

性格要素と外見要素の加減算による類似キャラクターの検索

小林 達哉^a 松下 光範^b

関西大学総合情報学部

a) k252475@kansai-u.ac.jp b) mat@res.kutc.kansai-u.ac.jp

概要 本研究の目的は、ユーザの好みのキャラクターが登場するコミックを検索する技術の実現である。現在広く用いられているコミック検索では、作品名や著者名、ジャンル情報などコミックに付与された情報をクエリとしているが、コミックに登場するキャラクターの性格や外見といった情報は、コミックの選択に大きく影響するにも関わらず、それに基づく検索は十分に検討されていない。これを解決するための一助として、本稿ではキャラクターを表現する性格要素 (e.g., 真面目, 優しい) や外見要素 (e.g., 金髪, 長身) をクエリとして利用する検索手法を提案する。提案手法では、これらの要素を既知のキャラクターに加減算できるようにして好みのキャラクターを表現することで類似キャラクターの検索を可能にする。

キーワード コミック, キャラクター, 検索, 分散表現, IDF

1 はじめに

全国出版協会・出版科学研究所の調査 [5] によれば、コミックの新刊発行部数は毎年 1 万点以上にも及ぶ。日々増加を続けるこの膨大なコミックの中からユーザの興味や関心に合致した作品群に出会うのは困難であるため、そのような探索を支援するシステムの実現が求められている。

現状では、この膨大なコミックの中からユーザがコミックを検索しアクセスする手段として出版社のサイトや電子書籍販売サイト (e.g., コミックシーモア¹, Book-Live²) の検索機能が利用されることが多い。こうしたアクセスを行う際、ユーザはジャンル情報 (e.g., 異世界, 恋愛), 作品名, 著者名などをクエリとしてコミックを検索する。

しかし、これらのサービスではコミックに登場するキャラクターの性格や外見といった情報などは、コミックの選択に大きく影響するにも関わらず、キャラクター情報に基づく検索は難しい。例えば、「熱血で優しいキャラクターが登場するコミック」や「天才で知的なキャラクターが登場するコミック」のような、内容に関わる情報をクエリとした検索は現状では行われていない。こうした問題を解消するため、本研究では内容情報に基づくコミック検索の実現を試みる。

松井らはコミック検索支援システムの一つとして、コミックの登場キャラクターに着目した検索の手法を提案した [2]。この手法では、ユーザの既知のキャラクターを基に性格 (e.g., 真面目, 優しい) や外見 (e.g., 金髪, 短髪) 要素を加減算することで「コナンに熱血を足したキャラクター」のような検索を可能にしている。しかし、

この研究ではキャラクターの要素を数値化する過程において、対象とするキャラクターがもつ性格や外見要素などを one-hot で表現しているためにそれらの間に重要度の差がなく、些細な特徴が影響して意図に沿った検索結果が必ずしも得られない。そこで、抽出された要素のうちどの要素がそのキャラクターを表現するうえで重要な要素かを反映可能にする必要がある。

こうした背景の下、本稿ではキャラクターを構成する要素として性格要素と外見要素に着目し、それぞれを分けて数値化することでこの問題の解決を図る。

2 先行研究

朴らはキャラクター情報に基づくコミック検索を実現するために、Web 上から抽出したキャラクターの説明文を用いてエゴグラムによる性格分類を試みた [4]。この手法では東大式のエゴグラム [6] によりコミックのキャラクターの性格を五次元のベクトルで表現し、キャラクターの性格を推定した。被験者実験により、キャラクターの性格とユーザが認識するキャラクターの性格が一致するか検証したところ、55%の精度で推定できたことが報告されている。この結果は、キャラクターの性格に基づく類似度検索が可能であることを示しているが、現状では十分な精度とは言い難い。

松井らは性格情報だけでなくキャラクターの外見情報も必要であるとし、萌え要素の加減算に基づくキャラクター検索手法を提案した [2]。この手法では、ユーザの既知のキャラクターを基に要素を追加・削除することで「コナンに熱血を足したキャラクター」のような検索を可能にしている。しかし、この手法では性格や外見などの要素を区別せずに one-hot で数値化しているため、キャラクターを認識するうえで重要な特徴と些細な特徴が区別できず、検索者の意図に沿った検索結果が得られないことが

Copyright is held by the author(s).

The article has been published without reviewing.

¹<https://www.cmoa.jp/>

²<https://booklive.jp/>

ある。

この問題を解決するため、本稿ではキャラクターの性格と外見を分けて数値化することで、抽出された要素のうちどの要素がキャラクターにとって重視されているかを表現可能にする。

3 デザイン指針

本稿では、既知のキャラクターを基に性格と外見の要素を加減算することによって、類似キャラクターの検索を可能にすることを目的とし、以下の3項目に取り組む。

- (1) キャラクターの性格・外見に基づく数値化
- (2) キャラクターの性格・外見の加減算に基づくキャラクター検索
- (3) キャラクターの演算を可能とするプロトタイプシステムの実装

(1) ではキャラクターを数値で表現するための性格・外見に基づく数値化の手法である。(2) では(1)の数値化手法に基づきキャラクター要素の加減算を行う。これにより、例えば「キャラクターA + キャラクターB」の演算を行うことで、それぞれの特徴を持ち合わせたキャラクターを検索することが可能になる。また、「キャラクターA + 要素」を行うことにより、「コナンに熱血を足したキャラクター」のように演算を行うことができる。(3) ではキャラクター要素の加減算をするためのプロトタイプシステムを実装し、ユーザ実験を通じて提案手法の有効性や限界について検証する。

4 データセットの作成

データセットの作成は松井らの研究[2]を参考に萌え要素に着目して行った。本稿では、萌え要素のうちキャラクターの構成要素である性格と外見要素を対象にした。処理手順は(1)要素の収集、(2)辞書の作成、(3)要素の付与、の順に行った。各々の処理について以下に述べる。

4.1 要素の収集

萌え要素については、外見に基づく要素だけでなく性格の要素を含める[1]。これら萌え要素を表現する語は一個人の印象で確立されるものではなく、複数人の主観的な印象で構築される。そこで、要素のデータセットを構築するために複数のユーザが自由に書き込める自由参加型 Web 百科事典を対象とし要素を収集することにした。自由参加型 Web 百科事典はインターネットに接続することで誰もが記事を編集できる Web サイトである。本稿では Wikipedia³、アニヲタ Wiki⁴、ニコニコ大

百科⁵、ピクシブ百科事典⁶のサイトを利用した。自由参加型 Web 百科事典では、萌え要素・萌え属性は同義語として扱われているため、各自由参加型 Web 百科事典で「萌え要素」または「萌え属性」として扱われている語を合計で 1403 語収集した。そこから重複する表現や公序良俗に反する表現を削除し、要素を表す 1172 語を対象とした。

ここで、萌え要素は性格と外見の要素が含まれているだけでなく、表現単体で特徴を表していないもの(e.g., うなじ, 髪の色)、キャラクターの性格と外見の特徴を表していない表現(e.g., 温泉回, 世界観)がある。そのため、著者を含めて合計7人のアノテーション作業により、性格及び外見を端的に表現した単語(e.g., 明るい, 冷静, 金髪)、短文(e.g., 世話好き, 情が深い)のみを抽出した。抽出する条件として、4人以上が性格または外見要素であると判断した場合その要素を抽出することとした。これにより85語の要素を削除し、1087語をキャラクターの性格または外見の特徴を表している要素の表現として採用した。

4.2 辞書の作成

収集した要素を対象に、表2のように表記揺れの統一を行った。漢字・ひらがな・カタカナの表記揺れが存在する表現については、各表現ごとに weblio 辞書と自由参加型 Web 百科事典の記事を確認し、表記揺れとして記載されている表現を一つのカテゴリに統一した。例えば、統一前の「たれ目, ”たれ目, ”“垂目, ”“垂れ目”」は意味的に同じであるが、漢字・ひらがな・カタカナの表記揺れにより複数の単語が存在することになるため、カタカナに統一した。このような統一処理を行うことで1087語の要素表現を578語(性格単語数:286語, 外見単語数:292語)に絞り込んだ。

4.3 要素の付与

キャラクターに対して要素を付与するため、本稿では「漫画全巻ドットコム」⁷の歴代発行部数の中から200作品に登場する1438キャラクターを対象とし、自由参加型 Web 百科事典を用いてキャラクターの説明文が含まれた文章を収集した。収集した文章を形態素解析である MeCab⁸を用いることにより「名詞, 形容詞」を抽出する。この時、固有名詞などを充実させた mecab-ipadic-neologd⁹の辞書を用いることによる表現の統一を行った。抽出された単語の中から本稿で作成した萌え要素の辞書に該当する単語をそのキャラクターに付与し、付与された一例を表1

⁵<https://dic.nicovideo.jp/>

⁶<https://dic.pixiv.net/>

⁷<https://www.mangazanken.com/ranking/books-circulation.html>

⁸<http://taku910.github.io/mecab/>

⁹<https://github.com/neologd/mecab-ipadic-neologd>

³<https://ja.wikipedia.org/wiki/>

⁴<https://w.atwiki.jp/aniwotawiki/>

表1 キャラクタに要素を付与させた時の一例

キャラクタ名	うずまきナルト	孫悟空
性格	目立ちたがり, わがまま, 冷静, 明るい, 負けず嫌い	明るい, 穏やか, 素直, 優しい, 純粋
外見	金髪, ゴーグル, ツンツン, 碧眼, 長身	筋肉質, ツンツン, 糸目, 小柄, 尻尾

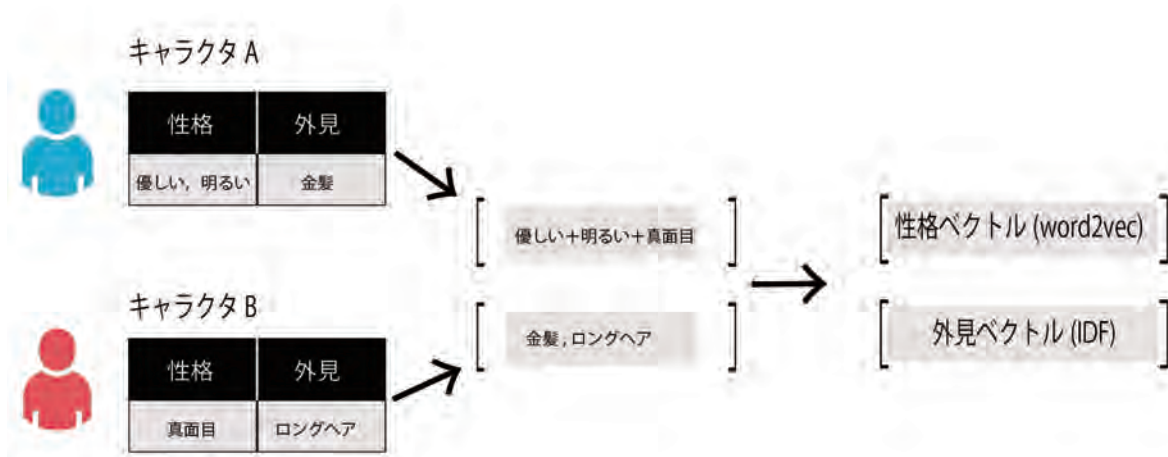


図1 キャラクタ同士の演算によるベクトル作成

表2 表記揺れ統一例 ([2] より表引用)

統一前	統一後
タレ目, たれ目, 垂目, 垂れ目	タレ目
狐耳, きつね耳, キツネ耳	狐耳
ロングヘアー, ロングヘア, 長髪	ロングヘア

に示す.

5 提案手法

本稿ではキャラクタを構成させる要素のうち性格と外見に着目して数値化を行い, 要素を用いて加減算することで該当する別のキャラクタを検索する. 加減算することで, 「キャラクタ A に熱血を足したキャラクタ」や「キャラクタ A とキャラクタ B を足し合わせた」ような性格や外見に着目した類似キャラクタ検索を可能とする(図1 参照).

5.1 性格要素のベクトル化

キャラクタの性格は複数の側面を持ち合わせて構成されているため, 加減算した単語の意味を複数考慮して性格を表現する必要がある. そこで, 性格要素の意味を反映することができる単語の分散表現を用いて, キャラクタに単語ベクトルを加えることにより複数の性格の意味を反映させて性格を表現する.

以上の考えに基づき, キャラクタの性格要素は word2vec の手法により単語の分散表現でベクトルを獲得する. ここで, ベクトルの作成には Skip-gram[3] を用いた. これにより文章の語順に対応した学習が可能となり, 周辺

の単語分布が似ている単語ベクトルが似た値をとるようになる. ただし, キャラクタに付与した性格要素は辞書に該当する単語を抽出しているため語順の関係がないため, 前後関係を考慮するような学習方法は不適切となる. そこで, 語順学習をさせないため学習途中で性格要素の順序をランダムに変化させ語順の関係性を考慮させないよう学習した. word2vec の実装には python のライブラリである gensim (ver.3.8.3)¹⁰を用いた. パラメータは, 次元数を 100 次元, ウィンドウサイズを 8 単語, 学習 epoch 数を 100 に設定した. 性格要素の語順をランダム化させる時は epoch 数を 5 毎にランダム化した.

cos 類似度に基づく学習単語の類似単語結果の一例を表3に示す. この結果を見ると, “明るい” は「“元気,” “天真爛漫,” “快活”」などの単語が高い類似度を示していることが確認できる. これらの語について weblio 辞書¹¹で類似語検索を行い類語として分類されている語を確認したところ, 入力単語である “明るい” という単語が類語として存在することが確認できた.

5.2 外見要素のベクトル化

キャラクタの外見は, 場面や状況などのコミックの進行状況により衣装が変更されるように, 外見要素の意味を考慮して構成されていない. そのため, 要素の意味を反映して性格を表現する単語分散表現の手法は不適切である. そこで, キャラクタの外見を表現するために, 対象とするキャラクタが持つ外見要素を数値化する.

¹⁰<https://radimrehurek.com/gensim/>

¹¹<https://thesaurus.weblio.jp/>

表 3 単語の類似結果

単語 順位	明るい		冷淡		真面目		ムードメーカー	
	類似単語	類似度	類似単語	類似度	類似単語	類似度	類似単語	類似度
1	元気	0.675	非情	0.754	堅物	0.706	人懐っこい	0.610
2	天真爛漫	0.672	無慈悲	0.647	頑固	0.566	明るい	0.542
3	快活	0.612	冷徹	0.606	優柔不断	0.507	快活	0.536
4	人懐っこい	0.604	冷静	0.552	律儀	0.481	明朗	0.537
5	活発	0.601	慎重	0.483	強情	0.464	男気	0.479

本稿では、キャラクターの外見要素に対して特徴語を考慮するために、「多数のキャラクターに登場する要素は、キャラクターの特徴語にはならない」という仮定に基づき、IDF法を用いることとした。この手法により、少数しか出現しない要素を特徴語として利用できる。

6 プロトタイプシステムの実装

6.1 システム構成

本稿で示したデータ作成及び提案手法に基づきキャラクター検索のプロトタイプシステムを構築した。システムの実装にはフロントエンドとしてVue.js (ver.2.6.11)¹²、バックエンドとしてDjangoRestFrameWork (ver.3.11.0)¹³、データベースとしてMySQL (ver.8.0.21)¹⁴を利用した。処理手順を図2に示す。

6.2 インタフェース

本稿で作成した検索フォームは2つ存在し、1つ目がキャラクター検索であり2つ目が要素検索のフォームである。キャラクター検索フォームに、ユーザの既知のキャラクターのフルネーム及び一部の名前を検索フォームに入力することにより、名前に該当したキャラクターが検索候補として出力される。検索候補として提示されたキャラクターをクリックすることによりキャラクターの持つ性格と外見の要素が可視化される(図3参照)。その中からユーザが加減算する要素を選択するために、要素をクリックすることで加減算を可能とさせた。要素をクリックすることで提示されている色が変更され、黄色が加算、青色が減算となる。

キャラクターの検索後サーバ側にキャラクターの演算式が送信される。サーバで要素同士の加減算が行われ、cos類似度が高い上位10キャラクターが検索キャラクターとしてユーザに提示される。その際に、提示されるのは「性格と外見、性格のみ、外見のみ」のそれぞれで計算された上位10キャラクターである。タブを切り替えることにより、それぞれ算出されたキャラクターを確認することができる。検索されたそれぞれのキャラクターは詳細ボタン

をクリックすることにより自由参加型Web百科事典から収集された文章とキャラクターに付与された要素を見ることができる。また、ピクシブのサイト¹⁵を利用することにより、任意のユーザが投稿したキャラクターのイラスト画像のページを閲覧可能となる。

7 実験

本稿では、キャラクターの構成要素の性格と外見に着目した加減算を行うことにより、類似キャラクターの検索が可能であるかを実験により確認する。実験では、上述のように構築したシステムをユーザに利用してもらうことによりキャラクター検索の有効性を検証した。実験協力者は普段からコミックを読む学生を対象とした。

7.1 実験手順

実験協力者は、関西大学総合情報学部在籍する学生の20名(男性10人、女性10人)であった。実験に先立ち、実験協力者に対して(1)実験の目的、(2)提案システムの操作説明、(3)実験課題の3つの項目を説明した。提案システムの操作説明では、キャラクター検索(e.g., 入力フォームにキャラクターの一部の名前を入れると検索できる、キャラクターの名前をクリックすると要素が可視化され取捨選択できる)を説明した。また、検索されたキャラクターに対して、自由参加型Web百科事典のキャラクター情報を閲覧できること、キャラクターをピクシブ百科事典のイラストページを用いることで、キャラクターのイラスト画像を確認できることを伝えた。これにより、全ての実験参加者が課題に用いるシステムの機能を理解している状態にした。

本実験の課題は「ユーザの既知のキャラクターを基に性格と外見の加減算に基づくキャラクター検索の有効性の検証」である。この点について有効性を示すため提案システムを実験者に利用してもらった。実験者にはまず、検索キャラクターとしてあらかじめ用意されたコミックの主人公15人の中から、ユーザ既知のキャラクターを演算してもらい、次に、用意された以外のキャラクターのうち実験者の既知のキャラクターを演算してもらった。演算では「キャラクター同士」と「キャラクター+要素」の2つの演

¹²<https://jp.vuejs.org/index.html>

¹³<https://www.django-rest-framework.org/>

¹⁴<https://www.mysql.com/>

¹⁵<https://www.pixiv.net/>

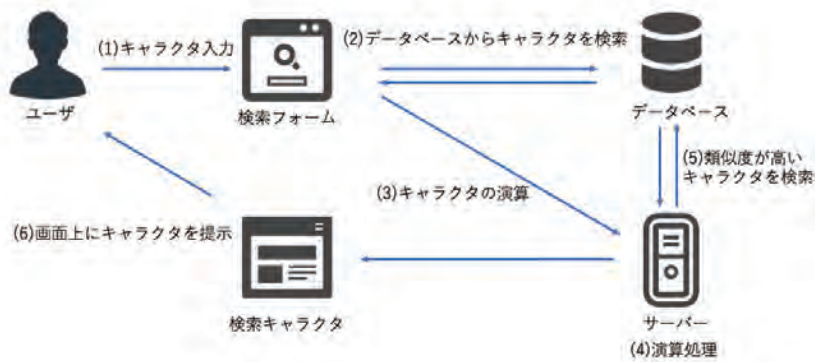


図2 作成したプロトタイプシステムの処理手順

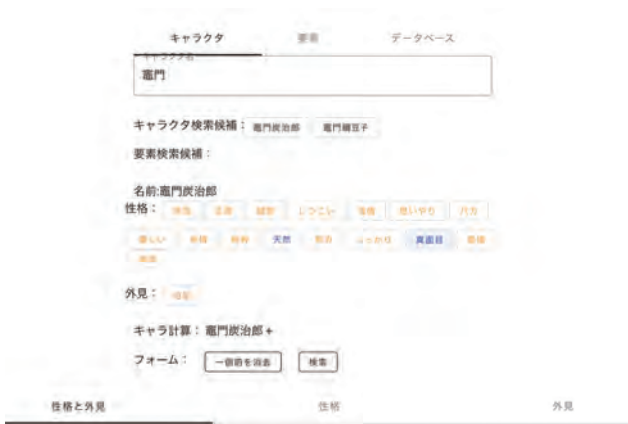


図3 プロトタイプシステムの検索フォーム

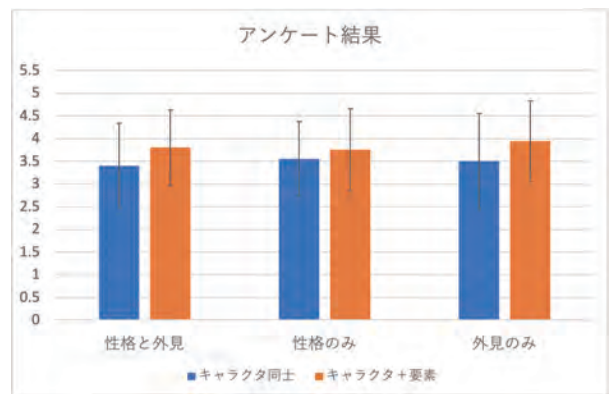


図4 ユーザ実験によるアンケート結果

算を行うように指示した。実験中、演算の結果検索されたキャラクターに関して思ったことや感じたことなどについてあれば、自由にメモを記述するように促した。実験では、システムを利用する時間を20分経過した時点で終了とした。実験終了後、検索されたキャラクターに対して検索精度についてアンケートを回答してもらった。アンケートは「キャラクター同士」の演算において演算したキャラクターの双方の特徴が反映されていたか、「キャラクター+要素」において演算した要素の特徴が反映されていたか、についてそれぞれ5段階評価で回答してもらった。アンケートには「キャラクター同士」と「キャラクター+要素」の両方で「性格と外見、性格のみ、外見のみ」のそれぞれで回答項目を用意した。

7.2 実験結果

「キャラクター同士」と「キャラクター+要素」の演算についてそれぞれのアンケート結果を図4に示す。グラフの結果はユーザの平均を表し、エラーバーは標準偏差である。表4にはアンケート結果の平均と標準偏差を示す。

7.3 考察

ユーザアンケート結果の評価値と自由回答からキャラクターの構成要素である性格と外見の加減算に基づくキャラクター検索の有効性について考察する。

表4を確認すると、「キャラクター+要素」の結果が「キャラクター同士」と比べて「性格と外見、性格のみ、外見のみ」の平均値を0.2ポイント程度上回り、「キャラクター+要素」の方が要素の特徴を反映した結果となった。

「キャラクター+要素」の検索結果が「キャラクター同士」を上回った理由として実験者の自由回答を確認したところ、「性格に関しては少し片方の特徴が反映されたキャラが多く感じた」「どちらか一方のキャラクターに偏っている場合もあったと感じた」「適切なキャラが推薦されたと感じるがたまに一つのキャラクターの特徴に左右されている」という意見が見られた。「キャラクター同士」の演算では両方のキャラクターの特徴ではなく片方のキャラクターの特徴が反映された結果の一部になっていたことが伺えた。これは、「キャラクター同士」の演算を行った時、キャラクターの要素が同等の数キャラクターに付与されていないことが考えられる。また、「外見がぴったり当てはまっても性格があまり演算されていないように感じた」「帽子などの装飾品の要素が強かった」という意見が得られ、出力結果が一部の要素に強く影響されることが原因である。

一方で、「キャラクター+要素」の検索結果では「キャラクター

表4 ユーザアンケート結果の平均

演算方法	性格と外見	性格のみ	外見のみ
キャラクタ同士の演算	3.40 ± 0.94	3.55 ± 0.83	3.50 ± 1.05
キャラクタ+要素の演算	3.80 ± 0.83	3.75 ± 0.91	3.95 ± 0.89

ラに外見の要素を足すとしっかり同じような性格のキャラで外見の違うキャラが出てきてとても精度が高いと感じた。性格の要素を足しても同じように感じた」「外見の特徴がそっくりな「優しい」要素を足したキャラクタを検索した時に想定したキャラクタが一番出てきた」「選択したキャラクタをベースに要素を付け加えるので思った通りのキャラクタが出やすいように思った」というように出力された結果が想定したものになりやすい傾向があった。これらの結果から、「キャラクタ+要素」によって、出力されたキャラクタを想定できることが示唆された。しかし、実験者数が少数であるため実験結果の平均値に大きな差が表れなかった。そのため今後は実験者数を増やすことを検討する。

8 まとめ

本稿では、キャラクタ情報に基づくコミック検索の実現を目指している。その端緒として、キャラクタを構成する要素の中でも性格と外見に着目したベクトル作成を試み、要素を加減算することにキャラクタを検索するプロトタイプシステムを提案した。ユーザ実験より、「キャラクタ+要素」は出力されたキャラクタを想定できることが示唆され、性格と外見要素の加減算の有効性を確認できた。

今後の展望としてプロトタイプシステムの改善を検討する。本稿で行った実験の自由回答からは「外見でも性格でも抽出されたキャラクターが何の要素から導いたものなのかが分からなかった」「キャラクタのどの要素を重視しているかという個人差が大きい」という意見が得られた。構築したプロトタイプシステムではキャラクタに付与された要素を可視化するに留まり、キャラクタが出力される際にどの要素が重視されているかユーザに提示されていなかった。そのため出力されたキャラクタがどのような観点で出力されたかユーザは理解することが困難となる。以上の理由からも、今後はどの要素を重視してキャラクタが出力されたかを理解することができるシステムの構築を目指す。

参考文献

- [1] 倉本到: 萌え擬人化キャラによるインタラクティブシステムの理解促進, 情報処理学会研究報告 (EC), Vol. 2012-EC-23, pp. 1-6 (2012).
- [2] 松井俊樹, 朴炳宣, 松下光範: 萌え要素の加減算に基づくキャラクタの類似度判定手法の提案, 第21回イン

タラクティブ情報アクセスと可視化マイニング研究会, SIG-AM-21-01, pp. 1-6 (2019).

- [3] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality, *CoRR*, Vol. abs/1310.4546 (2013).
- [4] 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の性格分類, 人工知能学会全国大会論文集, 1J3-02 (2018).
- [5] 出版科学研究所: 出版月報 (2020年2月号), 全国出版協会 (2020).
- [6] 和田迪子, 渡部麻美, 市村美帆, 松井豊: Web調査による新しいエゴグラムの尺度開発, 筑波大学心理学研究, Vol. 53, pp. 63-71 (2017).