卒 業 研 究 報 告 書

題　　目

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

研 究 グ ル ー プ　知能情報第 1 グループ

指 導 教 員　森 直樹 教授

平 成　32　年（　2020　年）度 卒 業

（No.　1171201102　）　高山 裕成

大 阪 府 立 大 学 工 学 部 知 能 情 報 工 学 科

目 次

図 目 次

表 目 次

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 [1] 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データが十分ではないという点が挙げられる. また, 漫画に含まれるテキストには, 口語表現, 擬音語, 表記揺れといった漫画特有の言語表現を含み, これらの扱いについて考慮する必要がある. そして, 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も課題となっている.

本研究では人工知能を用いた漫画の内容理解のために, 漫画におけるキャラクタのセリフのマルチモーダルな感情推定を目的とする. まず自然言語処理を用いた漫画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化について検討する.

以下に本論文の構成を示す. まず, 2 章ではコミック工学に関連するデータセットについて, また 3 章では本研究で用いる要素技術について概説する. 次に, 4 章で関連研究について概観する. さらに, 5 章では漫画のセリフのマルチモーダルな感情推定を行うための提案手法について述べる. そして, 5 章において, 実験手法とその考察を示す. 最後に, 6 章で本研究の成果をまとめた上で, 今後の課題について述べる.

2　コミック工学に関連するデータセット

2.1　Manga109

Manga109 [2] は,

2.2　4 コマ漫画ストーリーデータセット

4 コマ漫画ストーリーデータセット [3] は, 本章では，実験に関連する要素技術について説明する.

3　要素技術

本章では，本研究に関連する要素技術について説明する．

3.1　自然言語処理に関する要素技術

自然言語の単語や文を計算機上で表現するための分散表現獲得手法について説明する．

3.1.1　形態素解析

形態素とは日本語などの自然言語において意味を持つ最小の単位のことであり, 文を形態素に分割し, 各形態素の品詞などを判定する技術を形態素解析という．英語の文では, 単語と単語の区切りがほとんどの箇所で明示的に示される．このため、形態素への分割処理は簡単なルールに基づいて行われる場合が多い．一方で, 日本語の文は単語間の区切りが英語ほど明確でないため, 形態素への分割は困難かつ重要である．

形態素解析器としては, MeCab [4] や Juman++[5] などが存在する．

3.1.2　局所表現

自然言語の単語を計算機上で表現する手法として, 最もシンプルなものが局所表現である．単語の代表的な局所表現の 1 つに One-hot 表現がある．One-hot 表現は単語をベクトルの各次元に 1 対 1 対応させる表現方法である．非常に単純な手法であり, 実装が容易であるという利点がある．一方で, One-hot 表現では語彙数とベクトルの次元数が等しくなるため, 語彙数の増大とともにベクトルの次元数も増大し, ベクトル空間がスパースになってしまう問題がある．また, 各単語がベクトル空間上で等距離に配置されてしまうため, 単語間の意味的な関係性については定義できないことも大きな問題である．

3.1.3 分散表現

局所表現の問題点を解決するために考案された手法が分散表現である. 分散表現は各概念をベクトルの単一次元ではなく複数次元の実数で表す. 単語の分散表現は，類似した文脈で使用される単語は類似した意味をもつ，という分布仮説を基盤としている. 単語を実数値密ベクトルで表現することにより, 単語間の意味的な関係性をベクトル空間上での類似度として定義できるという大きな利点がある.

3.1.4 Word2Vec

Word2Vec [6] は単語の分散表現を獲得する手法の 1 つである. この手法は, 同じ文脈で出現する単語は類似した意味を持つと予想されることに基づいており, 写像されたベクトルは, One-hot 表現のような局所表現と異なり, 単語間の意味を考慮した類似度測定や,「王様」−「男」+「女」=「女王」のような単語間の意味における演算などができるようになる.

Word2Vec では, 自己から周りの単語あるいは周りの単語から自己を予測することにより分散表現を獲得する. 前者の手法を Skip-gram といい, 後者の手法を Countinuous Bag-of-Words (CBOW) という.

3.1.5 Doc2Vec

Doc2Vec [7] は Word2Vec をベースとした, 文書をベクトル空間上に写像して分散表現を得る自然言語処理の手法である. Paragraph ID は各文書と紐づいており, 単語の学習時に一緒にこの Paragraph ID を学習することで文書の分散表現を獲得する. このベクトルを用いると文書間の類似度の算出や文書間での加減算が可能になる.

CBOW を拡張したモデルを Distributed Memory モデルといい, Skip-gram を拡張したモデルを Distributed Bag-of-Words という.

3.1.6 BERT

Bidirectional Encoder Representations from Transformers (BERT) [8] は, 2018 年に Google が発表した言語モデルであり, 複数の双方向 Transformer に基づく汎用言語モデルである. これまでの言語モデルは特定の学習タスクに対して 1 つのモデルを用いてきたが, BERT は大規模コーパスに対して事前学習を施して, 各タスクに対して fine-tuning をすることで, さまざまなタスクに柔軟に対応することができる. さらに, 以前はモデルごとに語彙を 1 から学習させるため, 非常に多くの時間とコストがかかっていたが, BERT ではオープンソースで公開されている文脈を既に学習させた Pre-Training BERT モデルを使用することで短時間で学習ができる.

BERT の 事前学習では, 周囲の単語からある単語を予測する Masked Language Model (MLM) と 2 つ目の文章が 1 つ目の文章の次の文章であるかを予測する Next Sentence Prediction (NSP) によりモデルを学習する.

3.2 画像処理に関する要素技術

画像処理に関する要素技術について説明する.

3.2.1 VGG

3.2.2 illustration2vec

4　関連研究

本章では，実験に関連する要素技術について説明する．

5　数値実験

本章では，実験について説明する．

5.1　実験概要

BERT および MLP を用いて新聞記事データの段落間の接続詞の有無を推定する実験をした．

5.2　データセット

5.2.1　使用データ

本稿では叙述的な文章として毎日新聞データセット[1] の新聞記事を用いた．このデータセットにはジャンルごとに 2008 年から 2012 年までの記事がある．そのなかのジャンルが国際のもので本文が 10 行以上ある 5000 記事を用いた．それぞれの文章に対し，“■”や “◇” , また感嘆符といった記号を除去し，“＜＞” や “《》” の間に書かれる注釈や作者名等を除いた．その上で，データセットの性質上，数字の羅列などを含む記事や，箇条書された記事が含まれているので，そのような記事を取り除き，文章として乱れていない記事を使用データとして扱った．

5.3　実験準備

実験では新聞記事データの 1 から 3 面の本文を用いた．まず，新聞記事データの各本文を段落ごとに分けて，すべて分かち書きする．次に，はじめの段落を除いて，各段落のはじめに接続詞があるものについてはその接続詞を “[MASK]” に変換する．そうでないものは段落の始めに “[MASK]” という単語を追加する．続いて，接続詞の直後に “、” があるものは，それを取り除く (接続詞があることが容易に推定できてしまうため)．前後 2 段落を “[SEP]” でつなげたものを 1 データとして扱う．これらのトークンを含めて単語数が 248 個より多いデータを除く．

[1] http://www.nichigai.co.jp/sales/mainichi/mainichi-data.html

これにより各データには 2 段落あり, それらは “[SEP]” でつながっている. さらに, その直後には必ず “[MASK]” がある. その “[MASK]” は, 通常の “[MASK]” のように, 接続詞を隠したものもあれば, ダミー (その部分には単語は入らない) もある.

5.4　実験

実験準備で得られた各データに対して, そのデータの 2 段落間の接続詞の有無を推定した. 今回は, データの不均等性に対してはデータ数を 1：1 にすることで対応した. また, 今回の実験では, 前回までで精度の良かった “[MASK]” 部分の分散表現に着目する方法のみをした. まず, 先程のデータを BERT の入力として, 得られた各単語ベクトルから “[MASK]” 部分の分散表現を得る. それを 3 層 MLP によって 2 次元にして, 接続詞の有無を推定した. 表 5.1 に実験時のパラメータを示す. 学習は BERT の最終層および MLP に対してした. 学習率および最適化アルゴリズムは optuna によって調整した.

表 5.1: 実験時のパラメータ

パラメータ	値
入力層の次元数	768
隠れ層のノード数	768
出力層の次元数	2
バッチサイズ	2
BERT の学習率	0.000105748
MLP の学習率	0.00004541588
最適化アルゴリズム	Adam
活性化関数 (隠れ層)	ReLU
活性化関数 (出力層)	Softmax function
目的関数	categorical cross entropy
学習終了条件	2 epoch

5.5　実験結果

データ数があまり多いわけではないので, 5 分割検証を行い, その際の精度の平均値, 標準偏差を比較した. また, ベースラインは, すべてをランダムに選択した際の期待値とした. 表 5.2 に 5 分割交差検証をしたときの平均及び標準偏差を示す.

表 5.2: 交差検証の結果

パラメータ	値
正解率	0.7502 (0.0176)
F 値	0.6656 (0.06164)

6　結論

本研究では,

謝辞

年 月 日

参考文献

[1] 松下光範. コミック工学：マンガを計算可能にする試み. 日本知能情報ファジィ学会 ファジィ システム シンポジウム 講演論文集, Vol. 29, pp. 199–199, 2013.

[2] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, Vol. 76, No. 20, pp. 21811–21838, 2017.

[3] 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.

[4] T. KUDO. Mecab : Yet another part-of-speech and morphological analyzer. *http://mecab.sourceforge.net/*, 2005.

[5] 京都大学大学院情報学研究科黒橋・河原研究室. 日本語形態素解析システム juman++ version 1.0. 2016.

[6] Greg Corrado Tomas Mikolov, Kai Chen and Jeffrey Dean. Efficient estimation of word rep-resentations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[7] Quoc V. Le and Tomás Mikolov. Distributed representations of sentences and documents. *CoRR*, Vol. abs/1405.4053, , 2014.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.