

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

Determining similarities between comic personages based on addition and subtraction of Moe-factors

松井 俊樹^{1*} 朴 炳宣² 松下 光範¹
Toshiki Matsui¹ Byeongseon Park² Mitsunori Matsushita¹

¹ 関西大学総合情報学部
Faculty of Informatics, Kansai University

² 関西大学大学院総合情報学研究科
Graduate School of Informatics, Kansai University

Abstract: The aim of our study is to calculate similarities between Comic personages by focusing on Moe-factors. “Moe” means strong feelings of affection, often towards an ideal personage in comic and anime. In the existing similarity calculation method between comic’s personages, it is difficult to search a similar personage based on the operation of the constituent factors such as “a comic personage like Conan, but with passionate.” To meet the requirement, this paper proposes a similarity calculation method by adding and subtracting Moe-factors to a known comic personage by focusing on Moe-factor as a component of the user’s interest. Our user study showed that the proposed method enabled calculating similar comic personages with an accuracy of 70.0%.

1 はじめに

近年では、アニメやコミックに登場するキャラクタに対する注目が高まっている。また、東によると現代ではキャラクタの特定の見た目(e.g.,セーラー服)や内面特徴(e.g.,無口)といった断片的な構成要素が重要視されている[1]。キャラクタの構成要素の中でも特に、ユーザの関心を引きつける要素(e.g.,猫耳,ツンデレ)は、この萌えという言葉を用いて、「萌え要素」や「萌え属性」と呼ばれて親しまれており、ユーザのキャラクタの好みに大きく影響を与えている。本稿では、この萌え要素に着目し、ユーザが好むキャラクタを検索する手法の実現を目指す。萌え要素をクエリとしてキャラクタを探す場合、ユーザは自身が好む萌え要素を列挙して、挙げられた要素を持つキャラクタを探す方法が考えられる。しかし、ユーザが必ずしも自らが好む萌え要素を明確に認識しているとは限らず、外在化が難しい場合も存在する[2]。そのため、情報検索手法の研究においては、ユーザの情報要求の曖昧性に対応した検索が可能であることが望ましい[3]。クエリ入力の負担を減らし検索を効率的に行う手段として、ユーザの既知のキャラクタをもとに類似したキャラクタを探す方法がある。例えば「荒っぽく無鉄砲だが、正義感

の強いキャラクタ」が好きなユーザが、好みにあったキャラクタを探す際、すでにその好みに該当するキャラクタを知っているなら、そのキャラクタと似ているキャラクタを探すことで、「荒っぽい」、「無鉄砲」、「正義感が強い」などの要素を列挙する必要がなくなる。本稿では、このような既知キャラクタをもとに類似したキャラクタを探す検索方式に着目する。

類似キャラクタを判別する手法として、朴らは性格診断法であるエゴグラムを用いて類似する性格のキャラクタを判定する手法を提案している[4]。しかし、朴らの手法で扱うことができるのはキャラクタ構成要素のうち性格のみであり、キャラクタの見た目に基づく構成要素は考慮されていない。また、類似度判定の際に「コナンに熱血を足したキャラクタ」のような、既存キャラクタの構成要素に操作を加えた情報を検索の基準にすることができない。

そこで本研究では、キャラクタの内面だけでなく見た目も含めた、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を提案する。

2 関連研究

河谷らは萌えを基準としたキャラクタの検索・分類の方法についてまとめている[5]。この研究では、萌えの対象となるキャラクタの特徴を画像から抽出し、そ

*連絡先：関西大学総合情報学部
〒569-1095 大阪府高槻市霊山寺町 2-1-1
E-mail: mat@res.kutc.kansai-u.ac.jp

の特徴をもとにキャラクタを検索する手法を提案している。キャラクタの顔画像から、視覚的に判別可能な萌え要素である目、顔、髪、の3つの領域を対象とし、減色処理と色クラスタリング処理によって各領域をレイヤーに分け、それぞれの特徴量を求めた。これにより目領域、顔領域、髪領域の特徴が抽出された。その後、抽出した特徴をキャラクタの画像に紐づけることにより、好みのキャラクタと似た特徴を持つキャラクタ画像の検索を実装した。しかし、河谷らの手法は画像をもとにしたものであるため、扱うことができる特徴が見た目で判別できるものに限定されている。

白石らは、絵本に関するレビューの文章から絵本が持つ感情的特徴を抽出し、感情を加味した絵本の分類手法を試みている [6]。まず、WordNet Affect というソーラスで単語間の関係を保持している概念辞書をもとに、喜び、悲しみ、恐れ、怒り、嫌悪、驚きの6つの感情を示す単語から感情単語表を作成した。次に、形態素解析により分割された絵本のレビューに対し感情単語表にある単語との一致数を感情ごとにカウントし、カウント結果を用いて絵本ごとの特徴ベクトルを作成し、クラスタリングによる分類を行なった。この手法による分類の分析の結果、絵本が有する感情の一端を捉えることができた。本稿ではこの手法をもとに、萌え要素の辞書を作成し、キャラクタごとの萌え要素の特徴ベクトルを作成することでキャラクタの類似度判定を行う。

キャラクタ情報に基づくコミック検索の実現を目的として、朴らはキャラクタの構成要素の中でも性格に着目した類似キャラクタ判定手法を検討している [4]。この手法では、性格診断法であるエゴグラムをもとに、性格を表す単語・短文にベクトルを付与した辞書を作成した。次に、ユーザによる編集が可能な自由参加型 Web 百科事典のキャラクタに関する文書から辞書に含まれる語や短文を抽出し、それらの語が持つベクトルの値を集計した。その後、その値をもとに類似度を測りキャラクタの性格タイプを導き出した。朴らの手法による分類の結果を、人間がキャラクタの立場でエゴグラムを用いた性格タイプの判別を行なった結果と比較し、55.0%の分類精度を示した。これにより、自由参加型 Web 百科事典から抽出した情報に基づき作成されたベクトルを用いることで、類似キャラクタの判別が可能であることが示唆された。本稿ではこの手法をもとに、自由参加型 Web 百科事典のキャラクタの説明に関する文書から特徴を抽出を行う。

3 データセットの作成

キャラクタ間の類似度判定を行うにあたり、キャラクタが持つ萌え要素をベクトルにより表現する。その

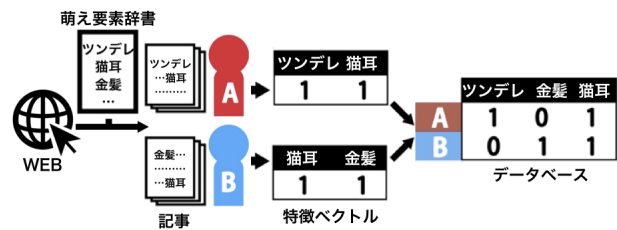


図 1: 特徴ベクトルの作成

ため、キャラクタの内面、見た目の特徴を表す萌え要素を収集し、萌え要素辞書を作成する。次に、キャラクタに関する説明が含まれた文書から、萌え要素辞書に含まれる表現を抽出し、抽出された表現を基底としてキャラクタごとの特徴ベクトルを作成した (図 1 参照)。本章では、上記手順の実装方法について示す。

3.1 対象データ

キャラクタが持つ萌え要素を抽出するため、萌え要素を示す表現を抽出し辞書を作成する。萌え要素を表現する語には、アニメやコミックの文化で生まれた独特の表現 (e.g., ツンデレ, ボクっ娘) が存在する。また、萌え要素は作品の制作者側ではなく、ユーザ側によって見出されていくものである。そのため本稿では、ユーザによる編集が可能であり、独特の表現に関する記述が豊富に存在する自由参加型 Web 百科事典の Wikipedia¹, ニコニコ大百科², ピクシブ百科辞典³, アニヲタ wiki (仮)⁴を対象とし、それらの Web サイトで萌え要素として扱われている表現を収集した。

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収集する。各 Web サイトごとに、本稿における萌え要素の収集条件を以下に示す。

- Wikipedia: category 「萌え属性」に属する語
- ピクシブ百科事典: 「萌え要素」の記事でまとめられた語
- ニコニコ大百科: 「萌え要素・萌え属性の一覧」にまとめられた語
- アニヲタ wiki (仮): 「萌え要素」, 「萌え属性」のタグがついた語

¹Wikipedia (<https://ja.wikipedia.org/wiki/>: 2018 年 10 月 28 日時点確認)

²ニコニコ大百科 (<http://dic.nicovideo.jp/>: 2018 年 10 月 28 日時点確認)

³ピクシブ百科事典 (<https://dic.pixiv.net/>: 2018 年 10 月 28 日時点確認)

⁴アニヲタ Wiki (仮) (<https://www49.atwiki.jp/aniwotawiki/>: 2018 年 10 月 28 日時点確認)

表 1: 表記揺れ統一例

統一前	統一後
タレ目, たれ目, 垂目, 垂れ目	タレ目
ドジっ子, ドジっ娘, ドジっ娘	ドジっ子
地味子, 地味っ子	地味子
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

	減算する要素 黒髪	赤髪	加算する要素 金髪	...
操作前	1	1	0	...
操作後	0	1	1	...

1から0へ 0から1へ

図 2: ベクトルの操作

3.2 萌え要素辞書作成

収集した 1182 語の萌え要素のうち、重複する表現、公序良俗に反する表現、表現単体で特徴を表していないもの (e.g., うなじ, 髪の色), キャラクタの内面, 見た目の特徴を表していない表現 (e.g., 温泉回, 世界観) を取り除いた 854 語を萌え要素辞書に使用する表現とした。漢字・ひらがな・カタカナの差や、表記揺れ、別名称が存在する表現については、各表現ごとに weblio 辞書⁵と 3.1 節で使用した自由参加型 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 (表 1 参照)。統一前の表現が抽出された際、統一後の表現が抽出されたものとして処理する。例えば、表 1 に示した「タレ目」の場合「タレ目」、「たれ目」、「垂目」、「垂れ目」のいずれの表現が抽出された場合においても、「タレ目」が抽出されたとする。

また、萌え要素として収集した表現の中には「純情」、「純真」、「純粹」のように、意味の類似した語が存在する。これらの語については、weblio 辞書で類似語検索を行い、類語として分類されている語を、同一の意味カテゴリの要素として統一した。辞書作成に用いた単語・短文が weblio 辞書の類語検索辞書に存在しなかった場合、語の意味を weblio 辞書と自由参加型 Web 百科事典で検索し、説明文の中に類語や別表記が存在しないかを確認した (表 2 参照)。

これらの処理により、萌え要素の辞書として 465 の意味カテゴリを持つ「萌え要素辞書」を作成した。

3.3 特徴ベクトル作成

「漫画全巻ドットコム⁶」の漫画発行部数ランキングに掲載された 102 作品に登場する 508 キャラクタを対象とし、キャラクタに関する説明が含まれた文書を萌え要素辞書作成と同一の Web サイトから収集した。まず、キャラクタごとに収集した文書から、萌え要素辞書に含まれる表現を抽出する。この際、語尾の表記揺

れに関しては MeCab による形態素解析を行い原形に統一した。

次に、各キャラクタごとに抽出された表現を基底として one-hot ベクトルを作成した。このとき、抽出対象となる辞書として使用した 465 次元から、抽出数が 0 の基底語を除去することで、362 次元に圧縮した。本稿ではこれをキャラクタの特徴ベクトルと呼ぶ。

4 提案手法

提案手法では、前章で作成したキャラクタの特徴ベクトルを用いて、ユーザが指定したキャラクタ (以下、基準キャラクタと記す) と類似するキャラクタを提示する。ユーザは検索条件を入力する場合、「ルフィ + 金髪」、「悟空 - 熱血」のように、基準キャラクタに対して加減算を行う萌え要素 (以下、変更要素と記す) を設定する。基準キャラクタの特徴ベクトルに対して、加算する変更要素の次元を 1 に、減算する変更要素の次元を 0 に操作する。例えば、特徴ベクトルの基底語として「黒髪」を持ち、「金髪」を持たない基準キャラクタに対し、「金髪」を加算する変更要素として、「黒髪」を減算する変更要素に指定した場合、図 2 のように操作する。その後、操作された特徴ベクトルと類似度が高い特徴ベクトルを持つキャラクタを判定し、ユーザに対して提示する。

4.1 関連要素の設定

ユーザが変更要素を指定する際、内面特徴については表現の多様性が高いため (e.g., 「根暗」と「陰気」)、ユーザが指定した変更要素がキャラクタの基底語ではない場合が想定される。しかし、ユーザが表現の多様性を考慮して変更要素の指定を行うのは困難である。そこで、変更要素と関連の強い基底語 (以下、関連要素と記す) についても同様の操作を行う。例えば、「無愛想」の関連要素として「無口」と「冷淡」が設定されていた場合、「無愛想」が加算する変更要素として指定された場合、「無愛想」、「無口」、「冷淡」の次元を 1 に変更する。関連要素を設定するにあたり、内面特徴を示す

⁵weblio 辞書 (<https://www.weblio.jp/>: 2018 年 11 月から 2018 年 12 月確認)

⁶<https://www.mangazentan.com/> (2018 年 12 月 20 日時点確認)

表 2: 辞書検索の例 (weblib 辞書)

収集した萌え要素	意味	類似語
冷血漢 脳筋 おちゃっぴい	人情味にかけている人。冷酷な人間 思考が単純・バカ, といった性格の者を評して言う語 おしゃべりで, 茶目っ気のあるさま	冷酷 単純, バカ おしゃべり, 茶目っ気

表 3: パターンの例

基準キャラクタ	プラスする要素	マイナスする要素	正解キャラクタ
サンジ うずまきナルト 三千院ナギ 結城リト 高良みゆき	真面目・無愛想・常識人 短気・傲慢 冷淡 気弱・勇気・面倒見が良い あざとい・ツインテール	女好き 孤独・ツツンヘア 自己中心的・お調子者 鈍感 賢い・眼鏡	佐藤潤 エドワード・エルリック ソーニャ 沢田綱吉 吉川ちなつ

萌え要素間での関連性の強さを分析する。一般に、文書中で共起関係がある語は同じ文脈を共有していると考えられている。よって、共起ネットワーク上でエッジが結ばれたノード同士には意味的な関連性が認められる場合がある [7]。そこで本稿では、各キャラクタを説明する文書においての内面特徴の共起関係に着目し、共起ネットワークを作成することで関連の強い基底語の分析を行った。共起ネットワークとは、文書に出現する語をノードとし、共起関係があるノード同士をエッジで結びつけることにより、ノード同士の関係性を可視化するネットワークである。共起ネットワークの作成には、樋口が作成・公開しているテキストマイニングツール、KH Coder を使用した [8]。KH Coder はテキスト型のデータを統計的に分析するためのフリーソフトウェアである。共起関係は、Jaccard 係数によって算出した。

共起ネットワークの作成に使用するデータは、特徴ベクトルを作成する際に収集した 508 キャラクタの文書と、特徴ベクトルの 362 次元のうち、内面特徴を表す 153 の要素を対象とした。KH Coder により文書を形態素解析し、辞書に含まれる表現のうち内面特徴を示す表現の共起回数を抽出する。共起関係の集計単位はキャラクタごととした。不必要に多くのノードと結びつき、関連要素の数が膨大になることで、萌え要素の加減算の結果がユーザの意図と離れることを防ぐため、本稿では最小限に簡潔化されたネットワークの作成を目指した。そのため、KH Coder による共起ネットワークの作成条件は出現数 5 以上、Jaccard 係数 0.1 以上の最小スパニング・ツリーとした。KH Coder の機能であるサブグラフ検出により、共起ネットワーク上で相対的に強く結びついているグループが自動的に色分けされており、ネットワーク全体から見ると同じ色で分けられた部分は強い共起関係がある。本稿では、作成した共起ネットワーク図上で隣接するノード同士を関連要素とした (図 3 参照)。例えば「内向的」の場

合は「穏やか」と「恥ずかしがり」が、「残忍」の場合は「戦闘狂」と「意地悪」が関連要素となる。

4.2 類似度の判定

ユーザにとっては、変更要素として指定した要素が、検索において最も重視する基底語である。そのため基準キャラクタと比較キャラクタの類似度は、特徴ベクトル中の変更要素の次元の一致率で算出した。この値が同じ比較キャラクタ同士の順位づけは、変更要素の次元以外の一一致率で決定した。一致率の算出には \cos 類似度を用いた。 \cos 類似度は、二つの n 次元のベクトル間の距離を測る際に用いられ、値が 1 に近いほど二つのデータが似通っていることを表す。

5 実験

5.1 実験目的

本稿の提案手法では、自由参加型 Web 百科事典から収集した萌え要素、キャラクタの説明に関する文書から抽出した情報に基づいて萌え要素の加減算を行なっている。人が判断する萌え要素の加減算の結果と、提案手法が提示する萌え要素の加減算の結果に一致性が見られるかを確認することで、提案手法の有効性を検証した。

5.2 正解データの作成

人が判断する萌え要素の加減算結果をアンケートで収集し、正解データとする。アンケートは 10 代から 20 代の男女 6 名 (男性 4 名, 女性 2 名) を対象として行った。アンケートでは「見た目と内面の特徴が近くなるよう、“基準キャラクタ + 眼鏡 - 恥ずかしがり屋 = 解

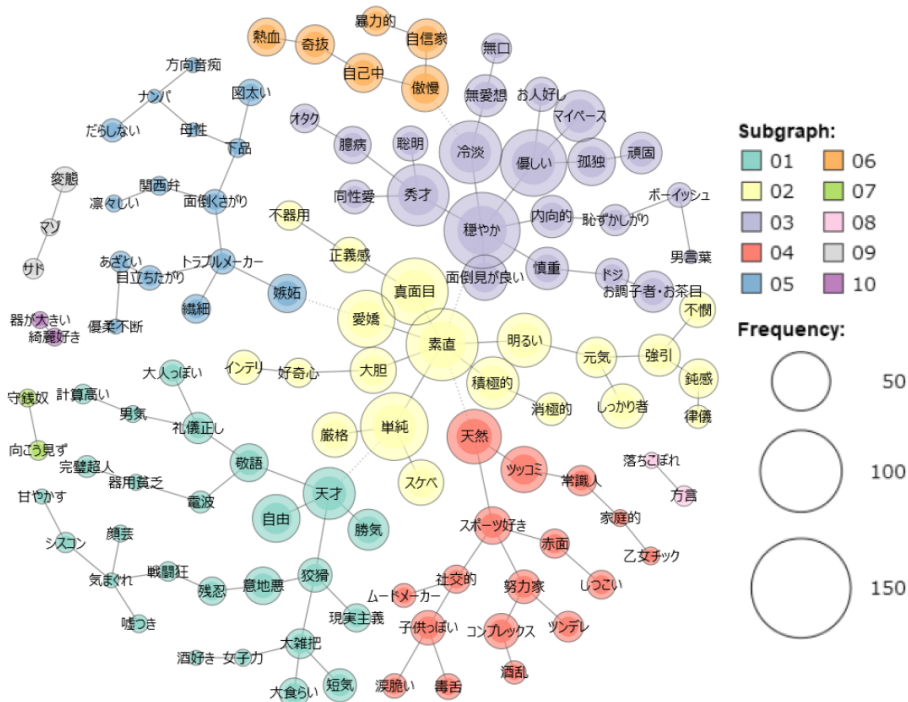


図 3: 共起ネットワーク

答キャラクター”という形式に当てはまる計算式を作ってください」という指示を与えた。このアンケートにより基準キャラクター、解答キャラクター、変更要素の組が20パターン収集された。得られたパターンの一例を表3に示す。

5.3 実験概要

実験は、3.3節で収集した508キャラクターに、アンケートで挙げられたキャラクターのうちでそれに含まれない13キャラクターを加えた521キャラクターを対象に、特徴ベクトルを作成した。アンケートによって収集された20の基準キャラクターと変更要素をもとに、4章で述べた手法を用いて基準キャラクターの特徴ベクトルを操作した。特徴ベクトルを操作した基準キャラクターと、特徴ベクトルを作成した521キャラクターの類似度判定を行う。類似度判定の結果、20の各パターンごとに520のキャラクターの中で上位3位までに解答キャラクターが提示されれば正解とした。

5.4 実験結果

提案手法と正解データを比較した結果、解答キャラクターが1位として提示されたのは45.0% (20キャラクター中9キャラクター)であった。2位として提示されたのは

20.0% (20キャラクター中4キャラクター)であった。3位として提示されたのは5.0% (20キャラクター中1キャラクター)であった。上位3位までに提示されたのは70.0% (20キャラクター中14キャラクター)であった。提案手法による結果を表4に示す。さらに、変更要素を加減算する前の基準キャラクターと解答キャラクターの類似度の順位平均と、変更要素に操作を加えた基準キャラクターと解答キャラクターの順位平均を比較した。変更要素の加減算する前の解答キャラクターの類似度の順位平均は126.3位 (最高4位, 最低432位, 分散10268.21)であり、変更要素の加減算を適用した後の順位平均は9.6位 (最高1位, 最低46位, 分散216.54)であった。

6 考察

実験により人手で作成された萌え要素の加減算結果を正解データとし、本手法の有効性を検証した結果、精度は70.0%であった。また、変更要素操作前から操作後で順位の分散は10268.21から216.54へと収束し、類似度順位が下がったのは20パターン中1パターンのみだった。これらの結果から、本稿で提案した萌え要素の加減算に基づくキャラクター類似度判定手法に有効性があることが示唆された。

一方で、解答キャラクターを上位3位以内に提示できなかった6パターンのうち5パターンでも解答キャラクターの類似度順位の上昇幅は限定的であった。これは、

表 4: 解答キャラクタ順位結果

パターン番号	変更要素操作前	変更要素操作後
1	130 位	1 位
2	149 位	38 位
3	10 位	1 位
4	165 位	2 位
5	79 位	1 位
6	4 位	43 位
7	54 位	1 位
8	322 位	2 位
9	180 位	46 位
10	41 位	1 位
11	59 位	3 位
12	72 位	21 位
13	113 位	2 位
14	33 位	2 位
15	179 位	11 位
16	82 位	1 位
17	171 位	1 位
18	160 位	13 位
19	91 位	1 位
20	432 位	1 位

アンケート回答者が想定した基準キャラクタ、解答キャラクタが持つ萌え要素と、提案手法が判断した基準キャラクタ、解答キャラクタの基底語に差異があったことで、変更要素の次元の一致率が下がったことが要因となっている。アンケート回答者と提案手法の判断に差異が生まれた原因として、キャラクタが所持する萌え要素を表現した特徴ベクトルを作成する際の、萌え要素の抽出手法の精度が影響していると考えられる。

萌え要素を抽出する際、提案手法では萌え要素辞書の表現と一致した表現のみを抽出している。そのため文脈によって表現された萌え要素が存在すると、不要な要素を抽出してしまう場合や、抽出すべき要素が抽出できない場合がある。そのため、萌え要素を抽出する際、萌え要素の抽出の際に行う形態素解析の後に構文情報を参考とした係り受け解析を行うことを検討する。

また、提案手法では特徴ベクトルを作成する過程において one-hot 表現を用いているため、特徴ベクトルの基底語間に重要度の差が存在しない。抽出された萌え要素のうちどの要素がユーザにとって重視されているのかを考慮した類似度判定を行うため、文書からキャラクタの萌え要素を抽出する際、Bag-of-words や TF-IDF を用いてキャラクタの各基底語に対して重み付けを行うことで、類似度判定において変更要素に指定された要素の一致率を算出する際、加算する変更要素として指定された要素の重みが重いものを上位に提示することで、よりユーザが重視した観点に基づいた類似判定が可能であるかについて検討したい。

7 おわりに

本研究では、既知キャラクタをもとに類似したキャラクタを探す検索方式に着目し、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を提案した。人手で作成したパターンを正解データとした実験から、提案手法の有効性が示唆された。今後は、文脈を考慮した萌え要素の抽出や、ユーザが重要視するキャラクタの基底語に対して重み付けを行い、萌え要素の加減算結果の精度向上を目指す。

参考文献

- [1] 東 浩紀: 動物化するポストモダン—オタクから見た日本社会—, 講談社 (2001).
- [2] 土方 嘉徳: 推薦システムにおけるインタラクション研究へのいざない, ヒューマンインターフェース学会誌, Vol. 15, No. 2, pp. 131–134 (2013)
- [3] R. S. Taylor: Question-negotiation an information-seeking in libraries, *College and Research Libraries*, (1968).
- [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, 第 32 回人工知能学会全国大会論文集, 1J3-02 (2018).
- [5] 河谷 大和, 柏崎 礼生, 高井 昌彰, 高井 那美: アニメにおける人物顔画像の萌え因子特徴評価と検索分類システムへの応用, 映像情報メディア学会技術報告, Vol. 34, pp. 113–118 (2010).
- [6] 白石 諒, 大竹 恒平, 生田目 崇: レビューに含まれる感情を用いた児童向け絵本分類の検討, 情報処理学会第 79 回全国大会 (分冊 2), pp. 533–534 (2017).
- [7] 佐藤 進也, 福田 健介, 菅原 俊治, 栗原 聡: 文書ストリームにおける語のバーストと共起ネットワークにおけるクラスタ構造の関係について, 情報処理学会論文誌データベース (TOD), Vol. 48, No. SIG14(TOD35), pp. 69–81 (2007).
- [8] 樋口 耕一: 言語研究の分野における KH Coder 活用の可能性, 計量国語学会, Vol. 31, No. 1, pp. 36–45 (2017).