深層学習に基づく 4 コマ漫画の感情推定とマルチモーダル化への検討

第 1 グループ　高山 裕成

1. はじめに

近年, 深層学習を始めとする機械学習技術の大きな発展を受けて, 人工知能を用いた創作物理解が注目されている. しかし, 創作は高次の知的活動であるため, いまだに実現が困難なタスクである. 人の創作物の理解に関する分野の中でも漫画を対象とした研究は, 絵と文章から構成される漫画を対象とするため, 自然言語処理と画像処理の両方の側面を持つマルチモーダルデータを扱う分野である.

漫画を対象とした研究分野では様々な研究が報告されているが, その多くは画像処理に基づいた研究であり, 自然言語処理による内容理解を目指した研究は少ない. その大きな原因としては, 漫画が著作物であることに起因する研究用データの不足や漫画特有の言語表現の複雑さが挙げられる.

本研究では, 人工知能を用いた漫画の内容理解のために, まず自然言語処理を用いた漫画のキャラクタのセリフの感情を推定して, その上で漫画のコマの画像情報を加えたマルチモーダルな推定手法について検討する. そして, 実験結果からマルチモーダル化が精度にどのような影響を与えるのかについて考察した.

2. 要素技術

2.1. BERT

Bidirectional Encoder Representations from Transformers (BERT) [1] は, 2018 年に Google が発表した言語モデルであり, 文書分類や質疑応答といった様々な自然言語処理の幅広いタスクにおいて公開時点での最高性能を達成している. 本研究では日本語 Wikipedia より全 1800 万文を用いて事前学習させたモデル [2] (以下, 京大 BERT) 及び, 大規模日本語 SNS コーパスによって事前学習させたモデル, hottoSNS-BERT [3] を用いた.

2.2. illustration2vec

illustration2vec [4] は Saito, Matsui らが提案した画像のベクトル化手法であり, 既存の画像認識モデルのほとんどが ImageNet などの実画像を評価対象にしており, アニメや漫画といったイラストに対して評価をしていなかったことから, イラストのより合理的なベクトル化が期待できる. 本研究では筆者らが公開している事前学習済みモデルを用いた.

3. 提案手法

本研究では, 上野によって作られた 4 コマ漫画ストーリーデータセット [5] を用いて, 各セリフにアノテートされた感情ラベルを推定するタスクを解き, その精度を確認する.

そのマルチモーダルな推定手法として, 図 1 に提案手法の概要を示す. Text Embedding 層への入力として, あるセリフを選んだ時に, Image Embedding 層への入力をこのセリフが含まれているコマの画像全体とする. そして, それぞれの層から得たものをセリフベクトル・コマベクトルとし, これらを結合したものを識別器への入力とすることでセリフの感情ラベルを推定する.

4. 実験

本研究で用いるデータセットには 7 種類の感情ラベル（ニュートラル, 驚愕, 喜楽, 恐怖, 悲哀, 憤怒, 嫌悪）と, アノテーション不備によるラベル不明 (以下, “UNK” とする) の全 8 種類が含まれているが, データ数と解析の難しさの問題から, 喜楽のみを正例, その他の感情ラベルを負例とする 2 クラスに分類し, これを推定する.

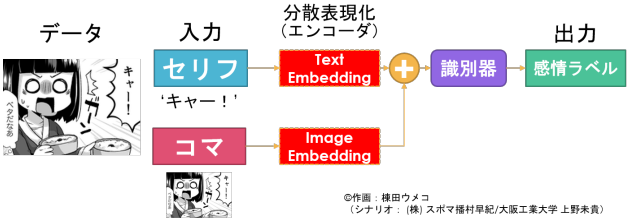

図 1: 提案手法の概要

4.1. セリフ 1 文のみを入力とする感情推定

...hottoSNS-BERT vs 京大 BERT

4.2. マルチモーダルな感情推定の検討

...hottoSNS-BERT vs 京大 BERT

5. まとめと今後の課題

実験結果からマルチモーダルな感情推定の場合, ...

hottoSNS-BERT の方が... 合理的...

今後の課題として, 現在の 4 コマ漫画ストーリーデータセットのみでの訓練・テストでは精度の向上が困難であることから, データセットの拡張が急がれる. この問題の解決策としては Manga109 やその他データセットを併用した半教師あり学習や, 人手による新しいデータの作成が挙げられる. データ作成の際にはオリジナルのデータからセリフ部分の文字を白抜きにしたコマの画像に対して, 物語の一貫性などの制約を設けた上で新たなセリフを作成し, 対応する感情ラベルを付与してもらうことで, このデータセットの “作者による感情ラベルのアノテーション” という特徴を保ったままデータ拡張が可能であると考えられる. また, セリフから得たベクトルとコマの画像から得たベクトルの結合方法やネットワークの構造の最適化についても更なる工夫が必要である.

参考文献

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[2] 黒橋 禎夫. Bert による日本語構文解析の精度向上. 言語処理学会 第 25 回年次大会, pp. pp.205–208, 2019.

[3] Sakaki, Takeshi, S. Mizuki, N. Gunji. Bert pre-trained model trained on large-scale japanese social media corpus. 2019.

[4] M. Saito and Y. Matsui. Illustration2vec: A semantic vector representation of illustrations. In *SIGGRAPH Asia 2015 Technical Briefs*, SA ’15, New York, NY, USA, 2015. Association for Computing Machinery.

[5] 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, 2018.