

漫画のセリフと発話者対応付けデータセットの構築と その分析

櫻井 翼^{†,a} 伊藤 理紗[†] 阿部 和樹[†] 中村 聡史[†]

[†] 明治大学 総合数理学部

a) ev190522@meiji.ac.jp

概要 漫画の発話者推定などの研究を進展させるには、そのセリフを誰が発話したのかというデータセットが必要となる。本研究では、これまでに構築してきた Manga109 における漫画のセリフと発話者を対応付けるデータセットを、アノテーション付与者数が平均 2 人から 5 人へとなるように拡張を行った。また、これらのデータセットにおける各セリフの評価人数や分散度合いなどを指標とした分析を行った。分析の結果、アノテーション付与者が 2 人と 5 人の時とでは 10% 近く完全一致率が下がることを明らかにした。また、評価一致度指標を算出することで、SF やバトルなどのシーンでは、アノテーション付与者にとって発話者の対応付けが困難であることなどを明らかにした。

キーワード コミック, セリフ, 話者, データセット構築

1 はじめに

電子書籍ビジネス調査報告書 2020[1]によると、電子書籍の市場規模は年々増加しており、その中でもコミックの市場占有率は 80% を超えている。また、2017 年には電子コミックの売上が紙媒体のコミックの売上を上回る結果となっている。2021 年現在、コロナ禍なども影響し電子コミックの市場規模がより拡大しており、電子コミックとして漫画を楽しむことがより一般的となりつつある。このような電子コミックの普及に伴い、今後デジタルならではの活用例や楽しみ方が増えていく。

デジタル端末を利用した電子コミックであることを活かした処理やシステムに関する研究も増加している。Mantra[2]ではコミックの画像情報やテキスト情報から文脈に則した自動翻訳を行っている。その他にも、漫画の検索や推薦・ネタバレ防止といった、コミックの内容を考慮した研究がなされている[3][4]。このような研究のためには、そもそもコミックを構成する様々な要素(コマの領域、セリフの領域、セリフの内容、擬音語や擬態語、登場人物の名前や顔、表情や服装、セリフの話者、登場人物の関係性など)について認識する技術を研究開発する必要がある。

こうした研究開発を推進するために、多くの人の手によってアノテーションが付与されたデータセットは重要であり、プロの漫画家によって描かれた 109 冊のコミックについて、アノテーションとともに公開されている Manga109 データセット[5][6]がある。他にも、Guérin ら[7]による eBDtheque データセットや Iyyer ら[8]による COMICS データセット、上野ら[9]による 4 コマ漫画に特化したストーリー構造のデータセットなどが挙げられる。このようにコミックに関する多くのデータセットが公開されており、様々な研究へと活用がなされている。

ここで、コミックの翻訳や内容に基づいた検索・推薦を高精度に行う際の課題の 1 つとして、セリフの発話者を推定することが挙げられる。我々はこれまでセリフからの話者推定に向けて、セリフ内のコミックの話者を手軽にアノテーション可能な、セリフ・話者対応付けデータセット構築システムを開発してきた[10](図 1~4)。また、コミック内のセリフの話者を読み取ることの困難性について整理するとともに、Manga109 データセットには含まれない、セリフと話者の対応付けを可能とするデータセットを構築してきた。さらに、得られたデータを踏まえ、セリフの話者自動推定に関する検討を行ってきた[11]。しかし、推定精度は全ての手法を組み合わせた 70.7% が最も精度の高いものであり、現状の手法のみで自動推定を行うには十分とはいえない結果であった。また、コミッ

クの特性を分析する際に参考としたセリフ・発話者対応付けデータセットは、2名のみで付与されたアノテーションであったため、その信頼性に問題があった。

そこで本研究では、我々が Manga109 をベースとして実現してきたセリフと発話者の対応付けを可能とするデータセット構築システム[10]を利用し、アノテーションを行う付与者の数を増やしてセリフ・発話者対応付けデータセットの拡張を行う。また、拡張を行ったデータセットについて複数の指標をもとに分析することで、人による評価のゆらぎや、アノテーション付与における話者推定が困難な場面について明らかにする。

2 関連研究

コミックを構成する要素について認識する技術に関する研究は様々なものがある。Nguyen ら[12]は、コミックにおけるコマの定義についての再検討およびその抽出手法についての提案を行っている。また、三原ら[13]は画像認識とクラウドソーシングを組み合わせたマイクロタスクにより、コミックにおけるコマの抽出システムの開発を行っている。田中ら[14]は、機械学習を用いてセリフの文字位置判定を行うことで吹き出し候補を検出し、その吹き出しの選択および分類を 86%の精度で実現している。また、Chu ら[15]はキャラクタの顔認識を、Tolle ら[9]はセリフの認識を高精度に行う手法を提案している。これらの研究は、コミックの画像やテキスト情報を用いて認識を行っており、より簡易で精度の高い手法が確立されつつあるといえる。こうした研究を更に推進するには、その正解となるデータセットが必要となる。本研究はこうした研究に資するものである。

コミックの研究開発を推進する技術として、平岡ら[17]はコマ同士のリンク関係を用いてコマの重要度を推定する手法を提案している。朴ら[18]は漫画に登場するキャラクタを用いたコミック検索システムの実現に向けて、キャラクタ特性分析を行っている。こうした研究は、コミックの内容を考慮した開発を行う際に必要な技術となり、その中の1つとして話者推定が挙げられる。話者推定に向けた研究としては、Rigaud ら[19]による吹き出しのしっぽからの距離による推定手法や、山本ら[20]がキャラクタの表示サイズやしっぽの方向などを用いた手法を提案し、その有用性を検証している。これらの研究ではコミックにおけるセリフと発話者の関係について、あらかじめ分析を行い、手法を確立する必要がある。本研究ではコミック内でのセリフと発話者の関係を分析することで、高精度な話者推定に向けたコミックの特性を明らかにしていくことを目指している。

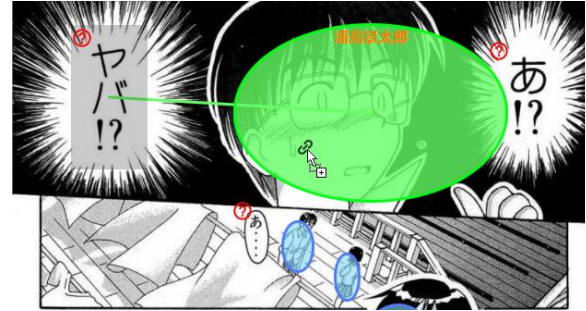


図1 ドラッグ・アンド・ドロップによるセリフ・発話者対応付けシステム

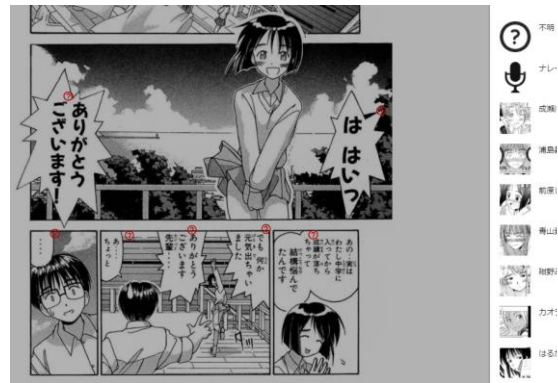


図2 キャラクターリストから発話者を選択する画面



図3 進捗状況を色により提示

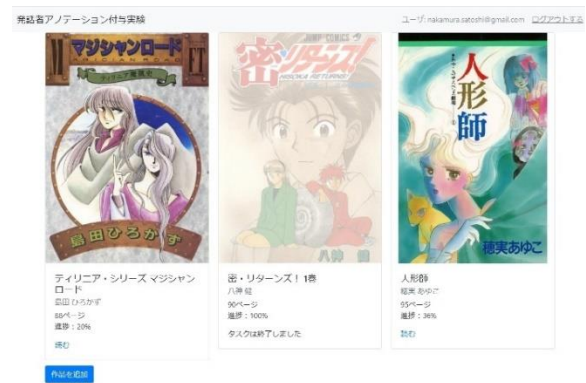


図4 アノテーション対象選択 UI

3 セリフ・発話者対応付けデータセットの拡張

我々が過去に行ったデータセット構築[11]では、Manga109の109冊全てのコミックに登場する1つのセリフに対して、2名がアノテーション付与を行っていた。先述の通り、1つのセリフに対してのアノテーション付与者が2名では、発話者の正誤判定を行うにはデータ数が不十分である。そこで、本研究ではデータ数をより確かなものにするため、我々がこれまでの研究で開発したセリフ・話者対応付けデータセットの構築システム[11]を利用し、アノテーション付与における付与者数を増やすとともに、その分析を行う。

3.1 セリフ・発話者対応付けシステム

我々が過去に実現したセリフ・発話者対応付けデータセットの構築手法および開発したシステムでは、Manga109データセットに含まれる109冊の作品を対象としており、総発話者数は147,918件となる。これらの各セリフについて、1セリフずつ順に登場人物のリストから手作業でアノテーションを付与していくのは手間のかかるものであり、現実的ではない。

そこで我々は、セリフは発話者の近くにあることが多いということに着目し、セリフをキャラクタの顔領域+体領域にマウスでドラッグ・アンド・ドロップすることにより割り当てを行う手法を提案し、アノテーション付与システムとして実現してきた(図1)。このシステムでは、割り当てを行う際にセリフと発話者が同一ページに登場していないことも考慮し、必要に応じてキャラクタリストから発話者を選択することも可能である(図2)。なお、発話者が分からない場合はキャラクタリストに存在する「不明」を選択し、状況の説明や注釈のようなキャラクタの発していないと考えられるセリフは「ナレーション」として付与してもらうものとなっている。また、コミックを読み進めることで、前のページにおける発話者が推定できるという状況を考慮し、ページ単位での管理インタフェースでコミック全体を把握しながらアノテーションの付与を行うことが可能となっている(図3)。

また、データセット構築においては、109冊もの膨大なコンテンツを対象とするものであるため、アノテーション対象を選択するUIも用意している(図4)。

3.2 データセットの拡張結果

前回のデータセット構築では、33人によりデータセットが構築されていたが、今回は新たに23人に依頼し、データセット構築の合計協力者数は56人となった。

この56人により付与されたセリフ・発話者のペアは合計749,856件となっており、Manga109における全てのコ

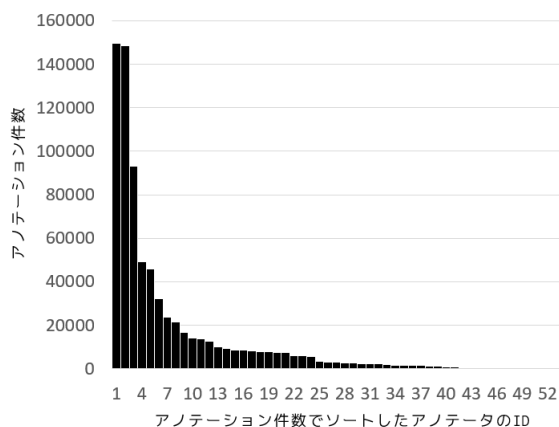


図5 各評価者のアノテーション付与件数

ミックの総発話が147,918件であることを考慮すると、1つセリフに対して平均約5人の付与者がいることになる。

図5にアノテーション付与者ごとのアノテーション件数を示す。この図において、横軸はアノテーションを行ったアノテーション付与者のID(アノテーション件数で降順ソート)、縦軸はそのアノテーション付与者のアノテーション件数を意味している。

この結果より、2名のアノテーション付与者が、それぞれ15万件近いアノテーションを付与していることがわかる。対象となるコミックの総セリフ数が147,918件であることから、ほぼすべての作品のセリフに発話者のアノテーション付けを行っていることがわかる。以上のことから、我々のシステムは、膨大なアノテーション対象があっても、セリフと発話者の対応付けを十分にこなすことが可能であると考えられる。一方、半数以上の付与者はほとんどアノテーションを付与していないため、その付与にはブレがあり、参考にならない可能性がある。

4 データセットの分析

4.1 前回のデータセットとの比較

表1に前回のデータセットでのアノテーション付与の一致度合い、表2に今回のデータセットでのアノテーション付与の一致度合いを示す。前回は付与者が2人で、今回は付与者が平均5人と増えているため、一致度合いは下がる。実際、同じ人物を選択していたものも86.9%から71.1%まで下がっていることがわかる。

表3は、コマ内に含まれるセリフと、そのセリフのあるコマ内に存在するキャラクタの数についての件数およびその割合を、前回と今回とで比較するため併記したものである。ここでは、該当セリフがあるコマに対する発話者候補の存在の有無についても示している。なお、コマ内に含まれるかの判定は各フレームの±50px内に該当セリフおよびキャラクタの描写がなされているかで行った。

¹ <https://nkmr.io/comic/speaker-dataset/>

表1 アノテーション付与結果
(前回の2人での付与結果)

意見	内容	データ数	割合
一致	同じ人物を選択	128,502件	86.9%
	「ナレーション」を選択	3,531件	2.4%
	「不明」を選択	414件	0.3%
不一致	異なる人物を選択	9,720件	6.6%
	「不明」を選択	2,433件	1.6%
	その他を選択	3,318件	2.5%

表2 アノテーションの付与結果
(今回の5人での付与結果)

意見	内容	データ数	割合
一致	同じ人物を選択	105,238件	71.1%
	「ナレーション」を選択	9,893件	6.7%
	「不明」を選択	3,616件	2.4%
不一致	異なる人物を選択	27,590件	18.7%
	「不明」を選択	1,456件	1.0%
	その他を選択	125件	0.1%

表3 セリフがあるコマとキャラクタの関係

	前回	今回
コマにキャラクタがない	8.5%	10.8%
コマにキャラクタが1人	37.1%	35.0%
コマにキャラクタが複数(2人以上)	54.4%	54.3%
発話者候補のみが存在	32.2%	37.0%
コマに発話者が存在している	85.3%	75.4%

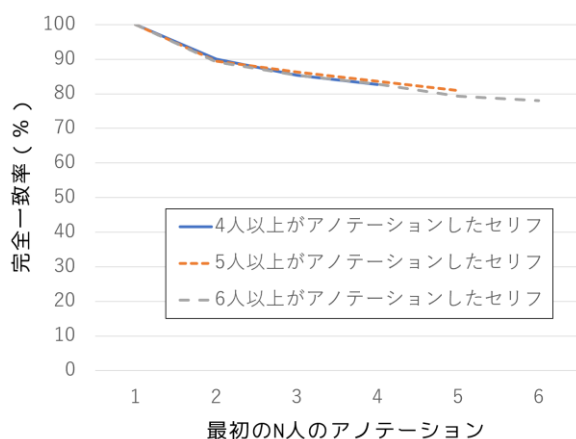


図6 アノテーション付与者N人目での完全一致率

また、発話者候補はアノテーション付与が1番多かったキャラクタとし、そのコマの登場キャラクタと比較することで判定を行った。

この結果より、コマにキャラクタがない割合やコマにキャラクタが複数といった項目については変化がないものの、コマに発話者候補のみが存在については増加し、コマに発話者が存在しているの評価については大きく

減少していることがわかる。この5~10%の値の変化は、正解としての設定誤りが生じる可能性があるため注意が必要である。

4.2 完全一致率と付与者数の関係

何人以上がアノテーションを付与すると妥当なものであると考えられるかを検討するため、付与者数ごとの完全一致率を比較する。

ここで図6は、4人以上がアノテーション付与したセリフ、5人以上がアノテーション付与したセリフ、6人以上がアノテーション付与したセリフそれぞれについて、最初のN人目までがアノテーションを付与したときの完全一致率を求めたものである。図の横軸は、最初のN人目(最初の2人目、最初の3人目など)がアノテーション付与した状況を意味しており、図の縦軸は完全一致率を示している。この結果より、4人、5人、6人いずれの条件においてもグラフはほぼ類似しており、アノテーション付与者が増えるに連れ、徐々に完全一致率が下がっていることがわかる。

ここで、前回構築したデータセットは2人のみによるものであったが、アノテーション付与者数を3人増やして5人にすると10%近く完全一致率が下がっていることがわかる。

4.3 発話者対応付けに関する評価指標と結果

セリフと発話者の対応付けに関する一致度の計算にあたり、全員の評価が完全一致している場合であれば問題ないが、完全一致は必ずしも評価指標として適切ではないと考えられる。また、一致しない状況でも、あるセリフの割り当てが2人の登場人物として付与されている場合はまだよいが、3人や4人の登場人物として付与されている場合は難易度が高くなると考えられる。

そこで、あるセリフに対して何人の発話者と間違えられたのかという variation, またあるセリフに対して最大でどの程度のアノテーション付与者が一致していたのかという max_match という指標を導入する。ここで variation は単純にあるセリフに対して付与された発話者の種類数, max_match は各アノテーション付与対象への付与人数を合計付与者で割ることで割合を算出したものの最大値である。例えば、あるセリフに対して、計10名の付与者がAというキャラへ8名、Bというキャラへ2名のアノテーション付与を行った場合、variation は2, max_match は0.8となる。なお、これらの項目において、コミックに登場するキャラクタへのアノテーション付与のみを分類対象とし、評価対象がナレーション・その他・不明と評価されたものについては除外して整理を行う。

次に、上記のデータセットの整理を行ったデータに対して、下記の評価一致度指標を定めた(式(1))。

$$\text{評価一致度指標} = \frac{2 \times \text{max_match} \times \frac{1}{\text{variation}}}{\text{max_match} + \frac{1}{\text{variation}}} \quad (1)$$

ここで評価一致度指標とは、max_match と variation の調和平均をとったものであり、アノテーション付与における評価のブレを表す指標である。本データにおいて、max_match の数値が低く、variation の数値が高いと人による評価差が大きいと考えられ、本指標の数値は低くなる。そのため、評価一致度指標による数値が高ければ、そのセリフに対するアノテーションが容易であり、一方で値が低い場合にはそのセリフのアノテーション付与が困難であるということになる。

図7は、Manga109の全てのセリフに対する評価一致度指標の値を導出し、作品ごとの評価一致度指標の値の平均を算出したものである。この結果より、1.0となる漫画は存在していないことがわかる。一方、0.8を割るような作品も存在していないことがわかる。

評価一致度指標が効果的に働くかどうかを確認するため、2つの作品(ARMS, 女王蟻)をピックアップし、各ページにおける評価一致度指標の値の平均を算出したものを図8および図9に示す。これらの図において、横軸はページ数を、縦軸は評価一致度指標の平均を意味している。これらのグラフより、ページによって評価一致度指標の値は大きく異なることがわかる。

図8および図9において、評価一致度指標の値が0.7以下であったページの一部を図10および図11に示す。この結果より、バトルシーンや暗闇のシーン、複数人が乗り込む宇宙船などが登場するシーンなどはアノテーション付与者にとって、セリフと発話者の対応付けが困難であることがわかる。

5 考察と展望

表3で示す通り、コマ内に発話者候補が描かれているコマの割合が、2人のアノテーション付与者のときと、5人のアノテーション付与者のときとでは大きく変わっていた。そのため、コマ内に複数の発話候補者がおり、複数の選択肢が生まれる場面において、アノテーション付与者が増えるとブレが生じるものと考えられる。

図6より、アノテーション付与者が2人のときに、セリフに対する発話者の対応付けが完全一致したからといって、その評価が適切でないことが明らかになった。具体的には、アノテーション付与者が最初の2人の場合と最初の5人の場合とで、10%も完全一致度が異なる。そのため、適切なアノテーション付与者数は十分に検討する必要がある。

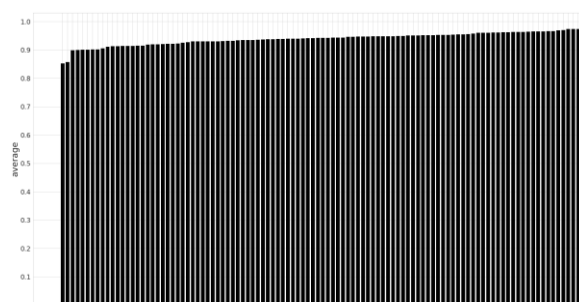


図7 各コミックにおける評価一致度値の分布

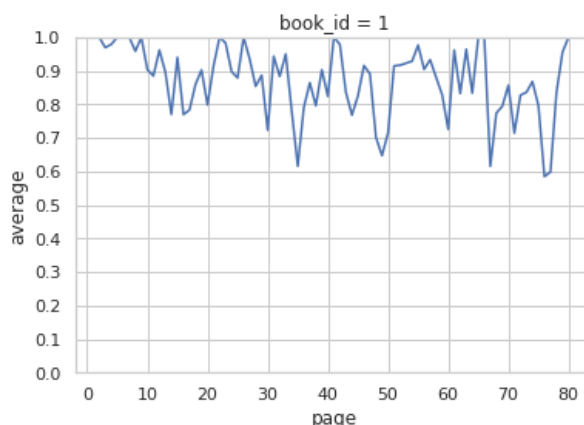


図8 ARMSにおける評価一致度値の分布

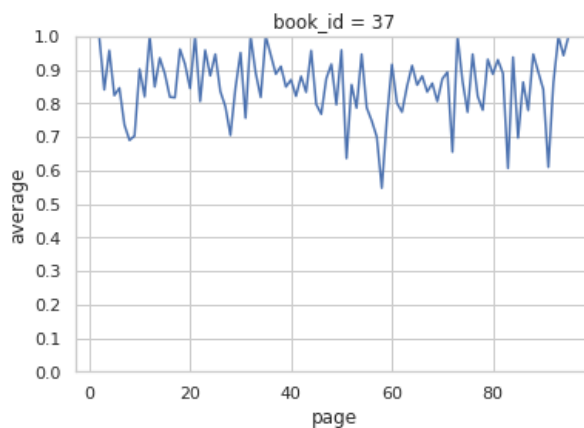


図9 女王蟻における評価一致度値の分布

完全一致によらない新たな評価一致度指標を定義し、各作品における評価一致度指標の値の平均を図7に示した。この結果より、その値が1.0や限りなく1.0に近い値はないため、全ての作品において評価一致度指標の値が低くなる場面が存在することがわかる。また、評価一致度指標は、SFやバトルといったジャンルにおいて特に低くなる傾向があり、図8や図9で示すように急激に値が低くなるページ(場面)が存在することがわか

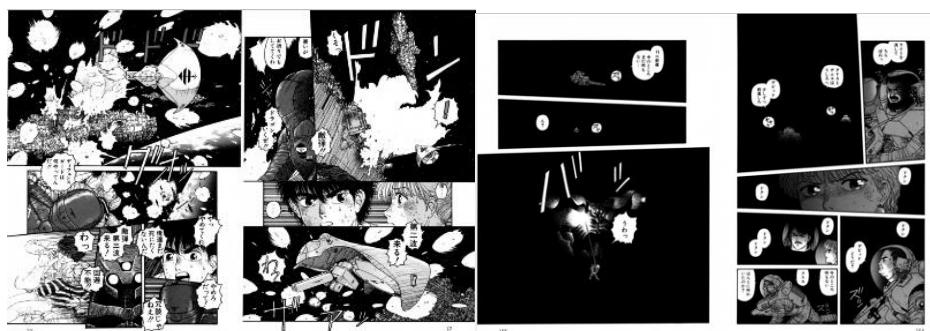


図10 戦闘シーンや暗闇のシーンは対応付けが困難 ©加藤 雅基「ARMS」



図11 戦闘シーンは対応付けが困難 ©大井 昌和「女王蟻」

った。また、それらの具体的なシーンでは、図10や図11に示すような戦闘シーンや暗いシーンといった、状況の把握が難しい場面であることが多かった。これらの場面ではコマ中にキャラクタが存在しない、または複数人存在しているといった特徴が見られたり、セリフにおいては、内言や身体状態を示す表現が多く存在し、吹き出しのしっぽがない場合やしっぽがキャラクタへは向けられていない、吹き出し自体が存在しないといった描かれ方の特徴も見られた。つまり、セリフ・発話者対応付けデータセットを構築するうえでは、単純で誰もが簡単に判断できるシーンでは1~2名のアノテーション付与者だけを割り当て、図10や図11などの難易度の高いシーンにはアノテーション付与者を多数割り当て、多数決などで判断する必要があると考えられる。

こうしたアノテーション難易度の自動推定については、今回構築したデータセットをもとに検討を行い、実現していく予定である。アノテーション難易度の自動推定が可能になれば、こうしたシーンはマイクロタスクで、こうしたシーンは1人で、こうしたシーンは人数を多くかけて多数決で、こうしたシーンは熟達者による判断でといったように、必要に応じてリソースを割り振り、効果的にデータセットが構築できると期待される。

6 おわりに

本研究では、コミック工学の研究に貢献することを目指して、Manga109におけるセリフと話者の対応付けを可能とするデータセットの拡張およびその分析を行った。ここでは、アノテーション付与者56人により、合計749,856件のアノテーションまでデータセットの拡張をした。また、これまでのデータセットと、新たに構築したデータセットとの比較を行い、アノテーション付与者が2人である場合の問題点、また人数を増やすとどのように完全一致率に変化が生じるかなどを明らかにした。さらに、完全一致率によらない新たな評価一致度指標を設定し、その指標を用いた分析を行うことで、人手によって発話者を推定することの困難な場面についての検討を行った。その結果、SFやバトルといったジャンルにおける、戦闘シーンや状態把握の難しいシーンで評価一致度指標の値が低くなることが分かった。また、それらのシーンにおけるコマやセリフの特徴について明らかにした。

今後は、各セリフに対してアノテーション付与の難易度を明確にしていくことで、発話者の自動推定へ向けての貢献を目指す。また、人による評価の差を明らかにしていくことで、人手によるアノテーション付与の必要数を動的に切り替える手法などを検討していく予定である。

謝辞

本研究の一部は JSPS 科研費 JP20K12130 の助成を受けたものです。

参考文献

- [1] インプレス総合研究所: 電子書籍ビジネス調査報告書 2020 入手先, <https://research.impress.co.jp/report/list/ebook/500995>, (参照 2021-10-20).
- [2] Mantra: マンガの超高精度な自動翻訳(オンライン), 入手先, <https://mangahot.jp/>, (参照 2021-10-20).
- [3] 佐藤剣太, 牧良樹, 中村聡史. 続巻への動機付けのためのコミック読書進度に応じたシーンのデータセット構築と自動推薦に関する検討, 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), No. G3-5, pp. 1-7, (2019).
- [4] Maki, Y. and Nakamura, S.: Do Manga Spoilers Spoil Manga?, The Sixth Asian Conference on Information Systems, pp.258-262, (2017).
- [5] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T. and Aizawa, K.: Sketch-based manga retrieval using manga109 dataset, Multimedia Tools and Applications, Vol. 76, No. 20, pp.21811-218388, (2017).
- [6] Ogawa, T., Otsubo, A., Narita, R., Matsui, Y., Yamasaki, T. and Aizawa, K.: Object Detection for Comics using Manga109 Annotations, arXiv:1803.08670, (2018).
- [7] Guérin, C., Rigaud, C., Mercier, A., AmmarBoudjelal, F., Bertet, K., Bouju, A., Burie, J., Louis, G., Ogier, J. and Revel, A.: eBDtheque: A Representative Database of Comics, 12th International Conference on Document Analysis and Recognition, pp. 1145-1149 (2013).
- [8] Iyyer, M., Manjunatha, V., Guha, A., Vyas, Y., BoydGraber, J., Daumé III, H. and Davis, L.: The Amazing Mysteries of the Gutter: Drawing Inferences Between Panels in Comic Book Narratives, IEEE Conference on Computer Vision and Pattern Recognition, pp. 7186-7195, (2017).
- [9] 上野未貴: 創作者と人工知能: 共作実現に向けた創作過程とメタデータ付与 4 コマ漫画ストーリーデータセット構築, 人工知能学会全国大会論文集, Vol. JSAI, (2018).
- [10] 阿部和樹, 中村聡史. 漫画における台詞発話者の自動判定に向けた技術的困難性による整理とデータセット構築手法の検討, 第 2 回コミック工学研究会発表会, pp.7-14, (2019).
- [11] 阿部和樹, 中村聡史. コミックのセリフと発話者対応付けデータセットの構築とその困難性, 第 3 回コミック工学研究会発表会, pp.7-12 (2020).
- [12] Nguyen Nhu, V., Rigaud, C. and Burie, J.: What do We Expect from Comic Panel Extraction?, 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), Vol. 1, pp. 44-49 (2019).
- [13] 三原鉄也, 石川夏樹, 豊田将平, 永森光晴, 杉

本重雄: 画像認識とマイクロタスク型クラウドソーシングを組み合わせたマンガのコマ領域の判定, 人工知能学会全国大会論文集 (JSAI2018), No. 4M105, (2018).

[14] 田中孝昌, 外山史, 宮道壽一, 東海林健二: マンガ画像の吹き出し検出と分類, 映像情報メディア学会誌, Vol. 64, No. 12, pp. (1933-1939).

[15] Chu, W. T. and Li, W. W.: Manga face detection based on deep neural networks fusing global and local information, Pattern Recognition, Vol. 86, pp. 62-72.

[16] Tolle, H. and Arai, K.: Method for Real Time Text Extraction of Digital Manga Comic, International Journal of Image Processing, (2011).

[17] 平岡 誉史, 山西 良典, 西原 陽子. コミックのコマ間のリンク関係によるコマの重要度推定についての一検証, 第 32 回人工知能学会, 1J3-01, (2018).

[18] 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, 第 32 回人工知能学会, Vol.JSAI2018, pp.1J302-1J302(2018).

[19] Rigaud, C., Le Thanh, N., Burie, J., Ogier, J., Iwata, M., Imazu, E. and Kise, K.: Speech balloon and speaker association for comics and manga understanding, 13th International Conference on Document Analysis and Recognition (ICDAR), pp.351-355, (2015).

[20] 山本和慶, 小川徹, 山崎俊彦, 相澤清晴. データドリブなアプローチを用いた漫画画像中の吹き出しの話者推定, 映像情報メディア学会技術報告, pp. 287-292, (2018).