

漫画における自動話者推定

山口理哉¹ 古田諒佑² 谷口行信¹ 日並遼太³ 石渡祥之佑³
 東京理科大学¹ 東京大学² Mantra 株式会社³

1 まえがき

近年、漫画のデジタル化と日本文化の広まりによって、国内外で漫画の人気が高まっている。それに伴い、レコメンドや自動翻訳等の関連サービス開発の需要が高まっており、機械による漫画の理解を可能にする技術が必要となっている。そこで本研究では、機械による漫画の理解と関連サービスへの応用を目的とし、その初期段階として自動話者推定手法の提案を行う。自動話者推定によって、漫画中のテキストとその話者(キャラクター)を紐づけることができれば、ストーリー分析に基づいた漫画の分類やレコメンド、キャラクターの個性や場面に沿った翻訳など、様々な関連サービスへの応用が期待できる。

2 提案手法

キャラクターとテキストの座標から計算した距離特徴と、テキストの内容から抽出した言語特徴の2つを用いた話者推定手法を提案する。距離特徴は画像内に存在する各キャラクターとテキストとのユークリッド距離を計算することによって得られる特徴である。また、言語特徴の抽出にはBERT[1]を用いた。最後に2つの特徴を統合して得られた特徴ベクトルによって多クラス分類問題を解き、話者推定を実現する。

距離特徴と言語特徴の統合の方法について以下の4つの方法を比較する。提案モデルを図1に示す。

- (1) 距離特徴と言語特徴を連結
- (2) 距離特徴をMLPに入力して得られた特徴と言語特徴を連結
- (3) 言語特徴から得た推定スコアと距離特徴の差を取る
- (4) 言語特徴から得た推定スコアと(1-距離特徴)の和をとる

3 実験

Manga109[2]に含まれるTaiyouNiSmashをデータとして使用し、キャラクター9人を話者推定の対象とした。加えて、話者推定のためのアノテーションデータとしてComic-Speaker-Dataset[3]を使用した。既存手法として[3]で提案された距離によるスコアリング方式の手法と比較し、さらに言語特徴のみを用いて推定した結果との比較も行った。評価指標には平均正解率を用いた。結果を表1に示す。既存手法の精度を超えることができなかった一方で、表2のように、キャラクターごとの正解率に注目した場合、発話に特徴のある一部のキャラクターについて改善が見られた。

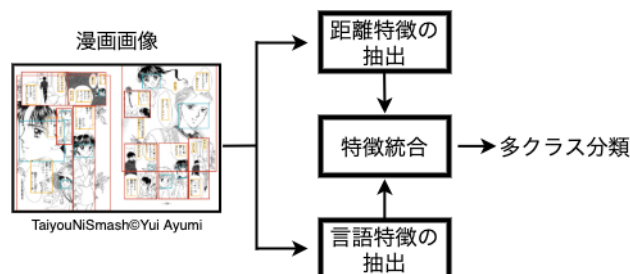


図1 提案モデル

表1 精度比較実験結果

手法	平均正解率
既存手法 [3]	0.673
言語特徴	0.153
提案手法 (1)	0.238
提案手法 (2)	0.320
提案手法 (3)	0.380
提案手法 (4)	0.429

表2 各クラスにおける正解率比較結果

キャラクター	A	B	C	D
テストデータ割合	0.12	0.01	0.55	0.01
既存手法 [3]	0.61	1.0	0.70	1.0
提案手法 (4)	0.56	1.0	0.47	0.0

...	E	F	G	H	I
...	0.15	0.01	0.07	0.04	0.03
...	0.64	0.0	0.82	0.33	0.80
...	0.72	0.0	0.91	0.0	0.20

4 まとめ

本研究では、機械による漫画の理解を目的とし、距離特徴と言語特徴を用いた自動話者推定手法の提案と、特徴の統合方法について検討を行った。今後はキャラクターの表情などのイラストに着目する手法や会話の流れを考慮した手法を開発し精度の向上を目指す。

参考文献

- [1] Jacob Devlin, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proc. NAACL*, pp. 4171–4186, 2019.
- [2] Kiyoharu Aizawa, et al. Building a Manga Dataset Manga109 with Annotations for Multimedia Applications. *IEEE MultiMedia*, Vol. 27, No. 2, pp. 8–18, 2020.
- [3] 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, pp. 7–12, 2020.