

ファインチューニングを用いた畳み込みニューラルネットワークによる皮肉検知

大原 虎太郎 平原 誠
法政大学大学院

1. はじめに

皮肉検知を目的とする先行研究は数多く行われているが、日本語の皮肉を検知する研究は少ない。本研究では、日本語テキストに対してCNNを用いて皮肉の検知を行う。また、ファインチューニングによって精度の向上を図る。

2. ファインチューニング

ファインチューニングの概略を図1に示す。ファインチューニングは、目的のタスクに関連する別タスクの学習済みネットワークのパラメータを、目的のタスクのネットワークのパラメータの初期値として学習を行う手法である。ファインチューニングを行うことで、少ないデータでも高い精度を出すことができる。

3. 提案手法

公開されている皮肉・非皮肉テキストデータセットは存在しないため、データセットの作成から行う必要がある。ニューラルネットワークにおいてデータ数は非常に重要であるが、大規模なデータセットの作成は設計者に極めて大きな負担がかかる。一方で、テキストのネガティブ・ポジティブを識別するネガポジ分析は、多く通り組まれているタスクであり、データセットも存在する。

そこで本研究では、「皮肉」は感情に起因すると仮定して、日本語テキストのネガポジ分析を行うCNN（以下、ネガポジ分析CNN）の事前学習を行う。その後、日本語テキストの皮肉と非皮肉を識別するCNN（以下、皮肉検知CNN）を下流タスクとしてファインチューニングをする。

4. データセット

ネガポジ分析では、Twitter日本語評判分析データセット[3]を使用した。データ数はネガティブ・ポジティブともに6,400件で、そのうちトレーニングデータを5,120件、バリデーションデータとテストデータをそれぞれ640件とした。

皮肉検知では、Twitterの投稿を収集し使用した。先行研究[1]に従い、「(皮肉)」という表現を含む投稿を皮肉テキスト、含まない投稿を非皮肉テキストとして収集した。データ数は皮肉・非皮肉ともに1,280件で、そのうちトレーニングデータを1,024件、バリデーションデータとテストデータをそれぞれ128件とした。

5. 実験と結果

まず、ネガポジ分析のデータをCNNに入力し学習をさせた。ネガポジ分析CNNのネットワーク構成を図2に示す。出力はネガティブである確率とポジティブである確率の2出力である。

次に、ネガポジ分析CNNの学習済みパラメータを利用し、作成したトレーニングデータを入力して皮肉検知CNNをファインチューニングした。ネットワーク構成は図2に示したネガポジ分析CNNと同じであるが、出力は皮肉である確率と非皮肉である確率の2出力であるため、出力層のパラメータのみランダムな数値で初期化した。

また、ファインチューニングをせず、ランダムな数値で全てのパラメータを初期化して学習を行う同構成の皮肉検知CNNを作成し、学習結果を比較した。

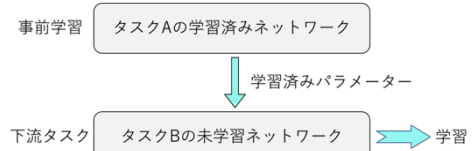


図1 ファインチューニングの概略

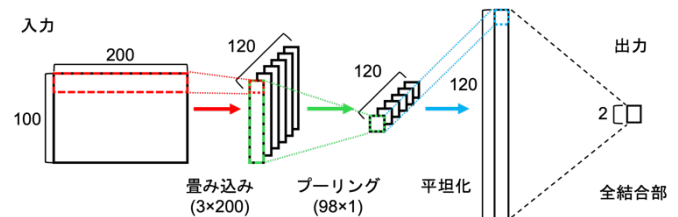


図2 ネガポジ分析CNN

表1 皮肉テキストのテストデータに対する識別結果

	正解率	適合率	再現率	F値
ベースライン(LSTM+Attention)		0.773	0.773	0.772
ファインチューニング無し	0.758	0.781	0.719	0.748
ファインチューニング有り	0.823	0.817	0.834	0.825

比較のため、先行研究[1]で行われたRNNベースのネットワークであるLSTMとAttention機構による皮肉検知の実験をベースラインとする。ここで、先行研究[1]で用いられたデータ数は42,000件である。

表1にベースラインの評価と、テストデータに対する提案手法の評価を示す。ファインチューニング無しのCNNと比べ、ファインチューニング有りのCNNは6%程度高い約82%の正解率を出せており、その他の評価指標についても、ファインチューニング有りのCNNの方が高くなった。

6. 考察

ファインチューニングを用いた機械学習では、多くの場合事前学習の精度が下流タスクの精度に影響する。そのためCNNによるネガポジ分析の精度を向上させれば、皮肉検知の精度も向上すると考えられる。

7. 今後の課題

事前学習であるネガポジ分析の精度を向上させ、下流タスクである皮肉検知の更なる精度向上を図る。

英語の先行研究[2]では、話し手の性格、テキストの「喜び」「驚き」「怒り」といった感情を予測する別のネットワークの出力を、皮肉検知の入力に加える手法が提案されている。このような、皮肉検知に有効な新たな特徴量を探し出すことも今後の課題とする。

参考文献

- [1] 肥合ほか, 言語処理学会発表論文集, pp.829-832, 2018.
- [2] Poria, *et al.*, Conference on Computational Linguistics, pp.1601-1612, 2016.
- [3] Suzuki Lab., Twitter日本語評判分析データセット, http://www.db.info.gifu-u.ac.jp/sentiment_analysis/, (参照2021)