

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コミック工学など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。

コミック工学の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。その大きな原因のひとつとしては、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ー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。また従来のニューラルネットワークを用いた自然言語処理モデルは、特定のタスクに対して 1 つのモデルを用いてきたが、BERT は転移学習により、1 つのモデルで様々なタスクに対応できる。

2.2. illustration2vec

illustration2vec [2] は Saito, Matsui が提案した画像のベクトル化手法であり、Danbooru と Safebooru から 100 万枚のイラストを用いて学習した事前学習済みモデルが公開されている。illustration2vec が扱った問題として、イラストに対する画像認識の難しさがあり、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。illustration2vec はそれらと比較してイラストのより合理的なベクトル化が期待できる手法である。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使って 4096 次元のコマ画像のベクトルを獲得する。

3. 実験

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 [3] を用いる。これは同一プロットの下、幾人かの漫画家によって描き下ろされた 4 コマ漫画で構成されており、各作者によってセリフの感情ラベルがアノテートされている。また、このデータセットには 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、

これを推定するタスクを解く。実験 1 では... 実験 2 では... 実験 3 では...

4. まとめと今後の課題

実験より、

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] 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.
- [3] 上野. 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.