大賞+感想	6
1	ネット小説大賞十感想	3876	ネット小説大賞五感想	5
1	ネット小説大賞十感想	3876	ネット小説大賞7感想	4
1	ネット小説大賞十感想	3876	ネット小説大賞6感想	2
1	ネット小説大賞十感想	3876	ネット小説大賞6感想	1
1	ネット小説大賞十感想	3876	ネット小説大賞8感想	1
2	ネット小説大賞十感想	3876	ネット小説大賞十SC	93
2	ネット小説大賞十感想	3876	ネット小説大賞十参加	13
2	ネット小説大賞十感想	3876	ネット小説大賞感想済	6
2	ネット小説大賞十感想	3876	ネット小説大賞感想七	2
2	ネット小説大賞十感想	3876	ネット小説大賞十1次	1
2	ネット小説大賞十感想	3876	ネット小説大賞九十想	1
2	ネット小説大賞十感想	3876	ネット小説九大賞感想	1
2	ネット小説大賞十感想	3876	ネット小説大賞感想六	1
2	ネット小説大賞十感想	3876	ネット小説大賞感想八	1
3	ネット小説大賞十感想	3876	ネット小説大賞八短編	422
3	ネット小説大賞十感想	3876	ネット小説大賞八短二	81
3	ネット小説大賞十感想	3876	ネット小説大賞応募作	9
3	ネット小説大賞十感想	3876	ネット小説大賞+SC	8
3	ネット小説大賞十感想	3876	ネット小説大賞受賞作	7
3	ネット小説大賞十感想	3876	ネット小説大賞+SC	3
3	ネット小説大賞十感想	3876	ネット小説大賞8受賞	2
3	ネット小説大賞十感想	3876	ネット小説大賞落選!	2
3	ネット小説大賞十感想	3876	ネット小説対象六感想	2
3	ネット小説大賞十感想	3876	ネット小説大賞六金賞	1
3	ネット小説大賞十感想	3876	ネット小説対象九感想	1
3	ネット小説大賞十感想	3876	ネット小説大賞九1次	1
3	ネット小説大賞十感想	3876	ネット小説大賞六落選	1
3	ネット小説大賞十感想	3876	ネット小説大賞八落選	1
3	ネット小説大賞十感想	3876	ネット小説大賞九落選	1
3	ネット小説大賞十感想	3876	ネット小説対象七感想	1
タグ長 11 同士の比較				
2	セ〇とはあえて言わない	1	〇ミとはあえて言わない	1
2	TM オマージュ小説企画	1	TMオマージュ小説企画	1

表 3.14: タグ間の編集距離: タグ長 12~21 文字

編集距離	タグ	使用小説数
タグ長 12 同士の比較 (一部)		
1	夏のホラー 2009 四夜目	12
	夏のホラー 2009 二夜目	10
1	夏のホラー 2009 四夜目	12
	夏のホラー 2009 三夜目	8
1	夏のホラー 2009 四夜目	12
	夏のホラー 2009 五夜目	4
1	夏のホラー 2009 四夜目	12
	夏のホラー 2009 一夜目	3
4	夏のホラー 2009 五夜目	4
	夏のホラー 2009 五夜目	1
2	言の葉の森企画「島作り」	1
	言の葉の森企画『島作り』	1
タグ長 13 同士の比較		
1	仮題「シャングリラ」第四章	1
	仮題「シャングリラ」第三章	1
1	仮題「シャングリラ」第四章	1
	仮題「シャングリラ」第二章	1
1	仮題「シャングリラ」第四章	1
	仮題「シャングリラ」第1章	1
1	仮題「シャングリラ」第三章	1
	仮題「シャングリラ」第二章	1
1	仮題「シャングリラ」第三章	1
	仮題「シャングリラ」第1章	1
1	仮題「シャングリラ」第二章	1
	仮題「シャングリラ」第1章	1
タグ長 19 同士の比較		
5	オリジナル青春野球部ラブストーリー小説	1
	オリジナルバンド青春ラブストーリー小説	1
タグ長 20 同士の比較		
5	心が忘れた温もりを取り戻しに行きませんか	3
	心が忘れた物を取り戻しに行きませんか…。	1
4	オリジナル軽音楽部青春ラブストーリー小説	1
	オリジナル青春軽音楽部ラブストーリー小説	1
タグ長 21 同士の比較		
4	BL オリジナル青春野球部ラブストーリー小説	1
	オリジナル BL 青春野球部ラブストーリー小説	1
4	オリジナル BL 小説青春野球部ラブストーリー	1
	オリジナル BL 青春野球部ラブストーリー小説	1

第4章 提案手法

本研究では、小説投稿サイトに投稿されている小説を読んだユーザが、その小説と類似した要素を持つ小説を検索するときに効果的な検索タグを推薦する手法について提案を行う。

4.1 概要

以下では、提案手法全体の概要について説明を行い、それぞれの詳細については、各節で説明を行う。まず、小説とその小説に設定されたタグの対から、タグの出現回数を基にタグの情報量を求める。求めたタグの情報量から、情報量の極めて小さいものと情報量の極めて大きいものを推薦の候補から取り除き、推薦するタグ候補の範囲として設定する。次に、小説本文の特徴量を生成し、その小説本文の特徴量を基に推薦候補の範囲にあるタグの特徴量を生成する。最後に、入力となる小説本文に対して、検索に効果的なタグの推薦を行う。検索に効果的なタグの選出には、上記で求めたタグの情報量と類似度を用いて行う。

提案手法の基本的なアイデアは、国際会議で発表 [8] しており、本論文は、基本アイデアを精緻化するとともに、より発展した内容となっている。

本提案手法によるシステムは、以下のような文脈のシナリオにおいて活用されると想定している。まず、ユーザが小説投稿サイトの1つのオンライン小説を読み終えたと仮定する。ユーザはその小説の内容に、面白い、感動したなどの感情を得たり、興味関心を覚えたりし、新たにその小説と類似した小説を読みたいと考える。しかし、小説投稿サイトに存在する数多くの小説の中から探索を行うことは困難である。そこで、ユーザの探したい小説にマッチした検索キーワードが必要になる。しかし、ユーザが自分の力で見つけ出すことは、小説投稿サイトのコンテキストを理解している必要があるが、日々新たな小説が投稿されることから正確なコンテキストを把握することは困難である。この時、ユーザは適切な検索用のキーワードが必要であるというニーズが生じる。この時、システムに読了済みの小説を入力することで、その小説と類似した小説が効果的に検索されるタグを複数推薦する。ユーザは、この推薦されたタグを選択して検索に用いることで、自分の読みたいと考える小説が探しやすくなると考えられる。

本研究における効果的な検索タグとは、以下の3つの要件を満たすタグである。

- 入力された小説の内容と合致したタグ
- 推薦されたタグ同士でタグの表す情報に重複がないタグ
- 複数のタグで絞り込みができるタグ

上記の要件を満たすタグとする理由をそれぞれ説明する。1つ目の「小説の内容と合致したタグ」という要件は、小説の内容に合致するすなわち、その小説もしくはその小説に類似した小説にはそのタグと同じ内容が含まれていると考えられるためである。そのため、推薦されるタグは、ユーザに入力された小説と似ている必要がある。2つ目の「推薦されたタグ同士でタグの表す情報に重複がないタグ」というものは、推薦されたタグ同士で内容が異

なっているほうが検索の際により効果的であると考えられるためである。小説とは、1つの作品中においても、作品内での場面や時間における変化など、様々な要素を持っていると考えられる。そのため、検索のためのタグを推薦するときに、同じ要素を示すタグが複数推薦されても、それで示すことができるものは、その小説の一部分に偏ってしまう。推薦するタグが異なる内容を表すようにすることで、小説の含む要素をまんべんなく示すことができる。また、推薦されたタグの中からユーザーが特に好ましく考える要素を選出して検索に活用することができる。3つ目の「複数のタグで絞り込みができるタグ」は、単一のタグでは絞り込みが難しいケースが想定されうるためである。タグの中には、多くの小説に設定された出現数が多いものがある。そのため、仮に1つだけタグを推薦するとき、出現数が多いタグが推薦されると、そのタグだけでは十分な絞り込みが行えず、結局多くの小説の中から類似した小説を探すことになり、ユーザーの検索を支援するという目的を達成することが難しくなる。そこで複数のタグを推薦することで、絞り込みが可能な検索を可能とする。

読了した小説と類似した別の小説を直接推薦するのではなく、検索するためのタグを推薦する理由は、計算量を抑え効率化を図るためである。オンライン小説は、日々内容を更新していく連載作品が多い。また、新たな作品が頻繁に投稿されている。これらをすべて推薦する候補とするために、毎回類似度比較のために計算をすることは非効率的であると考えられる。そこで、本研究では、事前にタグの特徴量を求めておくことで、計算量を抑えられると考えた。

4.2 タグの情報量

ある事象が発生したとき、その事象がどの程度まれに発生するかを示す尺度を情報量という。事象 e が発生する確率が $P(e)$ のとき、その事象の情報量 $I(e)$ は以下の式で示される。

$$I(e) = -\log P(e) \quad (4.1)$$

したがって、事象が発生する確率 $P(e)$ が低くなるほど、事象 e の情報量 $I(e)$ はより大きな値となる。逆に、その事象 e が頻繁に発生し確率 $P(e)$ が高くなるにつれて、事象 e の情報量 $I(e)$ は小さくなっていく。

[9]の研究で、Ouraらはオンラインショッピングサイトでのクエリの出現という事象を確率的な事象とし、検索クエリのワードの情報量を求めている。この手法を参考に、本研究では、この情報量の概念を小説投稿サイト内でのタグの出現数とし、タグの情報量を求める。タグが出現する、すなわち1つの小説に設定されることを1つの事象として扱う。タグ t が1つの小説に設定されるという事象が起きたとき、情報量 $I(t)$ は以下の式で表せる。

$$I(t) = -\log P(t) \quad (4.2)$$

タグの出現確率 $P(t)$ は、タグ t が出現する回数 $n(t)$ 、すべてのタグの総出現数 N としたとき、以下の式で求められる。

$$P(t) = \frac{n(t)}{N} \quad (4.3)$$

タグの出現回数の値が大きくなるとそれに伴い、タグの情報量は小さくなる。逆に、タグの出現回数の値が小さくなるほど、タグの情報量は大きくなる。

タグの情報量が小さくなるとき、このタグは非常に多くの小説に設定されている、普遍的な性質を持ったタグであるといえる。このようなタグは普遍的であるゆえに検索で絞り

込みが難しいと考えられる。また、普遍的なタグは、小説投稿サイトを利用するユーザーにとって、既知なタグであると想定される。これらのことから、情報量の極めて小さいタグは、ユーザーに推薦する候補から外す必要がある。逆に、タグの情報量が非常に大きくなる時、そのタグはごくわずかな小説にのみ設定されているものであり、特定性の強いという性質を持っている。このようなタグは、検索に用いてもそのタグが設定されているごくわずかな小説のみがヒットし、新たな作品の発見には適していないと考えられる。したがって、情報量の極めて大きなタグも推薦するタグの候補から外す必要がある。以上のことから、タグの情報量をもとにタグの推薦範囲を定める。

4.3 タグの特徴量

小説投稿サイトにおいて小説とタグは2部グラフの関係にある。タグは、その設定されたオンライン小説の世界観、テーマ、登場人物等の要素を示すものである。そのため、当然それらのタグが設定されている小説はタグが示す要素を内包しているといえる。したがって、タグは設定された小説と同様な性質を有していると考えられる。以上のことから、タグは設定された小説の特徴量を合成することで、そのタグの特徴量を求めることができると仮定する。タグの特徴量 \vec{T}_j は、そのタグが設定されている i 個の小説のベクトル \vec{N}_i を合成したものである。 \vec{T}_j は以下の式から求められる。

$$\vec{T} = \sum_{i \in D_i \rightarrow T_j} \vec{D}_i \quad (4.4)$$

推薦を行う前に、4.2節で定めた推薦する候補となるタグの特徴量を全て求めておく。タグの大きさを統一するために、単位ベクトルに変換を行う。

4.4 タグ推薦

ユーザーが入力する読了した小説の本文の特徴量を求め、4.3節で求めておいたタグの特徴量と類似度を比較し、タグの推薦を行う。類似度の比較には Cos 類似度を用いる。入力された小説本文のベクトルを \vec{D}_i 、類似度比較を行うタグのベクトルを \vec{T}_j とすると Cos 類似度は以下の式で求められる。

$$\cos(\vec{D}_i, \vec{T}_j) = \frac{\vec{D}_i \cdot \vec{T}_j}{|\vec{D}_i| |\vec{T}_j|} \quad (4.5)$$

Cos 類似度は -1 から 1 の範囲で表される。より 1 に近い値を取るものが、入力した小説本文により類似したタグである。よって、小説本文の持つ性質に合致したタグを推薦するときは、Cos 類似度が 1 に近い順にタグの推薦を行えばよい。しかし、そのまま類似度が高い順にタグの推薦を行った場合、タグ同士で Cos 類似度比較を行うと、高い類似度の値を示すタグが推薦される可能性が考えられる。タグ間で Cos 類似度が高いということは、小説の同じ要素を示すタグであるということになる。このままでは、検索に効果的なタグの要件を満たすことができない。したがって、文書と高い類似度を持つが、タグ間では異なった情報を持つタグを検索に有効なタグとして推薦する手法について考える必要がある。

4.4.1 手法 A

小説と高い Cos 類似度を持ちつつ、推薦されるタグ同士では類似性が低いタグの推薦をする。端的に示すと、タグの特徴量を座標上に位置した時、既に推薦されたタグと逆位置に存在し小説本文に類似したタグを推薦することで上記の推薦を行う。以下にその詳細を記す。

入力した小説本文 \vec{D} は推薦の基準として、 \vec{T}_0 とする。推薦するタグは 1 番目には選ばれるタグから N 番目のタグを、 $\vec{T}_1, \vec{T}_2 \dots \vec{T}_N$ とする。 \vec{T}_1 は、小説本文と最も類似度が高いタグを選出する。

$$\begin{aligned}\vec{T}_0 &= \vec{D} \\ \vec{T}_1 &= \arg \min_i |\overrightarrow{DT_i}| \end{aligned} \tag{4.6}$$

新たに推薦するタグが既に推薦されたタグと類似性の低い特徴量を持つようにするためには、今まで推薦されたタグ全てを考慮する必要があるが、計算量を抑えるために、直近に推薦された 2 つのタグを考慮する対象とする。 \vec{T}_i を次の推薦するタグとすると、直近 2 つに推薦されたタグは \vec{T}_{i-1} と \vec{T}_{i-2} と示せる。 \vec{T}_{i-1} と \vec{T}_{i-2} をつなぐ線を、小説本文 D から引いた二等分線の交点を \hat{T} とする。この \hat{T} の逆位置にあたる $-\hat{T}$ に近いものから推薦タグを選ぶ。 \vec{T}_i を $-\hat{T}$ と異なる象限へと写像するために、 $-\overrightarrow{DT}$ と $\overrightarrow{DT_i}$ がなす角 θ_i (θ_i は $0 \leq \theta_i \leq \pi$ の範囲) を二等分した位置に T_i を写像し T'_i とする。 T'_i から $-\overrightarrow{DT}$ へと伸ばした垂線の交点を T''_i とする。このとき、以下の式で T''_i を求められる。

$$\begin{aligned}\vec{T}_i'' &\leftarrow \arg \min_i |\overrightarrow{DT_i''}| \\ |\overrightarrow{DT_i''}| &= \frac{|\overrightarrow{DT_i'}|}{\cos(\frac{\theta_i}{2})} \end{aligned} \tag{4.7}$$

求められた T''_i の元となったタグが推薦されるタグへと選出される。図 4.4.1 が上記の図解である。この手順を N 番目まで繰り返し、ユーザに推薦するタグを決定する。

4.4.2 手法 B

手法 A では、タグの特徴量と入手された小説本文を比較することで、検索に効果的なタグの推薦を行った。この手法では、検索に効果的なタグの要件である「入力された小説の内容と合致したタグ」と「推薦されたタグ同士でタグの表す情報に重複がないタグ」の 2 つに重点を置いてタグの推薦を行っている手法である。しかし、もう 1 つの要件である「複数のタグで絞り込みができるタグ」については、4.2 で行った、タグの情報量によるタグの推薦範囲を定めただけに留まっている。そのため、手法 A では推薦したタグの中で情報量に偏りがみられる可能性がある。このとき、複数あっても検索するときあまり絞り込みができないケースなどが推測される。

手法 A では、推薦するタグを選出するときに、最も重要視したものは、入力された小説本文との特徴量との Cos 類似度の高さであった。そこで、手法 B ではタグの情報量という

軸からの視点を，手法 A へと取り入れる．タグの情報量を I とし，タグの特徴量を F としたとき，以下の式から推薦するタグ \vec{T}_i を求める．

$$\vec{T}_i = \alpha F + (1 - \alpha) \frac{1}{I} \quad (4.8)$$

このとき α は $0 \leq \alpha \leq 1$ とする．手法 A の小説本文の特徴量とタグの特徴量を比較するときに，上記の式を当てはめて，タグの推薦を行う．

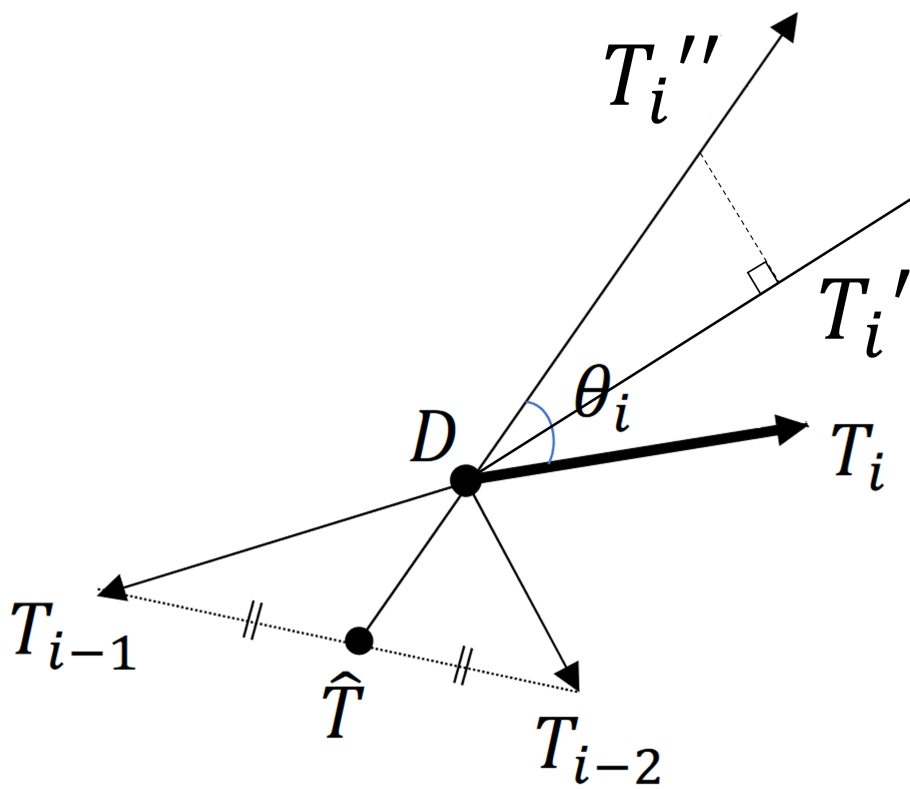


図 4.1: 推薦タグ T_i の図解

第5章 実験・評価

本章では、第4章で提案した手法について実験及び評価を行う。

5.1 データセット

本研究の実験・評価には、「小説家になろう」に投稿された小説およびそれに付随するタグを用いる。

5.1.1 タグ

小説に設定されているタグ（小説家になろうサイト上では、キーワードが相当する。）の収集を行った。タグの収集には3.1で前述したなろう API を利用した。小説家になろう全体の傾向を明確にするため、サイト内の小説を網羅的に収集を行い、661,495件の小説に設定されたタグ3,833,470個を収集した。タグの収集期間は、2019年6月27日である。

1つあたりの小説にどのくらいタグが設定されているかを調査した。図5.1.1はその結果である。図5.1.1から読み取れるように、タグが4つ設定されている小説が最も多くなっていることがわかった。また、設定されているタグの平均値を求めると5.80であった。以上のことから、タグはおよそ5個あればその小説の性質を表すことができると考えられる。そこで、本研究で効果的なタグとして推薦するタグの個数を5個とした。

小説家になろうでは、様々な種類のタグが存在する。具体的な種類は3章で挙げた通りである。本研究では、これらのタグをその性質によって、自由タグとシステムタグという2つの枠組みを作成、その2つに大別した。自由タグとは、著作者（ユーザ）の手によって決められる、自由度の高いタグのことである。小説家になろうにおいては、手動入力キーワードを自由タグに分類した。システムタグとは、サイト側が用意したタグのことである。手動入力キーワード以外のキーワード、登録必須キーワードをシステムタグとした。システムタグは、サイト側が用意したという特徴から、普遍的なもので多くの人にとって既知であったり、ある要素を含むものをゾーニングさせるためのものであったりするものが多い。図5.1.1は、小説家になろう上のタグの出現頻度を、すべてのタグ、自由タグのみそれぞれの様子をあらわしたものである。比較してみると、システムタグが上位に存在していることがわかる。これは、システムタグは選択式のため、情報量が小さいものが多いと推測でき、検索の絞り込みには向いていない。以上から、本実験で検索に効果的なタグとして推薦するものは、自由タグから選ぶものとした。

図5.1.1を見ると、自由タグのみの出現頻度も全体の出現頻度と同様に分布している様子がわかる。そのため、システムタグを取り除き自由タグのみを推薦する場合でも、情報量の極めて大きいものと小さいものは、推薦候補から除外する必要がある。本実験では、タグの出現頻度が10,000より大きいもの、出現頻度が5回未満であるものを、それぞれ情報量の極めて大きいものと小さいものとした。

推薦候補のタグに、Tから始まる数字8桁のコード、Tコードを識別子として与えた。タグの出現頻度から、タグの情報量を求めた。求めた情報量は、Tコードに紐づけて管理した。

5.1.2 小説本文

本研究では、小説の特徴量（ベクトル）を生成するために小説の本文を用いる。本研究では、「小説家になろう」に投稿された小説のうち大ジャンル「恋愛」に区分されているジャンル「異世界」に分類される小説を対象とした。小説本文の収集には、Beautiful Soupを用いた。Beautiful Soupとは、HTMLやXMLファイルからデータを抽出することができるスクレイピングに特化したPythonライブラリである。小説本文の収集期間は、2019年7月18日から2020年1月22日である。上記期間の中で、サイトの負荷を考慮し断続的に収集を行った。

5.2 ベクトル生成

5.1で準備したデータセットを基に小説本文及びタグのベクトルを生成する。

5.2.1 本文のベクトル生成

5.1.2節で収集した小説本文の特徴量を求めるために、本文のベクトル生成を行った。タグの推薦に用いるためには、小説とタグが2部グラフの関係である必要がある。したがって、収集した本文のうち、タグと2部グラフの関係にない、すなわち、タグが1つも設定されていない小説を取り除いた。また、小説を収集するときには、分類される大ジャンル・ジャンルでのみ、制限をかけているため、小説本文が数行程度など、非常に短い作品もみられた。これらの小説は、ほかの小説と比べて特徴量を抽出するために、分量が十分であるとはいえない。そのため、小説の本文が1,000文字以上の小説を対象とした。以上の条件を満たす小説本文26,904件の特徴量を求めた。

まず、小説本文の形態素解析を行った。形態素解析には、形態素解析エンジンであるMeCab [10]を用いる。形態素解析では、本文から取り出す単語を、名詞、形容詞、動詞、形容動詞の4つの品詞を取り出した場合と、上記に加えて連体詞、副詞、接続詞、感動詞の合計8つの品詞を取り出した場合の2通りで行った。このパターンをそれぞれ前者をパターン1。後者をパターン2と呼称する。この2通りの品詞取り出しパターンを比較することで、小説の特徴量をより適切に表すのに必要な品詞について比較する。品詞を取り出すときには、原形に統一し取り出しを行った。オンライン小説は、幅広い年代の人々が執筆し投稿されているため、型にはまらない様々な表現が用いられている。オンライン小説の内容から適切に特徴量を抽出するために、形態素解析には新語に対応可能な辞書であるNEologdを用いた。ストップワードとしては、SlothLib [11]が提供しているストップワードリストを利用した。

形態素解析した本文は、文書から分散表現を求める手法であるDoc2Vec [12]を用いて300次元のベクトルを生成した。小説本文の文量の差が、小説本文の特徴量へと影響を与えないようにするために、単位ベクトルへと変換した。

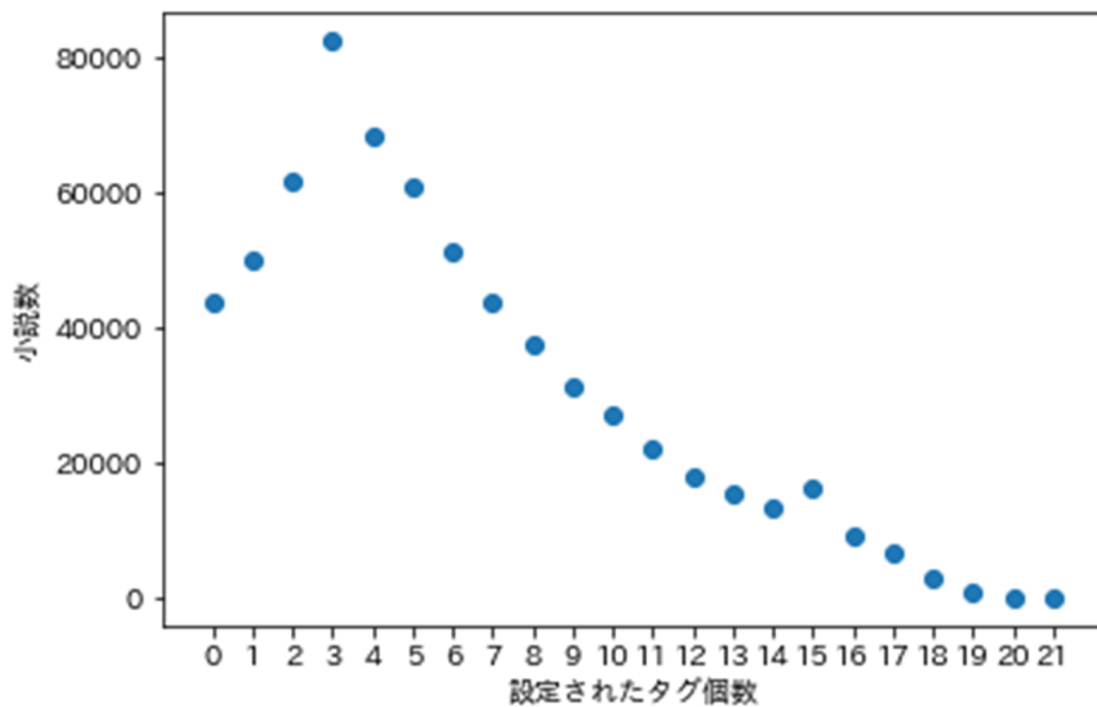


図 5.1: 1つの小説あたりに設定されたタグ数

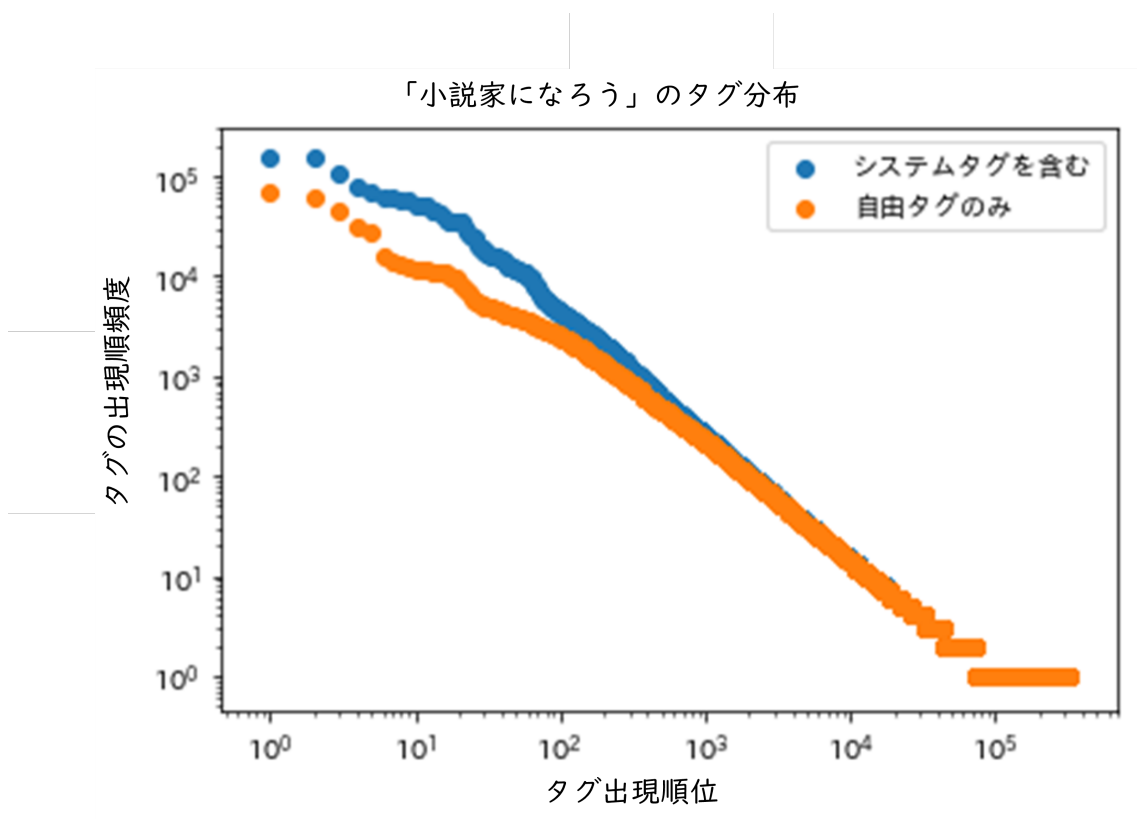


図 5.2: 「小説家になろう」内のタグの出現頻度分布

表 5.1: 小説ベクトルと手法 A の推薦タグベクトルの Cos 類似度平均値

推薦順	パターン 1	パターン 2
1	0.726	0.707
2	0.593	0.578
3	0.579	0.566
4	0.568	0.559
5	0.531	0.524

5.2.2 タグのベクトル生成

5.2.1 で生成した小説の本文ベクトルをもとにタグのベクトル生成を行った。タグのベクトルもその出現頻度に特徴量が影響を受けないように単位ベクトルへと変換を行った。タグのベクトルは、5.1.1 で作成したタグの識別子である T コードに紐づけることで管理を簡潔に行えるようにした。

5.3 タグの推薦実験

5.2.2 節で生成したタグのベクトルと、小説本文のベクトルを類似度比較し、検索に効果的なタグの推薦を行った。大ジャンル「恋愛」・ジャンル「異世界」に属する小説から無作為に選んだ 10 件の小説を対象とした。小説本文は、5.2.1 節と同様な手順で特徴量を算出した。それぞれ手法 A、手法 B ごとに推薦実験をパターン 1、パターン 2 それぞれで行う。

5.3.1 手法 A

手法 A の手順に従い、各小説にタグの推薦を行った。表 5.1 は、それぞれ小説本文のベクトルと推薦したタグのベクトルの Cos 類似度を求め、全体で平均したものである。表 5.1 のそれぞれパターン 1 とパターン 2 で、Cos 類似度に大きな違いは見られない。Cos 類似度は 0 から 1 の値であり、値が 1 に近づくほどそれらのベクトルは類似しているといえる。すなわち、Cos 類似度が高いほど、小説とタグは同様な性質を持っていると考えられる。そこで、手法 A で推薦したタグと各小説に著作者の手によって設定されたタグ（著作者タグ）と比較を行うことにした。著作者タグは、小説を作成した本人によって設定されたタグであるため、小説の持つ性質と同じ性質のタグであると考えられる。比較の結果を、表 5.2 に示す。表 5.2 から両パターンで、著作者タグと合致したタグが推薦できていることがわかる。このことから、手法 A で推薦したタグの中には小説と同様な性質を持っているタグが含まれていることが分かった。パターン 1 とパターン 2 を比較すると、パターン 2 のほうが著作者タグと合致したタグを推薦できていることが読み取れる。著作者タグと合致した推薦タグの小説本文との Cos 類似度は、表 5.1 と比較し平均値より高いものもみられたが、Cos 類似度が高くても著作者タグと合致しないタグも見られた。したがって、Cos 類似度が高いタグが著作者タグと合致するとは一概にいえない。

手法 A で推薦したタグ同士で異なった性質を持っているかを検証するために、小説と推薦したタグのユークリッド距離を算出し、多次元尺度構成法により可視化を行った。図 5.3.1 と図 5.3.1 は、N コードが N7805EI の小説に対して推薦されたタグの距離を可視化したものを一例として挙げる。大きな傾向を把握することは難しいが、タグ同士である程度離れて

表 5.2: 手法 A の推薦タグと著作者タグの比較

	パターン 1	パターン 2
総合致タグ数	6	11
最大合致数	3	4
合致小説数	4	6
非合致小説数	6	4

表 5.3: 小説ベクトルと手法 B 推薦タグベクトルの Cos 類似度平均値

α	パターン 1			パターン 2		
	0.1	0.5	0.9	0.1	0.5	0.9
推薦順 1	0.514	0.511	0.647	0.500	0.503	0.633
推薦順 2	0.516	0.517	0.504	0.502	0.500	0.500
推薦順 3	0.521	0.525	0.507	0.508	0.517	0.502
推薦順 4	0.534	0.521	0.514	0.513	0.506	0.503
推薦順 5	0.518	0.512	0.512	0.514	0.507	0.497

位置していることが読み取れる。このことから、手法 A ではタグ同士でタグの表す情報に重複がないタグを推薦できたと考えられる。

5.3.2 手法 B

手法 B でも 5.3.1 節で手法 A に行った実験及び検証を行った。手法 B では、タグの特徴量の類似度とタグの情報量のどちらに重点を置くかパラメータ α で決定する。 α は、0.1, 0.5, 0.9 の 3 つの値に分けて実験を行った。

表 5.3 は、それぞれ小説本文のベクトルと推薦したタグのベクトルの Cos 類似度から平均を求めた結果である。手法 A と比べて、全体的に Cos 類似度の値が小さくなっていることがわかる。パターン 1 とパターン 2 による目立った差異は見受けられない。また、全体的に 0.5 付近の値を示している。

手法 B で推薦したタグと著作者タグの比較も行った。その結果を、表 5.4 に示した。パターン 1 とパターン 2 での大きな差異は見られなかった。手法 A の結果である表 5.2 と比べると、著作者タグと合致した推薦タグが少なくなっている。しかし、検証の際に手法 A では、著作者タグと合致する推薦タグがなかった小説に対して、著作者タグと合致したタグを推薦できていたケースもあった。

小説と手法 B で推薦したタグのユークリッド距離を算出し、多次元尺度構成法によってマッピングをした。それぞれ、形態素解析の品詞パターンと α の値ごとに図を作成した。図 5.3.2~5.3.2 が作成した図のうち N コードが N7805EI の小説に対して推薦されたタグをマッピングしたものである。これらは共通して、小説とタグの位置が離れたところに位置しているように読み取れる。タグ同士は、小説との位置関係と比較して近い位置に存在している。 α の値が 0.9 の時、比較的小説の座標の近い位置にタグが存在していることがあるが、それ以外のタグは、離れた座標にある。これらのことから、手法 B では推薦されたタグ間で異なる性質のタグを推薦をできたかは明らかではなかった。

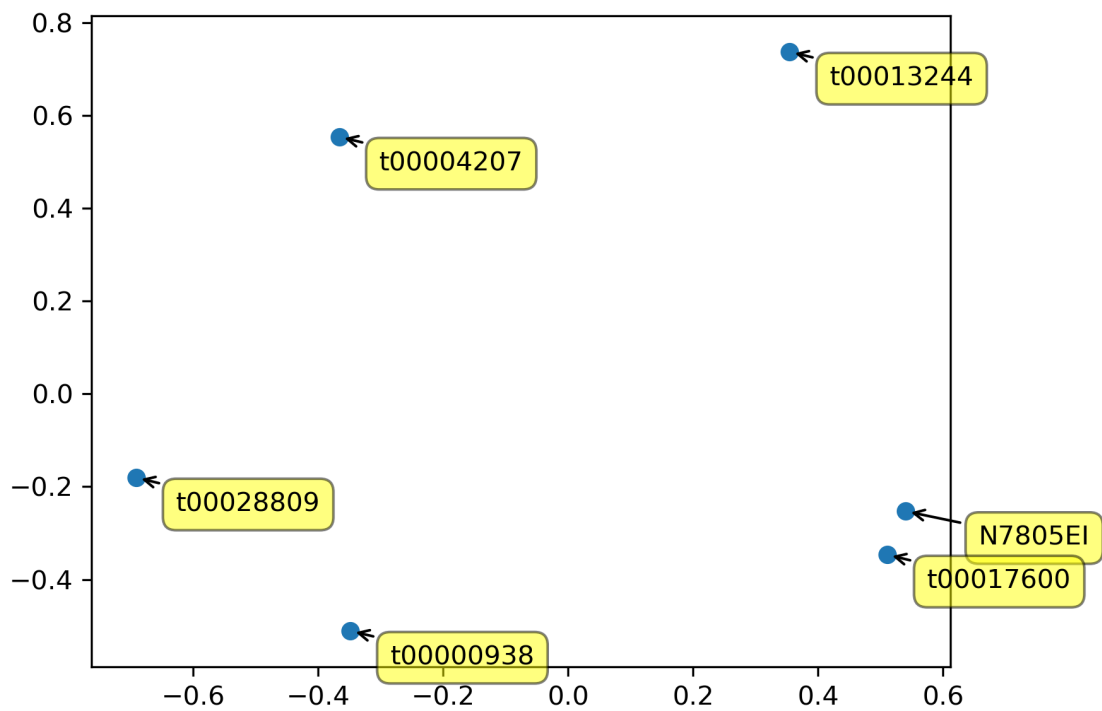


図 5.3: 小説 N7805EI と手法 A パターン 1 で推薦されたタグの距離

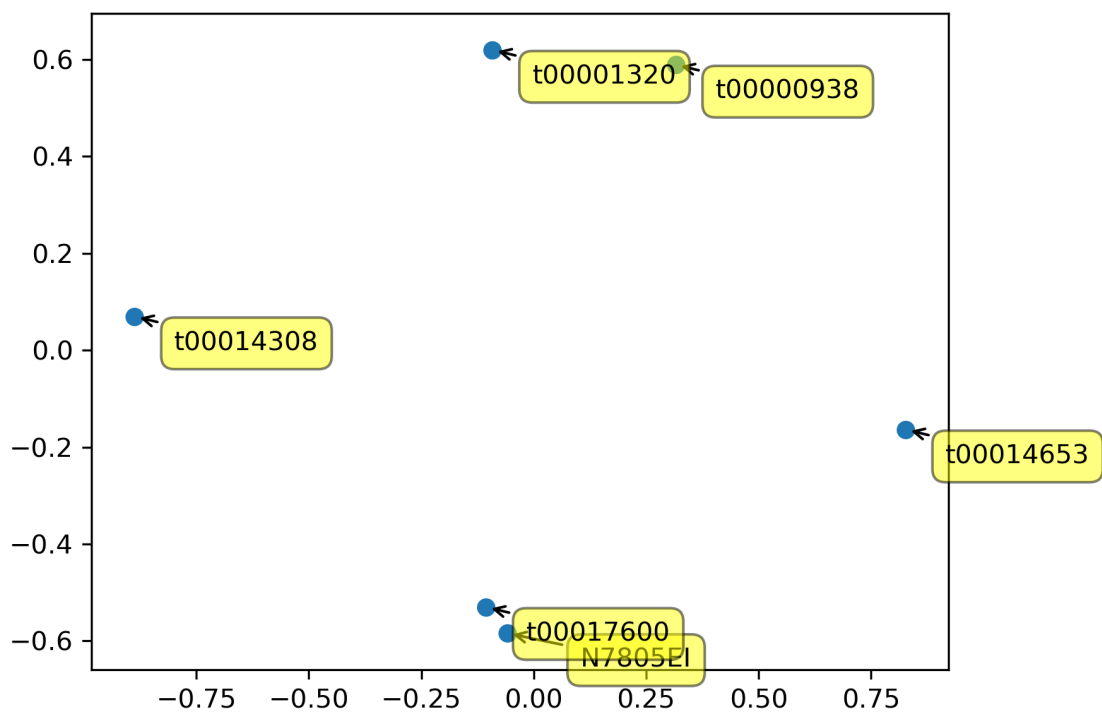


図 5.4: 小説 N7805EI と手法 A パターン 2 で推薦されたタグの距離

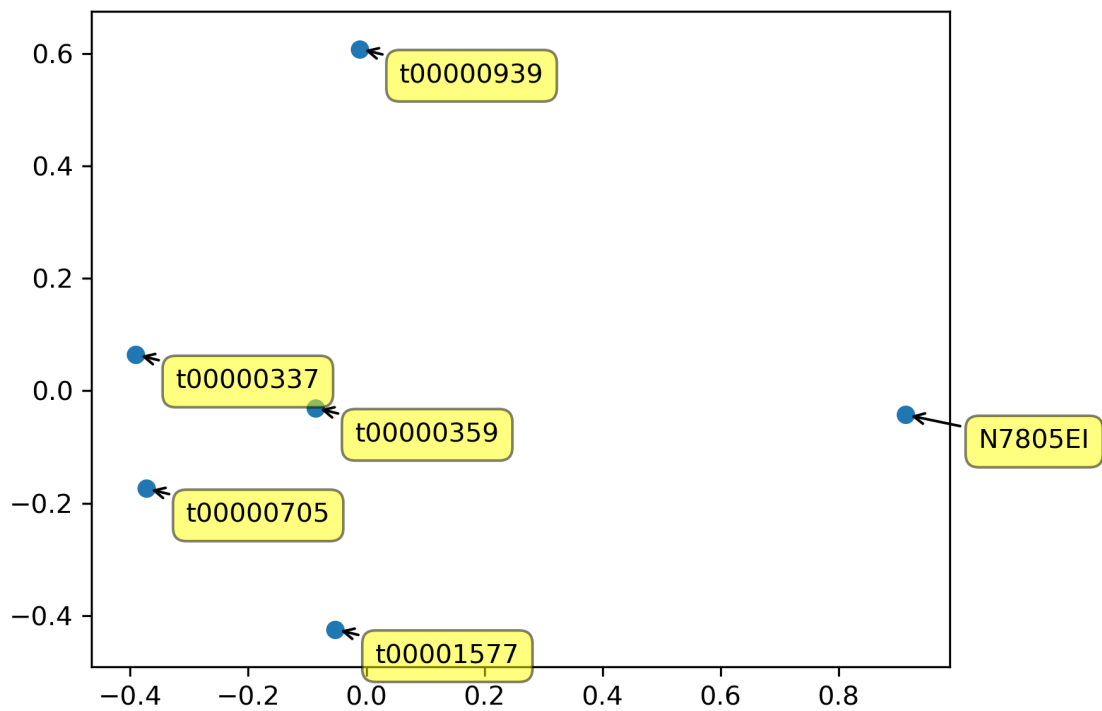


図 5.5: 小説 N7805EI と手法 B パターン $1\alpha = 0.1$ で推薦されたタグの距離

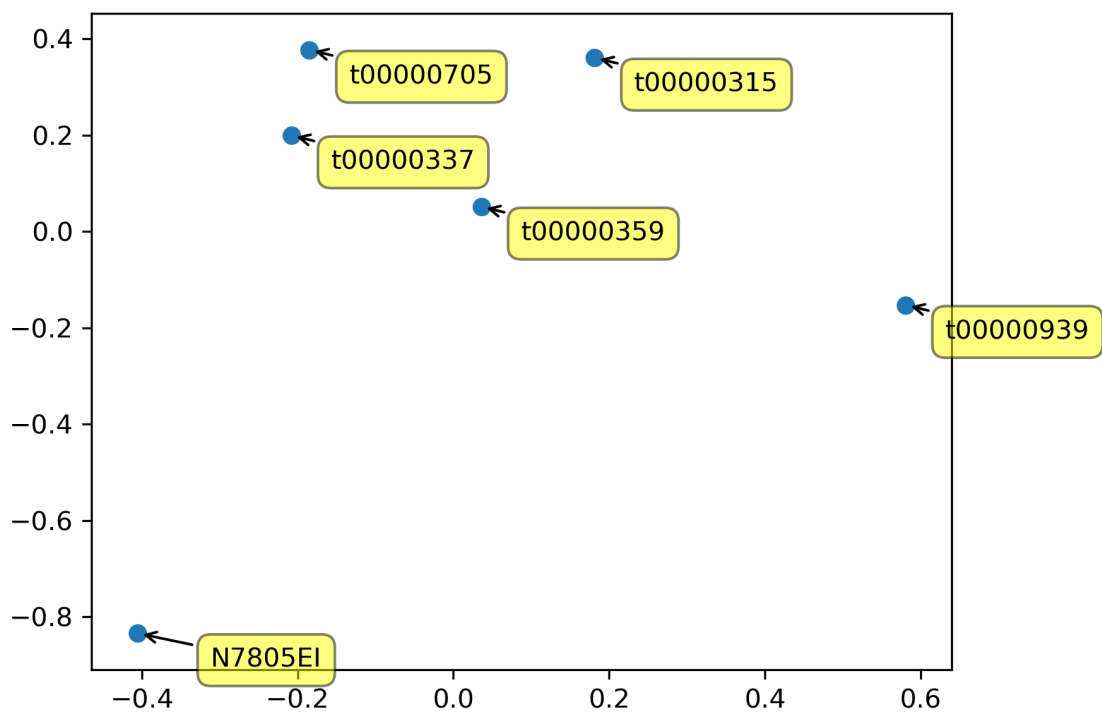


図 5.6: 小説 N7805EI と手法 B パターン $1\alpha = 0.5$ で推薦されたタグの距離

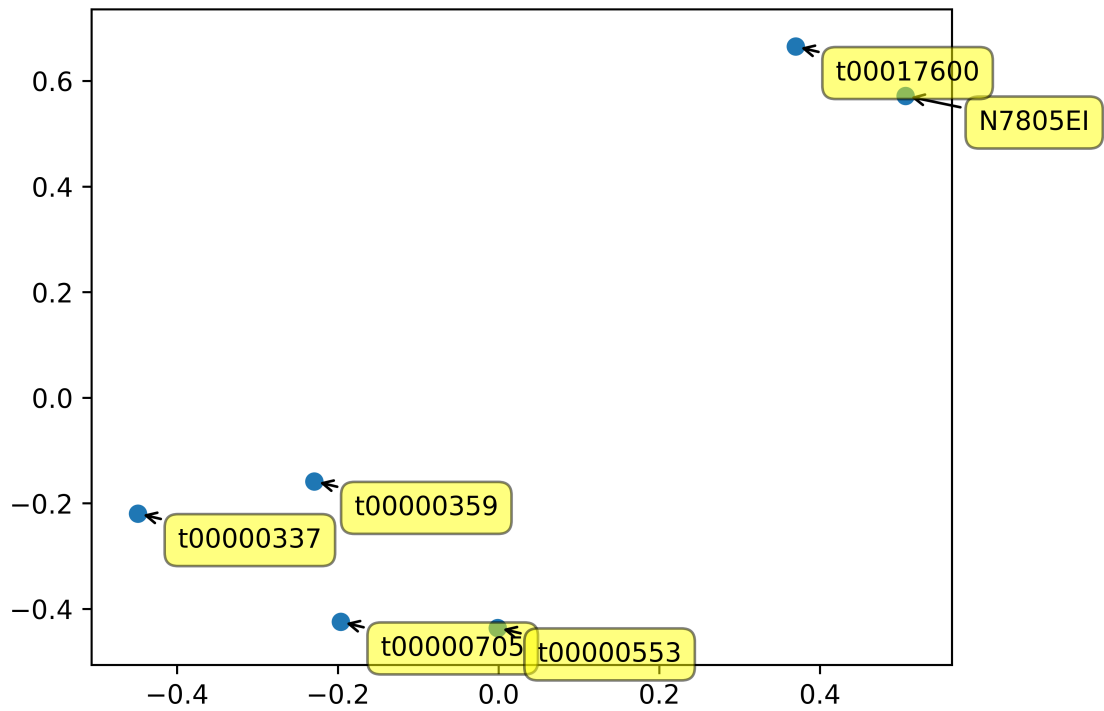


図 5.7: 小説 N7805EI と手法 B パターン $1\alpha = 0.9$ で推薦されたタグの距離

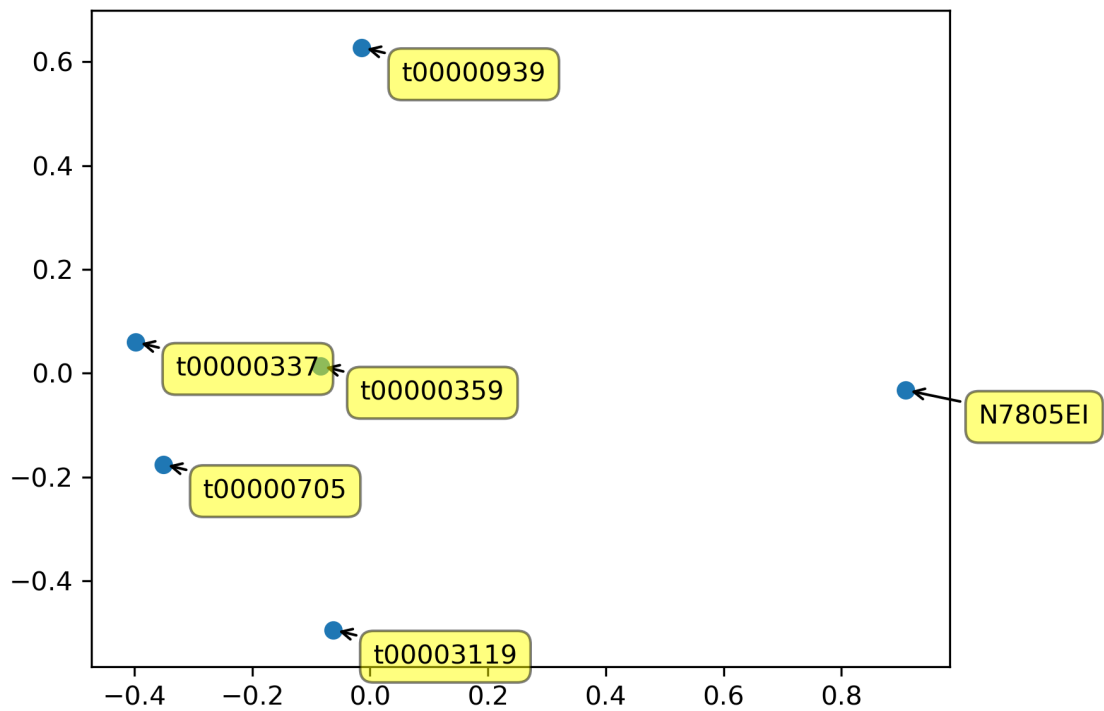


図 5.8: 小説 N7805EI と手法 B パターン $2\alpha = 0.1$ で推薦されたタグの距離

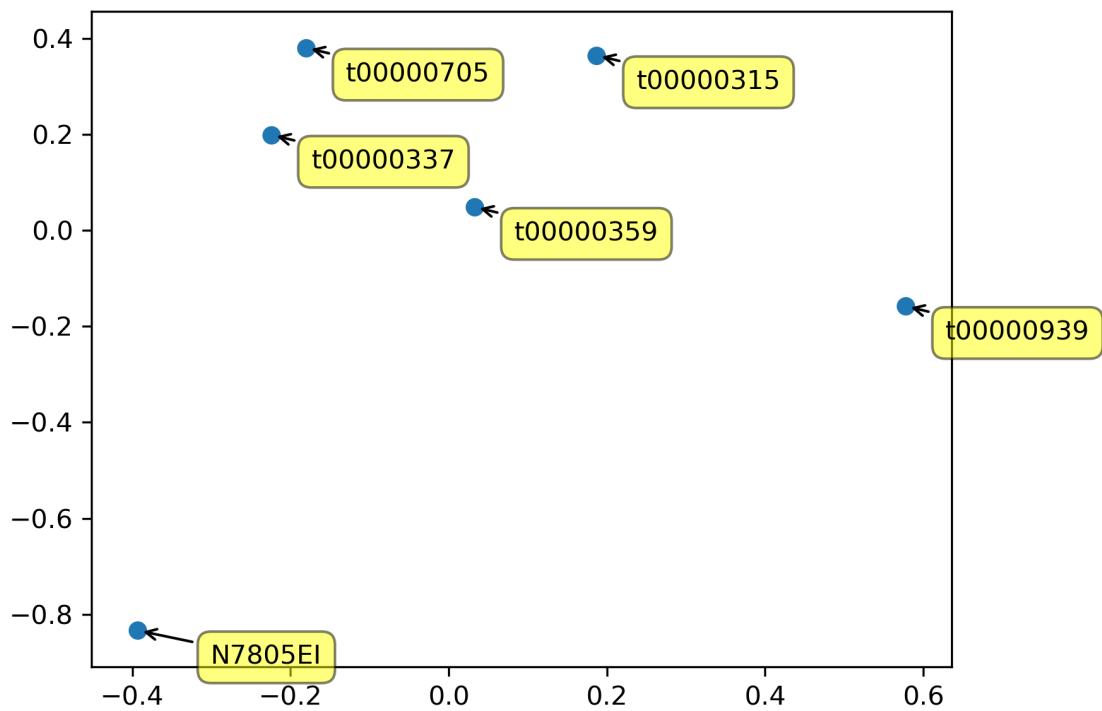


図 5.9: 小説 N7805EI と手法 B パターン $2\alpha = 0.5$ で推薦されたタグの距離

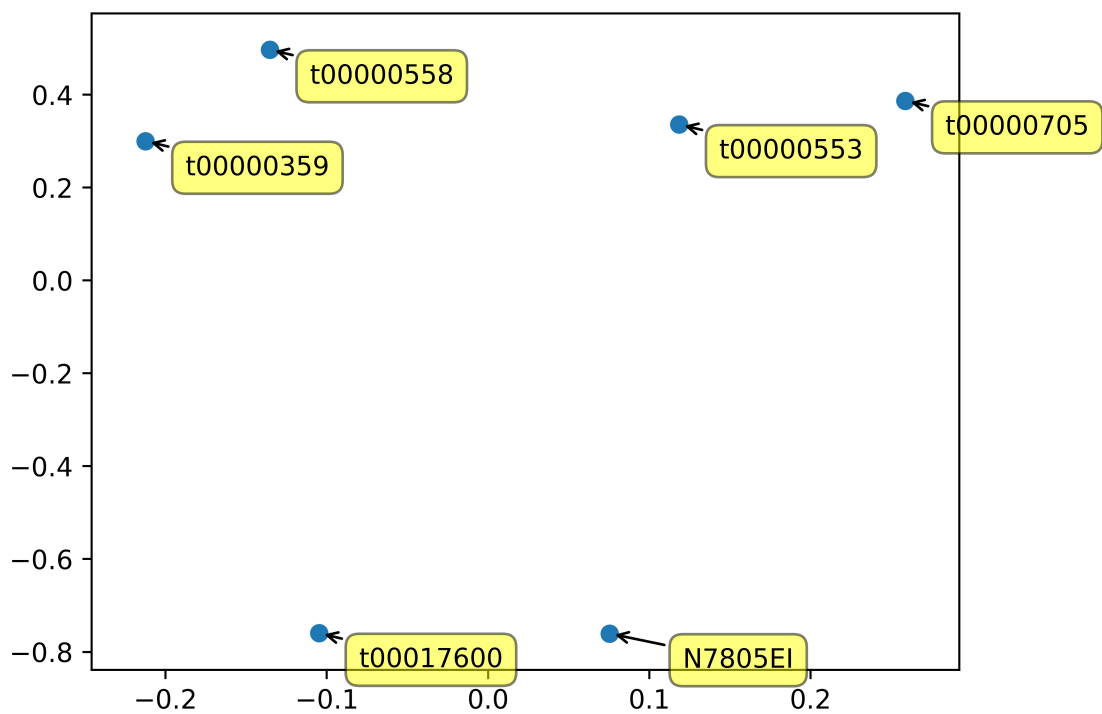


図 5.10: 小説 N7805EI と手法 B パターン $2\alpha = 0.9$ で推薦されたタグの距離

表 5.4: 手法 B の推薦タグと著作者タグの比較

α	パターン 1			パターン 2		
	0.1	0.5	0.9	0.1	0.5	0.9
総合致タグ数	4	4	5	4	4	5
最大合致数	3	2	3	2	2	3
合致小説数	4	3	3	3	3	3
非合致小説数	6	7	7	7	7	7

5.4 考察

本節では、5章で行った実験・評価について考察を行う。Mecabの形態素解析で用いた品詞によって2つのパターンで分けてタグの推薦を行ったが、推薦したタグのCos類似度には大きな違いは見られなかった。一方で、パターン2、名詞、形容詞、動詞、形容動詞、連体詞、副詞、接続詞、感動詞の8つの品詞を特徴量に用いたときのほうが、著作者タグと合致したタグを推薦できていたため、より小説の内容に沿ったタグを推薦できていたと推測できる。5.3節の実験では、入力の小説を無作為としてしまったため、小説の性質による推薦の違いなどの検証はできていない。文量や同じシステムタグが設定されているなどの制約をかけて分類した小説を入力とすることで、どのような小説に提案手法が有効であるかの検証が可能となる。また、著作者タグの個数もそれぞれで異なっていたため、統一することでより正しく評価ができると考えられる。

本研究では、小説の特徴量を算出するために小説本文を用いた。実験において小説の特徴量、それから作成したタグの特徴量によって、検索に効果的なタグの推薦が一部であるができたので、特徴量としては十分であると考えられる。しかし、小説本文は、作品によっては非常に多い分量を持つものもあり、事前の処理や小説本文を入力とする本手法では、時間や計算コストがネックになると考えられる。本文よりも短く小説について簡潔に書かれているあらすじを用いる方法や、ユーザが読了した小説の特に好んだ個所を抜粋して入力するなど、小説の特徴量の抽出とコストのバランスを考える必要がある。

第6章 おわりに

本研究では、小説投稿サイトに投稿されている小説を読了したユーザに対して、その読了した小説と類似した要素を持つ小説を検索するために効果的なタグを推薦する手法について、提案を行った。

具体的には、小説本文と小説に付与されているタグを行い、個々の小説の特徴量（ベクトル）は、小説本文に Doc2Vec を適用し算出した。一方、タグの特徴量は、そのタグが設定されている 2 部グラフの関係にある小説のベクトルを合成して算出した。本研究における検索に効果的なタグとは、入力された小説の内容と合致したタグ、推薦されたタグ同士でタグの表す情報に重複がないタグ、複数のタグで絞り込みができるタグの 3 つの要件を満たすタグとした。1 つ目は、入力された小説と Cos 類似度の高いタグを、2 つ目は、タグ同士で特徴量を比較し、互いに異なるタグを推薦することとした。3 つ目の要件は、出現頻度を用いて情報量を算出し、絞り込みに適さない情報量を持つタグを除外することとした。提案手法を小説家にならうに投稿されているオンライン小説に適用・評価し、提案手法の有効性を検証した。限定された範囲の小説に関してはあるが、効果的な検索タグを推薦できることができた。

今後の課題としては、提案手法の評価の改善、特に推薦タグの検索における絞り込みの効果の検証が十分ではないため、さらなる検証を行いたい。また、今回実験でもちいた小説と異なるジャンルにも手法を適応させ、その結果を考察することで小説投稿サイト全体でのユーザの検索行動を支援していくことが重要である。

謝辞

指導教員の佐藤哲司教授をはじめ，松村敦助教授には多岐にわたり熱心なご指導をいただきました。ここに記して，心から感謝申し上げます。また，研究生活および学生生活を共に過ごしたコンテンツ工学研究室の皆様にも感謝いたします。

参考文献

- [1] 高田 叶子, 佐藤 哲司. 文体の類似度を考慮したオンライン小説推薦手法の提案. 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, DEIM2017, B5-2, 高山グリーンホテル 岐阜県高山市, March 3 - 9, 2017. IEICE データ工学研究専門委員会.
- [2] 飯田委哉, 伊東栄典, 佐嘉田悠樹. クラスタリングによるオンライン小説の多様性動向分析. 火の国情報シンポジウム論文集, Vol. 2018, pp. 1-7, 2018.
- [3] 清水一憲, 伊東栄典, 廣川佐千男. 集合知に基づくオンライン小説のランキング手法の提案と評価. 情報処理学会研究報告, pp. B-3, 2013.
- [4] 伊藤志暢, 松村敦, 宇陀則彦. 語の結束度と感情を考慮したオンライン小説の段落分割手法の提案. 第81回全国大会講演論文集, Vol. 2019, No. 1, pp. 437-438, 2019.
- [5] 実崎直人, 伊東栄典. 回帰分析によるオンライン小説の人気度推定. 第81回全国大会講演論文集, Vol. 2019, No. 1, pp. 333-334, 2019.
- [6] 井上 優作, 若林 啓. 表記の多様性を考慮したハッシュタグ推薦. 第8回データ工学と情報マネジメントに関するフォーラム, DEIM2016, B6-5, ヒルトン福岡シーホーク 福岡県博多市, February 29 - March 2, 2016. IEICE データ工学研究専門委員会.
- [7] Xueting Wang, Yiwei Zhang, and Toshihiko Yamasaki. User-aware folk popularity rank: User-popularity-based tag recommendation that can enhance social popularity. In *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 1970-1978. ACM, 2019.
- [8] Mutsuki Yamazaki, Kazufumi Inafuku, and Tetsuji Satoh. Tag recommendation method for enhancing web novel retrieval. In Tokuro Matsuo, Kunihiro Takamatsu, Yuichi Ono, and Sachio Hirokawa, editors, *9th International Congress on Advanced Applied Informatics, IIAI-AAI 2020, Kitakyushu, Japan, September 1-15, 2020*, pp. 43-48. IEEE, 2020.
- [9] Genkou Ou, Kei Wakabayashi, and Tetsuji Satoh. Searching behavior analysis of online shopping based on information content of query words. In *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*, pp. 43-48, 2019.
- [10] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to Japanese morphological analysis. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 230-237, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [11] 大島 裕明, 中村 聡史, 田中 克己. Slothlib web 検索研究のためのプログラミングライブラリ. 日本データベース学会 letters, Vol. 6, No. 1, pp. 113-116, 06 2007.

- [12] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, Vol. 32 of *Proceedings of Machine Learning Research*, pp. 1188–1196, Beijing, China, 22–24 Jun 2014. PMLR.