

座標情報に基づく漫画のオブジェクト順位推定

元山 直輝^{1,a)} 岡田 真¹ 森 直樹¹

概要：本研究では、漫画における台詞、コマの読み順を座標情報を基に機械学習によって推定した。漫画見開き 2 ページに対する台詞、コマの通し順位推定の実験をした結果、それぞれ従来手法を上回る高い精度を得られた。さらに、漫画内の文書の言語的特徴について調査し、自然言語処理の活用の可能性について考察する。

キーワード：コミック工学, 機械学習, Random Forest, コマの順序推定, 台詞の順序推定, アノテーション

1. はじめに

近年、多様な分野において人工知能や機械学習を用いる試みが盛んに研究されており、人の創作物理解にも機械学習が適用され、大きな発展が見られている。特にコミック工学に代表される漫画を対象とした研究は、画像処理と言語処理が密接に結びついたマルチモーダルな分野として注目されている。漫画を対象とした研究では、画像処理に基づいた研究は数多くなされてきたが、対話理解にまで踏み込んだ研究はほとんどなされてこなかった。その一因はデータにある。漫画は画像データで提供されることが多く、漫画に含まれるテキストの情報を簡単に扱うことが難しかったためである。

コマや台詞の順序の推定は、読者が物語の文脈を踏まえて推定する必要があり、漫画のコマや台詞といったオブジェクトの順番に関するアノテーションデータが必要不可欠である。そこで、コマと台詞に人手で順位をつけるアノテーションツールを開発し、既存の漫画データセットを拡張した。そして、文脈を踏まえた漫画の理解の準備として、座標情報を用いた機械学習によるコマと台詞の順位推定手法を提案する。また、実験により提案手法の有効性と改善点について考察する。さらに、台詞の順位推定の実験で誤りの見られた例から、自然言語処理の活用の可能性について考察する。

2. Manga109

一般に公開されている漫画のデータセットとして

¹ 大阪府立大学
Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai,
Osaka 599-8531

^{a)} motoyama@ss.cs.osakafu-u.ac.jp

「Manga109」 [1], [2] がある。漫画の研究のために相澤らにより作られたもので、漫画 109 冊の画像データに加え、登場人物の名前や、台詞文、画像内における登場人物の顔、全身、コマ、台詞の座標のアノテーションデータが含まれる。Manga109 ではすべての漫画の画像データは見開き 2 ページ 1 枚の形式で保存されており、画像上での台詞やコマを含む矩形の座標情報がアノテーションによって付与されている。画像上の座標平面は左上を原点とし、右向きに x 軸、下向きに y 軸が設定されている。座標情報は台詞、コマを囲む矩形の x 座標の最小値、y 座標の最小値、x 座標の最大値、y 座標の最大値の 4 次元で表される。

本研究では、Manga109 を対象としたアノテーションツールを開発し、見開き 2 ページ内の台詞およびコマ全てに対して一意の通し順位と、それぞれの台詞、コマが 4 コマ漫画のものであるか否かのメタデータを付与した。

3. 先行研究

座標に基づく台詞順序推定の先行研究として、佐藤 [3] により提案された手法がある。この手法では、同一見開きページ内の任意の台詞 2 つを 1 組にした台詞組集合を作成し、素性ベクトルでそれぞれの台詞組を表現する。各ベクトルは座標 4 次元 \times 2 個とそれらの差分 4 次元の計 12 次元と、台詞の所属するコマについても同様の 12 次元による計 24 次元ベクトルとする。正解ラベルとして正順ならば 0 を、逆順ならば 1 を付与する。ここで正順とはオブジェクトが読む順番に並んでいる場合を指し、逆順はその逆を指す。これを見開きページ内の全台詞に対して適用して各々の台詞 2 つの順序識別し、ランキングを用いることで見開き 2 ページについて台詞の通し順位を推定する。なお、台詞が属するコマの条件は以下の通りである。

- (1) 台詞を囲う矩形の重心が内部に存在するコマ
 - (2) 1. を満たすコマが無い場合、矩形の四隅が最も多く属するコマ
 - (3) 2. を満たすコマが複数存在する場合、そのうちで台詞の矩形の重心とコマの重心の距離が最も小さいコマ
- また、アルゴリズムに基づくコマ順序推定の先行研究として、武井ら [4] の手法がある。この手法ではまず、対象ページ内の全てのコマの中で、コマ領域の左上の座標が一番上に位置しているコマを検出する。もし検出されたコマが 1 つであった場合は、そのコマの読み順序を「1」と確定する。検出されたコマが複数存在する、つまり複数のコマが一番上に位置している場合は、そのコマの中で一番右に位置しているコマの読み順序を「1」と確定する。読み順序確定後、確定したコマを除いたページ内の全てのコマを対象として、再度同じ処理を繰り返すことでページ内全てのコマの通し順位を確定する。

4. 提案手法

本研究の提案手法は以下の 3 通りに分かれる。

- (1) 4 コマ漫画を含んだ「台詞」通し順位推定
- (2) コマの座標情報による「コマ」通し順位推定
- (3) コマ通し順位推定を利用した「台詞」通し順位推定の補正

手法 1 は従来手法の問題点の改善、手法 2, は従来手法のコマへの適用、手法 3 は手法 1, 2 を踏まえた台詞の通し順位推定の改良をそれぞれ目的としている。

4.1 4 コマ漫画を含んだ「台詞」通し順位推定 (手法 1)

手法 1 は、従来手法の問題点の改善を目的とする。先行研究で示した座標に基づく台詞順序推定に対して、4 コマ漫画における台詞を含めた順序識別およびコマの所属判定の改良をした。そのコマが 4 コマ漫画のものであれば 1 を、そうでなければ 0 を表す 1 次元ベクトル \times 2 個を加えた計 26 次元素性ベクトルで台詞組を表現する。これを見開きページ内の全台詞に対して適用して各々の台詞 2 つの順序識別し、ランキングを用いることで見開き 2 ページについて台詞の通し順位を推定する。また、台詞が属するコマの条件は

- (1) 台詞の矩形の四隅が全て所属するコマ
- (2) 1. を満たすコマが複数存在する場合、コマのそれぞれの四隅と台詞の矩形の重心との距離の平均が最も小さいコマ
- (3) 1. を満たすコマが無い場合、矩形の四隅が最も多く属するコマ
- (4) 3. を満たすコマが複数存在する場合、そのうちで台詞の矩形の重心とコマの重心の距離が最も小さいコマとした。

4.2 コマの座標情報による「コマ」通し順位推定 (手法 2)

手法 2 では、手法 1 をコマの通し順位推定に適用する。手法 1 と同様の素性ベクトルでコマ組を表現する。各ベクトルは、コマについての座標 4 次元 \times 2 個およびそれらの差分 4 次元の計 12 次元に、4 コマ漫画の台詞であるかどうかを示す 1 次元ベクトル \times 2 個を加えた計 14 次元とする。通し順位推定手法については手法 1 と同様である。

4.3 コマ通し順位推定を利用した「台詞」通し順位推定の補正 (手法 3)

手法 3 では、手法 1, 2 を踏まえた台詞の通し順位推定の改良を目的とする。手法 1 によりコマ毎に台詞の通し順位推定し、手法 2 によるコマ通し順位推定を利用して台詞通し順位推定の補正をする。以下にその手順を示す。

- (1) 見開き 2 ページでコマの通し順位推定をする。
- (2) 見開き 2 ページ内の全台詞の所属コマを判定をする。
- (3) コマ毎にそのコマに所属する台詞のみで通し順位推定をする。
- (4) 1, 3 で得た結果を元に、各コマにおける台詞順位を見開き 2 ページ通し順位に並び替える。

5. 実験

本実験では、提案手法の有効性を確認するため、提案手法と従来手法のそれぞれについて通し順位推定実験をし、作品毎に 1 つ抜き交差検証をすることで精度を確かめた。従来手法として、台詞の通し順位推定では佐藤の手法を、コマの通し順位推定では武井の手法を用いた。実験で用いたデータは Manga109 より漫画 7 作品を引用した。表 1 に用いた作品を示す。なお、作品 No. 3, 4, 5, 6 は 4 コマ漫画作品である。また、各オブジェクト組の順序推定には Random Forest を用いた。表 2 にその条件を示す。

5.1 Random Forest

Random Forest[5], [6] は、2001 年に Leo Breiman によって提案された機械学習手法である。複数の決定木を生成し、個々の決定木を弱学習器としてアンサンブル学習することで、高い識別性能を持った識別器を構築する。単純な構造で良好な結果が得られるだけでなく、学習が高速であるという特徴を持つことから、パターン識別をはじめとして回帰やクラスタリングに利用される。

5.2 手法 1

手法 1 について実験した。実験には表 1 の 7 作品の台詞計 10936 個を用いた。表 3 に通し順位の識別結果および従来手法の結果を示す。表 3 より、従来手法に対して、4 コマ漫画の作品の識別率が大きく向上したことが確認できた。また、図 1, 2 に、手法 1 により斜めコマにおける台詞の所

表 1 使用した作品

作品 No.	作品名	台詞数	コマ数
0	AisazuNihaiIrenai	1474	1014
1	Arisa	1391	910
2	LoveHina vol01	1810	1077
3	KoukouNoHitotachi	2423	938
4	OL Lunch	1808	831
5	TetsuSan	1179	715
6	YouchienBoueigumi	851	363
合計		10936	5853

表 2 実験条件: RF

決定木の深さ	256
決定木の数	16
説明変数の最大使用数	8
データ分割最小数	2

表 3 台詞の識別率

作品 No.	従来手法	提案手法
0	0.953	0.984
1	0.928	0.941
2	0.861	0.927
3	0.974	0.976
4	0.543	0.765
5	0.927	0.998
6	0.945	0.951
全体の精度	0.868	0.930

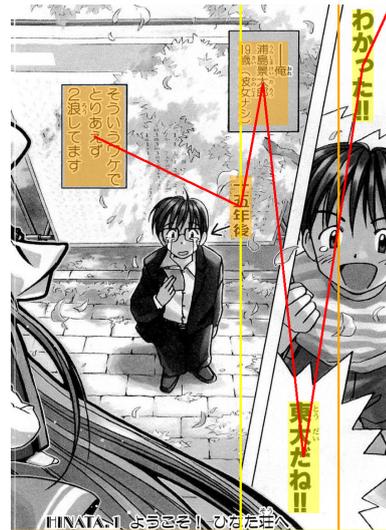


図 2 手法 1 による台詞の所属コマ判定 ([1], [2] より引用)



図 1 従来手法による台詞の所属コマ判定 ([1], [2] より引用)

属判定が適切になされた例を示す。図 1 は従来手法による台詞の通し順位推定の結果をプロットしたものであり、図 2 は手法 1 による台詞の通し順位推定の結果をプロットしたものである。それぞれの図において赤線は推定された通し順位を表しており、台詞領域の色と台詞の所属するコマの色は対応している。図 1, 2 より、台詞の所属コマが正しく判定されていることが分かる。台詞の所属コマ判定されたことから、順位が正しく推定されたと考えられる。



図 3 作品 No.4 の通し順位推定結果 ([1], [2] より引用)

次に、最も通し順位識別率が低かった作品 No.4 について考察する。図 3 に通し順位推定の結果の例を示す。それぞれの図において赤線は推定された通し順位を表しており、台詞領域の色と台詞の所属するコマの色は対応している。

この作品は 4 コマ漫画であり、見出しとなるイラストと台詞が各話の最初に設けられているが、この箇所にはコマの判定がないため台詞の所属判定が適切になされていなかった。また、各コマのタイトル部分はコマの判定がないため各話の 1 コマ目の台詞として扱われており、このコマの内部において漫画の原則 (右上, 右下, 左上, 左下) に従って読み進められているため、タイトル部分の誤判定が数多く見られた。これらのことから、他作品と比べて識別率を大きく下げることになったと考えられる。また、漫画の原則に則っていない読み方をするような台詞には誤識別が多く見られた。このような台詞では座標だけでその台詞順を判定することが困難であり、台詞文を用いた補正が必

表 4 コマの識別率

作品 No.	従来手法	提案手法
0	0.897	0.986
1	0.826	0.974
2	0.812	0.943
3	0.311	1.0
4	0.383	0.997
5	0.233	1.0
6	0.298	1.0
全体の精度	0.585	0.982

要であると考えられる。

5.3 手法 2

手法 2 について実験した。実験には表 1 の 7 作品のコマ計 5853 個を用いて通し順位推定をした。表 4 に通し順位の識別結果および従来手法の結果を示す。従来手法に対して、高い精度でコマの順位を推定できることが確認できた。ここで、識別率が高かった作品について考察する。識別率が 0.99 を上回った作品 No.3, 4, 5, 6 の作品は全て 4 コマ漫画であった。4 コマ漫画のようなコマ構造が単純な作品については、高い精度で推定できると考えられる。

次に、誤識別したコマについての例を考える。図 4, 5 に推定に誤りが見られたコマの例を示す。図 4 は他のコマに内包される関係を持つ被内包コマの例であり、図 5 は縦方向に隣接しており横方向に離れた位置関係を持つコマの例である。図 4 のような被内包コマは、今回用いたコマ 5853 個のうち 52 個存在した。また、誤識別したコマ 102 個のうち、被内包コマは 28 個存在した。これより、被内包コマの 53% を誤識別し、また全体の誤りの 27% が被内包コマであることが分かる。このことから、被内包コマの順位推定は困難であると考えられる。また、図 5 のような位置関係のコマも誤識別したコマの中に 10 個存在したこれらのコマに共通する特徴として、コマの面積が小さいことが挙げられる。そこで、コマの面積について考察する。図 6 に全コマにおける面積の分布を、図 7 に誤識別したコマにおける面積の分布をそれぞれ示す。横軸は面積の平方根を表しており、縦軸はコマの数を表している。図 7 より、面積 250 以下のところに誤識別したコマが集中していることから、小さいコマは誤識別しやすいと考えられる。このことから、学習データの不足が誤識別の一因として考えられるので、データをさらに拡張したのちに更なる原因の考察が必要である。

5.4 手法 3

手法 3 について実験した。表 5 に手法 3 による識別結果と、比較として手法 1 による識別結果を示す。全ての作品において精度が向上したことから、コマによる補正の効果が確認できた。図 8, 9 に補正が成功した例を示す。図 8 は

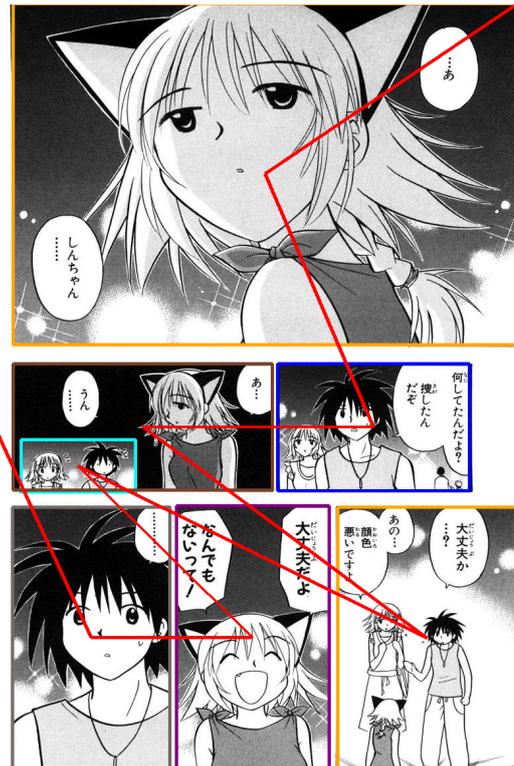


図 4 被内包コマの例 ([1], [2] より引用)



図 5 横方向に離れた位置関係のコマの例 ([1], [2] より引用)

手法 1 による通し順位推定の結果をプロットしたものであり、図 9 は手法 3 による通し順位推定の結果をプロットしたものである。これらより、コマ順を正しく順位推定できた場合、台詞の順序推定に対する補正の効果が期待できると考えられる。

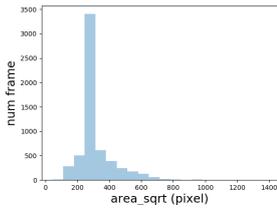


図 6 全てのコマの面積

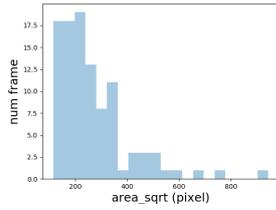


図 7 誤識別したコマの面積

表 5 コマで補正した台詞の識別率

作品 No.	手法 1	手法 3
0	0.984	0.993
1	0.941	0.950
2	0.927	0.938
3	0.976	0.977
4	0.765	0.770
5	0.998	0.998
6	0.951	0.958
全体の精度	0.930	0.934

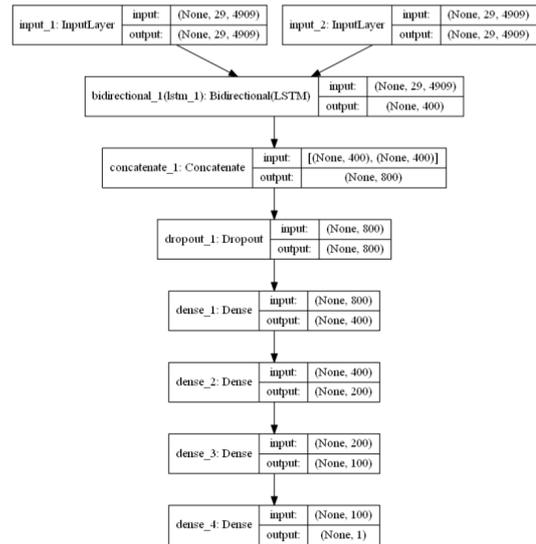


図 10 台詞の文を用いる順序識別ニューラルネットワークの構造 ([3] より引用)

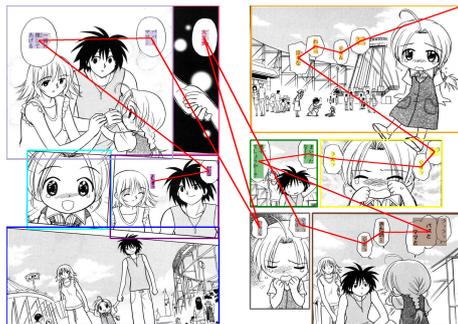


図 8 手法 1: 作品 No.2 ([1], [2] より引用)

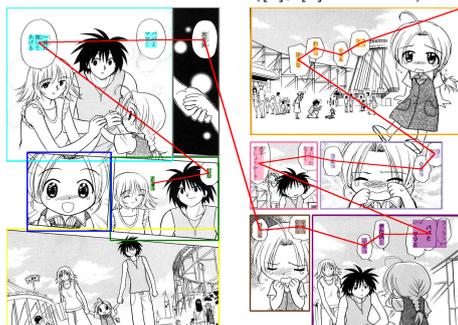


図 9 手法 3: 作品 No.2 ([1], [2] より引用)

6. 自然言語処理の活用に関する考察

本稿では、座標情報に基づいた漫画のオブジェクト順位推定およびコマの通し順位推定を利用した台詞の通し順位推定の補正をした。ここでは、手法 1 による台詞通し順位推定で誤りのみられた台詞の例を通して、自然言語処理の活用による台詞の通し順位推定の補正に関する考察を述べる。

6.1 先行研究

先行研究である佐藤 [3] の手法では、まず Manga109 でアノテートされているすべての台詞文について、形態素解析で見出し語に分割する。次に、見出し語から one-hot ベクトルを作成し、順に並べることで各台詞をベクトルの系列に変換する。そして同一見開きページ内で連続する 2 台詞を 1 組とし、RNN モデルで順序識別をする。図 10 にこのモデルの概略図を示す。

この手法における台詞組の順序推定の精度は 0.82 に留まっており、台詞の座標情報に基づいた推定結果よりも低い結果となっていた。このことから文情報のみによる順序識別は容易ではないことが分かる。この理由として、漫画特有の台詞の特徴が考えられる。漫画の台詞は小説と比較して口語表現が多様され、それぞれの文章は短文であることが多い。したがって、漫画の台詞の特徴を反映した解析が必要であると言える。

6.2 言語処理を用いた補正による効果が期待できる台詞例

次に、本実験で用いた漫画における台詞の順位推定で誤識別したものの中から、言語処理を活用により順位の補正が期待できる例について考察する。

Allen ら [8] によると漫画の読者は文字情報の種類により読む順番あるいは読み飛ばす文を決定していると考えられている。例えば、登場キャラクターによる会話文が優先的に読まれ、それ以外の文が読み飛ばされているという実験結果が報告されている。よって、漫画における台詞を特徴によって分類することは台詞の順序推定において重要であると考えられる。そこで、手法 1 による推定結果をプロットした図から、それぞれの台詞の特徴について述べる。

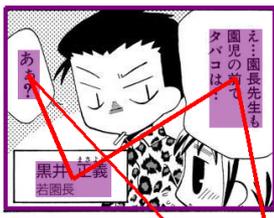


図 11 台詞文と地の文
 ([1], [2] より引用)

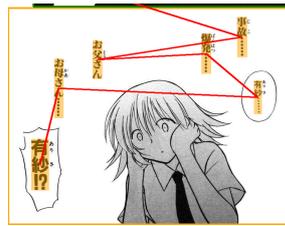


図 12 独白
 ([1], [2] より引用)

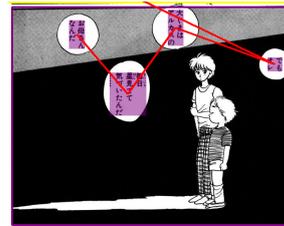


図 13 連続台詞文
 ([1], [2] より引用)

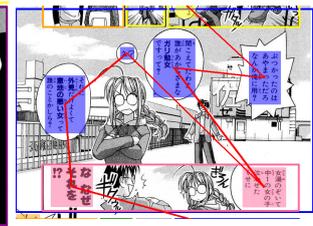


図 14 会話
 ([1], [2] より引用)

6.2.1 地の文

漫画の文情報は、台詞文 (キャラクターによる発言) と、地の文 (特定の発話者が存在しない文) に大別できる。ここまでにおいて、Manga109 でアノテーションされた文情報は全て台詞として扱ってきたが、以降は文情報と台詞は区別する。同一コマ内に台詞と地の文が混在していた場合、文情報の種類によりグループ分けして読み進めると考えられる。そこで、図 11 に文情報から台詞と地の文を区別することにより読み順推定の補正が見込める例を示す。この例において、台詞文と地の文が交互に読まれると推定されている。文情報から台詞文と地の文の判定することで、読み順の補正ができると考えられる。

6.2.2 独白文

漫画の文情報は、登場人物の心情を表現した独白文と、実際の発話文に分けることができる。このような例においても、それぞれの文情報のグループにより読み順が決まると考えられる。そこで、図 12 に独白文と発話文を区別することにより読み順推定の補正が見込める例を示す。この例では、独白文と発話文が交互に読まれると推定されている箇所が存在する。文情報から独白文と発話文の判定することで、読み順の補正ができると考えられる。

6.2.3 一人の話者による連続台詞文

漫画では、一人の話者の発言が複数の吹き出しに渡って連続するような表現が見られる。そこで、図 13 に台詞文の連続性の推定が可能な例を示す。この例において、「大ぐまはアルカスの」「お母さんなんだ」と、「でもオレ」「毎日星見てて気づいたんだ...」という 2 つの発言が見られる。このような発言の連続性を推定することで、文情報の繋がりを取得できると考えられる。

6.2.4 会話

図 14 に登場人物による会話の例を示す。この例では、誤識別により会話が途切れていることが分かる。そこで、会話の連続性について発話者の観点から考察する。漫画において、登場人物はキャラクター固有の語尾、言葉遣いで発話している作品も少なくない。よってこの例のように同一コマで複数の人物が複数の発言をしている場合においても文情報からの発話者の推定が可能であると考えられる。それぞれの台詞の話者を踏まえた上で、台詞同士の応答関係を明らかにすることで会話順が推定できると考えられる。

7. まとめと課題

本研究において、4 コマ漫画を含めた漫画オブジェクトの通し順位推定、およびコマを利用した台詞の通し順位推定の補正に関する手法を提案した。従来研究に対して台詞の所属するコマの判定を改良し、4 コマ漫画を表す特徴量を組成ベクトルを追加することで、高い精度で通し順位を推定できることが確認できた。また、コマを利用した台詞順位推定の補正についても精度の向上が確認できた。さらに、自然言語処理の活用に関する考察から、言語情報を用いた解析が台詞の順位推定の補正に効果が期待できると考えられる。今後の課題として以下の点が挙げられる。

- 台詞の所属コマ判定の改善
- 言語情報による台詞順序の補正
- アノテーションデータの拡張

参考文献

- [1] Y.Matsui, K.Ito, Y.Aramaki, A.Fujimoto, T.Ogawa, T.Yamasaki, K.Aizawa, Sketch-based Manga Retrieval using Manga109 Dataset, Multimedia Tools and Applications, Springer, 2017
- [2] T.Ogawa, A.Otsubo, R.Narita, Y.Matsui, T.Yamasaki, K.Aizawa, Object Detection for Comics using Manga109 Annotations, arXiv:1803.08670, 2018
- [3] 佐藤幸一. 自然言語技術を用いた人の創作物における対話データの解析および応用. MS thesis. 大阪府立大学. 2019
- [4] 武井響也, 三原哲也, 永森光晴, 杉本重雄. マイクロタスクによるマンガの暗黙構造についてのメタデータ作成 - マンガのコマを読む順序とテキストの話者について - 第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM2019) 予稿集. 2-146. 2019.
- [5] Leo Breiman. Random Forests. Machine Learning, Vol.45, No.1, pp.5-32, 1995.
- [6] 波部齊. ランダムフォレスト. 情報処理学会研究報告, vol.2012-CVIM-182, No.31, 2012.
- [7] UNSER-SCHUTZ, Giancarla. マンガにおける言葉の役割 - 構造・語彙・登場人物という三つの観点から -. Ph.D thesis, 一橋大学, 2013
- [8] Allen.K, Ingulsrud.J. Strategies used by children when reading manga. 神田外語大学紀要. 2018