

コミックの登場人物についての説明文からの性格タグ推定

樋口 亮太 山西 良典 松下 光範

関西大学総合情報学部

{k896930, ryama, m_mat}@kansai-u.ac.jp

概要 本稿では、ストーリーに基づいたコミックの検索を目指し、ストーリーの構成要素のひとつであるキャラクターの性格に着目する。ストーリーは、「どのようなキャラクター」が「どのように活躍するか」で捉えられると考えた。このうち、「どのようなキャラクター」に相当するキャラクターの性格などを示した文章は Web 上に存在するが、その記述は作品によって多様であるため、説明文自体をクエリとしてコミックの検索に用いることはできない。一方で、「どのようなキャラクター」であるかを作品を超えて同一の基準で端的に表すために性格タグを用いることもある。しかし、性格タグの付与には膨大なコストと性格タグに対する解釈の曖昧性が存在する。本稿では、単語分散表現を用いて表現した Web 上のキャラクター説明文と性格タグの関連性を機械学習によりモデル化し、説明文からの性格タグを推定する。推定結果の分析から、キャラクターの説明文の記述の傾向と性格タグとの関係性を考察する。

キーワード コミック工学, ストーリー情報, キャラクターの性格, コンテンツ検索

1 はじめに

コミックの発行部数は毎年1万点以上にも及び、膨大である。この膨大なコミックの中からユーザーが読みたい作品を選択する方法として、電子コミックや書籍販売のウェブサイトでは、コミックの検索サービスがある。現状の代表的なコミック検索法では、技術書や雑誌などの検索と同様に、書誌に関する情報 (e.g., タイトル, 作者, 掲載誌) やジャンル (e.g., ラブコメ, アクション・アドベンチャー, ヒューマンサスペンス) などがクエリとして利用されている。しかし、コミックは、ストーリー性をもったコンテンツであり、これらのメタ的な情報だけではコンテンツの内容を表現できているとは言えない。現状のメタ情報のみを用いた検索の枠組みでは、ストーリー情報 (例えば、「熱血なキャラクターが敵を倒すストーリー」や「優しいキャラクターが仲間を救うストーリー」) に基づいた検索は実現されていない。レビュー情報を用いてコミックの内容に基づいた検索 [3] も提案されているが、レビュー情報はコミックに対する読者の感想でありコンテンツの内容そのものを扱っているわけではない。また、レビュー情報にはネタバレが含まれる可能性 [1] もあり、コミックの内容そのものにフォーカスした検索手法が必要であると考えられる。

ストーリーに基づいたコミック検索を実現するために、まずコミックのストーリー性を構成する要素について検討した。本研究では、「ストーリーは「どのようなキャラクター」が「どのように活躍するか」で捉えられると考えた。本稿では、このうち「どのようなキャラクター」に着目する。キャラクターを説明する上では、性格や生い立ちなどの内面的な情報と、見た目などの外見情報が用

いられる。このうち性格は、キャラクターがどのように行動選択を行うかの大きな要因の一つとなり、ストーリー展開に大きく影響する可能性が高いと考えられる。性格に着目した研究の一つとして、朴らは性格診断に用いられるの一つであるエゴグラム¹の5つの性格的特徴を用いて、各性格特徴に対応した単語群を用意した辞書を作成し、キャラクターを説明する文章 (以下、キャラクターの説明文) から性格分類を行った [2]。しかしながら、あらかじめ著者らが用意した辞書を用いたアプローチであるため、辞書に含まれない多用な表現で記述されたキャラクターの説明文に対しての性格分類の有効性においては疑問が残る。また、性格もエゴグラム¹の5つの性格特徴を用いたベクトル表現としているが、様々なコミックのキャラクターの特徴を表現するうえで、5次元ベクトルで十分であるかについても検討しなければならない。ファンコミュニティの中では、キャラクターの性格を端的に表すために、性格タグが用いられることもある。タグでの表現により、異なる作品に登場するキャラクターの性格を多様性を担保しつつも、多くの人の中で共通の理解で評価可能になる。そこで、本稿では Web 上に複数存在するキャラクターを説明文そのものを入力として、ファンコミュニティの中で共通理解が得られていると考えられる性格タグを推定する。

提案手法では、単語分散表現を用いて表現した Web 上のキャラクター説明文からコミックキャラクターの性格タグを推定する。キャラクターの説明文と性格タグの関連性をモデル化できれば、未知のキャラクターに対しても説明文からタグを自動的に付与することが可能になる。これにより、キャラクターの性格をコミックの検索に導入し、ストーリーに基づくコミック検索の端緒となることを目指す。

Copyright is held by the author(s).

The article has been published without reviewing.

表1 キャラクターの説明文についての情報抽出源とそこから得られた説明文数.

自由参加型ウェブ百科事典	説明文数
Wikipedia ¹	535
ニコニコ大百科 ²	441
pixiv 百科辞典 ³	538
アニヲタ wiki (仮) ⁴	490
合計	2,004

表2 キャラクターに付与された頻出上位 20 件の性格タグと対応する説明文数.

性格タグ	説明文章の数
強気	1,059
異性が好き	1,038
優しい	905
俺	831
色恋に興味はある様子	742
気の強さは普通	685
プライド高い	670
精神年齢が高い	604
堂々とした	598
精神年齢が若い	562
まとも	530
ちょっと変	514
私・わたし	509
社交的	486
普通の上品さ	473
S	423
世話焼き	416
お調子もの	370
落ち着いた	356
紳士的	344

2 データセット

Web 上からキャラクターの説明文と性格タグを自由参加型ウェブ百科事典からそれぞれ収集し、データセットを構築した。キャラクターの説明文については、複数の自由参加型ウェブ百科事典から横断的に収集した。

キャラクターの説明文については、表 1 に示すように、複数の自由参加型 Web 百科事典を情報リソースとして、コミックのキャラクターについて記述している説明文を取得した。得られたキャラクターの種類数は 537 人であり、複数の情報リソースに記述されているキャラクターもあれば、いずれかの情報リソースにしか記述されていないキャラクターもいた。同じキャラクターについて説明していても、情報リソースによって説明文の記述には大きな差異が見られる。例えば、「ジョジョの奇妙な冒険」における「空条承太郎」について Wikipedia では設定に関する客観的な情報に関する記述が多い一方で、アニヲタ wiki (仮) では、

『道とは自分で切り開くもの』などその行動の基盤には確固たる信念があり、いかなる詭弁や恫喝を向こうに回しても一切のブレが無い。

のように、キャラクターのセリフが引用されて説明されている。そこで、本稿では同一のキャラクターを説明していたとしても異なる情報リソースから得られた説明文については、それぞれ異なる記述特徴を有した独立の説明文として扱うこととした。

キャラクターの性格タグについては、キャラ属性王国⁵を情報リソースとして収集した。キャラ属性王国では、コミックのキャラクターを説明するために、「ステータス」「容姿」などの外見情報や、「背景」「能力」などの内面的な情報をカテゴリとして様々なタグが付与されている。本稿では、このうちキャラクターの性格に関連する「基礎性格」「性格」のカテゴリに分類されているタグを収集した。上述の説明文が取得できたキャラクターの性格を表現するためには、合計で 225 種類のタグが用いられていた。本稿では、得られた性格タグのうち、表 2 に示す頻出上位 20 件の性格タグを対象とした。1 人のキャラクターには複数の性格タグが付与されているため、各説明文は 20 種類の性格タグのそれぞれが付与されたかバイナリで表現された 20 次元の性格ベクトルを持つことになる。

3 提案手法

提案手法では、以下の手順でキャラクターの説明文から性格タグの推定を行う。

1. 単語分散表現によって、説明文を多次元ベクトル化する
2. 入力を上記 1 で得られた多次元ベクトル、出力を性格タグを示す 20 次元のベクトルとしてキャラクターの説明文と性格タグの関係モデルを学習する。
3. 学習したモデルを用いて、キャラクターの説明文から性格タグを示すベクトルを推定する。

下節では、手順 1 および 2 の詳細について説明する。

3.1 単語分散表現の獲得

説明文の多次元ベクトル化に用いる単語分散表現を獲得するために、まず 2 章で用意した説明文 2,004 件を学習した。ここで、学習した全 2,004 件の説明文には、42,412 文、1,051,036 単語が含まれていた。学習に際して、説明文を形態素解析によって単語単位に分割し、その表層形を学習した。形態素解析器には MaCab (ver.0.996)、辞書には NEologd (ver.0.0.7)⁶を用いた。辞書 NEologd には、有名なコミックのキャラクター名も含まれている。

⁵<https://chara-zokusei.jp>

⁶<https://github.com/neologd/mecab-ipadic-neologd>

表 3 各性格タグの推定における 5 分割交差検証での平均評価：Precision, Recall, F1 値.

性格タグ	平均 Precision	平均 Recall	平均 F1 値
強気	0.699	0.724	0.699
異性が好き	0.718	0.712	0.718
優しい	0.658	0.621	0.658
俺	0.674	0.678	0.674
色恋に興味はある様子	0.640	0.654	0.640
気の強さは普通	0.577	0.603	0.577
プライド高い	0.575	0.664	0.575
精神年齢が高い	0.575	0.608	0.576
堂々とした	0.508	0.520	0.509
精神年齢が若い	0.488	0.597	0.488
まとも	0.518	0.606	0.519
ちょっと変	0.454	0.502	0.454
私・わたし	0.557	0.606	0.557
社交的	0.456	0.545	0.457
普通の上品さ	0.444	0.535	0.444
S	0.482	0.548	0.483
世話焼き	0.475	0.562	0.476
お調子もの	0.417	0.532	0.385
落ち着いた	0.417	0.532	0.418
紳士的	0.425	0.520	0.426

単語分散表現の学習では、ライブラリ gensim の W2V (ver.3.8.3)⁷ を利用した。このとき、単語分散表現の学習時の window サイズは 5、ベクトルの次元数は 100 とした。

3.2 説明文と性格タグの関係モデルの学習

キャラクターの説明文と性格タグの関係モデルを学習するためには、3 層のニューラルネットワーク (NN) を用いた。NN の学習には、ライブラリ PyTorch⁸ を用いた。

ネットワークの構成は、第 0 層 (入力層) : 100, 第 1 層 (隠れ層 1) : 500, 第 2 層 (隠れ層 2) : 250, 第 3 層 (出力層) : 20 とした。入力層には、3.1 節で得られた 100 次元の単語分散表現ベクトル、出力層には、20 次元の性格タグベクトルがそれぞれ対応する。推定においては、各性格タグ毎に推定モデルを構築する方法も考えられるが、推定モデルの構成については今後の検討課題とする。

4 実験

提案手法によって、キャラクターの説明文の記述から性格タグを推定可能であるか実験した。下節にて、実験設定と実験結果、考察を述べる。

4.1 実験設定

実験では、説明文 2,004 件を対象として、5 分割交差検証を行った。このとき、交差検証ごとに NN の学習モデルを構築し直し、テストデータには学習に用いていない説明文を用いた。

学習した NN では、説明文を入力すると、その説明文

に対応する性格タグの 20 次元のベクトルを出力する。実験の評価では、性格タグそれぞれについての Precision, Recall, マイクロ F-1 値を算出した。本稿では、これらの評価指標の 5 分割検証実験における平均値によって推定性能について議論することとする。

4.2 結果

表 3 に、推定結果を示す。F1 値を見てみると、「強気」「異性が好き」「優しい」「俺」などのタグについては比較的高い平均 F1 値が得られた。一方で、「お調子もの」「落ち着いた」「紳士的」などのタグについては、比較的低い F1 値の推定結果となった。

これらの結果から、高い推定性能を示すとは言いえないものの、多様な性格タグをキャラクターの説明文を入力として用いることで推定できる可能性が示唆された。一方で、分散表現の獲得や関係モデルを構築するネットワークの構成や学習方法などについては、さらなる検討が必要であると考えられる。

4.3 考察

表 3 の結果をもとに、キャラクターの説明文の記述について考察した。本稿で推定対象とした性格タグでは、対義関係にあるタグ「精神年齢が高い」と「精神年齢が若い」のペアや「お調子もの」と「落ち着いた」のペアが存在したり、どちらも第一人称である「俺」と「私・わたし」などがあつた。また、「ちょっと変」のような一見して性格の詳細とは紐つかないタグも存在した。

「俺」と「私・わたし」のペアはサンプル数の違いを考慮しても、「俺」と「私・わたし」の平均 F1 値の差が大きい。例えば、「俺」という性格タグが付与されているキャラクターには「鬼滅の刃」に「嘴平伊之助」が登場

⁷<https://radimrehurek.com/gensim/>

⁸<https://pytorch.org/>

する。このキャラクターの Wikipedia には、「戦う相手が居ない場でもその闘争心が収まらないらしく、大声を張り上げたり意味も無く木の幹に体当たりしたりしている。」という強気の男性とわかる記述がある。一方で、「私・わたし」という性格タグが付与されているキャラクターには女性にも多いが、中性的な男性にも付与されていることがある。例えば、「私・わたし」という性格タグは「七つの大罪」に登場する「エリザベス」に対して、「HUNTER × HUNTER」に登場する「クラピカ」にも付与されている。「俺」の性格タグが付与されているキャラクターは比較的強気の男性キャラクターに限定されるが、「私・わたし」の性格タグが付与されているキャラクターは、性別が混在しており、説明文の内容が限定されにくいため、推定結果に差が出たと考えられる。

「ちょっと変」という性格タグは全体的に低い推定結果になっている。この性格タグが付与されているキャラクター説明文には、他にも様々な性格タグが付与されている場合が多く、内容的にもまとまりがないことが低い原因だと考えられる。例えば、「ちょっと変」という性格タグが付与されているキャラクターの「新世紀エヴァンゲリオン」に登場する「碇シンジ」の pixiv 百科事典には「やや内省的で繊細な性格。自らの存在意義に思い悩んでおり、苛酷な状況に追い詰められた際などは極めて情緒不安定に陥る事も。」という記述がある。また、同じ作品に登場する「葛城ミサト」にも「ちょっと変」という性格タグが付与されている。「葛城ミサト」の Wikipedia には「私生活においては、非常にがさつかつ、ずぼらでだらしない面が多い。」と記述されている。同じ作品の同じ性格タグが付与されているキャラクターの説明文であっても、その説明の記述は大きく異なる。このような性格タグは説明文からの推定が難しいことがわかった。

5 おわりに

本稿では、キャラクターの説明文とキャラクターの性格タグをウェブ上の異なる複数の情報ソースから収集し、それらを組み合わせることによってコミックのコンテンツであるキャラクターの性格を推定した。キャラクターの説明文とキャラクターの性格タグの関係モデルの構築では、単語分散表現を用いて表現した Web 上のキャラクター説明文と性格タグの関係性を獲得した。実験の結果、多様な性格タグに対して、一定の推定性能で説明文から性格タグが推定できる可能性が示唆された。

提案手法による推定性能について考えると、推定器のネットワークの構成やパラメータの調整など検討すべき点は多い。これらについては、今後の課題とする。また、性格がストーリーにどのような影響を与えるのかについて調査し、コミックのストーリー検索の実現を目指す。

謝辞

本研究は、一部、文科省科研費基盤 C#20K12130 の助成のもと行われた。本稿の執筆にあたって、関西大学総合情報学部小林達哉の協力を得た。記して謝意を表す。

参考文献

- [1] 佐藤剣太, 牧良樹, 中村聡史: 未読および既読シーンの提示が読者のコミック閲覧意欲に与える影響, 情報処理学会研究報告 (EC), Vol. 2018-EC-47, No. 3, pp. 1-8 (2018).
- [2] 朴柄宣, 居林香奈枝, 松下光範: エゴグラムに基づいたコミックキャラクターの性格分類, 人工知能学会全国大会論文集, 1J3-02 (2018).
- [3] 山下諒, 朴柄宣, 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).