

皮肉データセットを用いた深層学習による皮肉推定手法

第 1 グループ 多田 瑞葵

1. はじめに

皮肉表現とは、文章の文字通りの意味と書き手の意図とが異なる修辞表現である。皮肉表現の理解は、人間にとっては凡そ容易である一方で、機械にとっては未だ困難である。そのため機械による皮肉推定は、文章の理解や感情分析等の分野でも課題となっている。

本研究では皮肉表現にドメイン依存性が存在するという仮説のもと、SARC データセット [1] を用いて文章の話題ごとの皮肉推定に取り組んだ。また文脈を考慮できるモデルとして深層言語処理モデル BERT [2] を使用した。

2. BERT

Bidirectional Encoder Representations from Transformers (BERT) は、Transformer による双方向のエンコーダーを用いた言語モデルである。文章を文頭と文末の双方向から学習することによって文脈を理解することを可能としている。様々な事前学習済みモデルが発表されており、ファインチューニングすることで各タスクに応用可能である。本研究では、Wikipedia の英語記事で事前学習されたモデルである bert-base-uncased¹ を使用した。

BERT は文の入力に対して、入力された文および文に含まれる各単語に対応する分散表現を出力する。また特殊トークンが用意されており、文頭の “[CLS]” トークンに対応する分散表現を入力された文全体の分散表現とみなして分類問題に使用することができる。

3. SARC データセット

本研究では、自己注釈付き Reddit コーパス (Self-Annotated Reddit Corpus; SARC) を使用した。このデータセットは Reddit² に投稿されたコメントから構築されており、皮肉ラベルが付与されている。各コメントには、投稿者の識別子、subreddit と呼ばれる投稿トピック、ユーザ評価、投稿日時、親投稿の情報も付与されている。親投稿とは、最初の投稿から各コメントに至るまでの一連の投稿であり、全てのコメントは 1 つ以上の親投稿を保持している。図 1 にデータの例を示す。本研究では、皮肉推定の対象となるコメントと、その直接の投稿元である 1 つの親投稿のコメントを実験に使用した。

本研究では SARC 2.0 main の均衡データセットから subreddit ごとのデータセットを作成し、テストデータが 1,500 件以上であった 4 つの subreddit (politics, AskReddit, worldnews, pcmasterrace) を対象に皮肉推定をした。また main からデータをランダムサンプリングした random データセットを作成し、比較対象とした。

4. 実験

データセットの各コメントが皮肉か非皮肉かの二値分類をした。まず、コメントを BERT に入力し、出力された分散表現から “[CLS]” トークンに対応する分散表現を取得した。そして取得した分散表現を線形層に入力し、皮肉か非皮肉かを推定した。その際、訓練データを用いて BERT をファインチューニングし、テストデータに対する精度を確認した。

subreddit : AskReddit	
親投稿	What's something that's completely legal, but that pisses you off when you see someone doing it?
	People reclining their seats fully on aeroplanes.
非皮肉	There is a special hell for these people.
皮肉	yes people who spell airplanes wrong are just horrible

図 1: データセットのデータ例

表 1: 実験結果

subreddit	Accuracy	Precision	Recall	F1 score
politics	0.730	0.722	0.748	0.735
AskReddit	0.619	0.630	0.580	0.604
worldnews	0.690	0.655	0.803	0.721
pcmasterrace	0.648	0.635	0.698	0.665
random	0.653	0.664	0.618	0.640

5. 結果と考察

表 1 にテストデータの実験結果を示す。太字の項目は random の評価値を上回ったことを示している。politics, worldnews, pcmasterrace の F1 値は random を上回り、politics ではどの評価値においても random を上回る結果となった。このことから、ドメインごとに皮肉を学習、推定することで、皮肉推定の精度が向上する場合があると考えられる。一方で、AskReddit ではどの評価値においても random を下回る結果となった。このことから、AskReddit における皮肉推定が困難であり、データセット全体の推定の際に精度を下げる要因の一つとなることが考えられる。

6. まとめと今後の課題

本研究では、BERT を用いた皮肉推定に取り組み、皮肉表現のドメイン依存性について確認した。皮肉表現にドメイン依存性が存在するという仮説のもと、subreddit ごとのデータセットを作成して皮肉推定をし、精度を比較した。結果として、subreddit ごとの皮肉推定することで精度が向上したものと低下したものがあり、ドメインを考慮して皮肉推定する手法の有効性と、皮肉推定の困難なドメインの存在を示すこととなった。

今後の課題として、ドメインごとの弱学習器を用いたアンサンブル学習によって皮肉推定に取り組むことや、皮肉推定に使用する文脈情報やメタデータを拡大することが挙げられる。

参考文献

- [1] M. Khodak, N. Saunshi, and K. Vodrahalli. A large self-annotated corpus for sarcasm. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA).
- [2] J. Devlin, M. Chang, K. Lee, and K. Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.

¹<https://huggingface.co/bert-base-uncased>

²<https://www.redditinc.com>