5P-01

キャプション生成を併用した画像感情分析

齋藤　優輝　　数藤　恭子

東邦大学

1　はじめに

感情分析の研究は近年盛んに行われている．中でも画像の感情分析は推薦システムや広告などといったマーケティングや，画像の共有・検索の円滑化などコミュニケーションへの応用も期待されている分野である．

画像の感情分析を行う手法として，画像に付随している文章情報も利用する方法がある．元々，感情分析は自然言語処理の領域で研究されていたこともあり，文章の感情分析の手法も使用できるこの手法は高い精度が出ている．

しかし，画像のみのデータセットではこの手法は使用できず，また画像に文章情報が付随していても，画像と関連が低い文章情報も少なくないため，ノイズとなるそれらをデータセット作成時に取り除く必要があり，かかるコストが大きいのが課題である．

そこで本研究では，画像からキャプションを生成し，これを文章情報として画像の感情分析に用いる方法を検討する．これによって，画像のみのデータセットを用いても，文章情報が付随したデータセットを学習した場合と同様に，感情分析の精度を高められるかどうかの検証を行う．

2　関連研究

画像の感情分析に使用されるモデルとして，Residual Network(ResNet)がある．残差学習の仕組みによって層を深くしても性能の悪化が起こりづらいモデルであり，クラス分類問題として感情分析を行うことができる．

先行研究[1][2]では，画像に付随するテキストやタグから，感情極性辞書を使って感情極性情報を抜き出し，それを画像と共に学習・推論させることで，画像単体を使用する場合と比べて画像の感情分析の精度が上がることが確認されている．また，画像情報，付随テキスト情報，感情極性情報を用いて学習させたモデルの推論時に画像情報のみを入力した場合でも，画像情報のみで学習させたモデルに画像情報のみを入力した場合より精度が向上することが示されている．

また、他の先行研究[3]では，本研究と同様に，画像からキャプション生成を経由して画像の感情分析を行っている．そこではあらかじめ用意した感情極性情報の多い形容詞と名詞のペア 1200 組から画像との関連が深い４組を抜き出し，生成したキャプションに追加することで，画像から生成したキャプションは感情極性情報が少ないという問題を軽減している．これに対し本研究では，キャプション生成を経由した感情極性と，元の画像からの感情極性を統合する効果を検証する．

Image sentiment analysis with caption generation

Masaki Saito, Toho University

Kyoko Sudo, Toho University

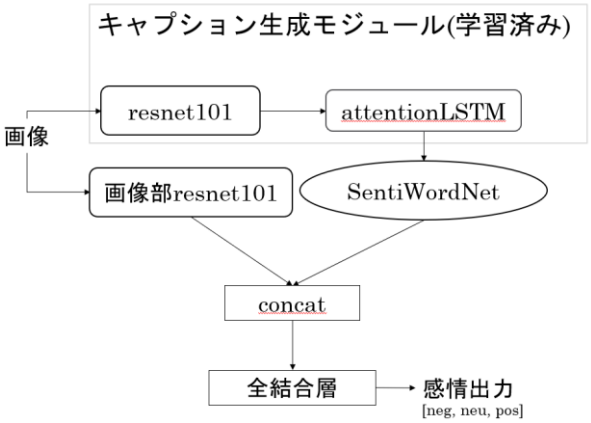

図１　提案手法全体図

3　提案手法

提案する感情分析モデルを図 1 に示す．画像を入力として，image sentiment score と caption sentiment score の２つの感情極性スコアを算出し，全結合層を用いて 2 つのスコアを統合し，最終的な感情極性を導く．最終出力である感情極性は Negative, Neutral, Positive のいずれかである．image sentiment score は，ResNet101(以下，画像部 ResNet101 と記す)を用いて入力画像から直接算出する．caption sentiment score は，学習済みキャプション生成モジュールによって入力画像から文章を生成し，その文章から算出する．

文章からのスコアの算出には，単語概念毎にポジティブ，ネガティブの極性スコアが付与されているデータである SentiWordNet[4]を用いる．1 つの単語に対して，ポジティブのスコアとネガティブのスコアが両方付与されている場合もあるため，文章を成す各単語のスコアは([ポジティブのスコア]-[ネガティブのスコア])として定義した．文章のスコア(caption sentiment score)はその合計として，実数値で出力される．

学習済みキャプション生成モジュールは ResNet101 をエンコーダー，Attention 付き LSTM をデコーダーとしたモデルである．キャプション生成を経由し，間接的に画像からスコア算出をする．２つのスコアを併用することで，生成キャプションに感情極性情報が少なかった場合でも従来通り画像から感情極性を抜き出し、精度を維持することを期待する．

4 実験

4.1　データセット

本研究では，画像とキャプションの組のデータセットと，画像に対応する感情極性がアノテーションとして付与されているデータセットを用いる．前者はキャプション生成モジュールの学習に用いるもので，COCO データセットを使用した．画像数は 2922，1 画像につき５つキャプションが付与されているためデータ総数は 14610 である．後者は ClowdFlower データセット[6]を改変したものであり，感情分析モデル全体の学習に用いる． 2400 の画像にクラウドソーシングによって 5 段階の感情極性いずれか

(HighlyNegative,negative, neutral, positive, HighlyPositive)が付与されていたが，そのうち HighlyNegative と HighlyPositive をそれぞれ negative, positive にマージし，3 値のデータに改変したものを使用した．

4.2　学習方法

提案する感情分析モデルは，事前学習済み ResNet101 によるエンコーダと， AttentionLSTM によるデコーダからなるキャプション生成モジュールに対し，COCO データセットを用いて学習する．その後，画像のみを入力とし，改変データセットを用いて感情分析モデル全体の学習を行う．この時，キャプション生成モジュールの重みは固定し，画像部 ResNet101 と全結合層のみが学習する．学習の際の損失は，最終的な出力である negative, neutral, positive それぞれへの確率と，画像に付与された正解ラベルとの交差エントロピー誤差とする．推論時は，学習時と同様に画像のみを入力とし，最も出力値が大きい感情をその画像の感情とした．

4.3　実験項目

今回は，以下の項目について実験を行った．

（実験 1）キャプション生成モジュールの精度評価

キャプション生成モジュール単体での性能について，positive と negative の適合率で評価する．

（実験 2）キャプションの極性の統合効果の評価

感情極性を無作為に判定した場合(random)，画像のみから感情極性を判定した場合(RN101only)，キャプション生成モジュールで生成したキャプションからの極性を統合し判定した場合(RN101+C&D)を比較し，提案手法の効果を評価する．

また，キャプションの極性の判定結果の変化が全体の判定結果に与える影響を調べるため，SentiWordNet[4]の代わりに，対象データのドメインとは異なるドメインの極性辞書 Lexicoder Sentiment Dictionary(LSD)[5]を用いた場合を比較する．

4.4　実験結果

（実験 1）キャプション生成部のみの感情極性評価の結果，True Positive Rate(positive と判断され実際に positive だった割合)は 24.0%，True Negative Rate(negative と判断され，実際に Negative だった割合)は 74.4%であった（neutral は除いて算出）．

（実験 2）

表 1 に実験 2 の結果を示す．「Dict [4]」，「Dict [5]」それぞれ感情極性辞書に SentiWordNet，Lexicoder Sentiment Dictionary を用いた場合である．

表 1　極性判定モジュールの組み合わせによる精度比較

極性判定モジュール	accracy
Random	0.3333
RN101only	**0.5652**
Caption&Dict [4]	0.3246
Caption&Dict [5]	0.2904
RN101+C&D[4](proposed)	**0.4958**
RN101+C&D[5]	0.4583

5　考察

実験 1 の結果から，postive の適合率は negative と比べて低く，negative と判断すべき画像も postive と判断してしまう傾向があることが分かった．

実験 2 の結果から，キャプションからの感情極性を用いる提案手法の結果は，resnet101 を単体で使用する手法，すなわち画像のみを用いる場合よりも精度が低下した．その要因は実験 1 で明らかになった，キャプション生成部の positive と negative での感情判定精度の偏りが原因だと考えられる．

また，極性評価辞書を SentiWordNet[4]から Lexicoder Sentiment Dictionary(LSD)[5]に変更した際の極性辞書の変化によって，Caption&Dict[4]から Caption&Dict[5]へと，RN101+C&D[4]から RN101+C&D[5 へとは]同程度の値の推移を見せている．この結果から，キャプションの極性が全体の判定結果に寄与していることが確認された．

先行研究[2]では，画像からの感情分析精度 50%前後，付随文章からの感情分析精度90%前後で，両者を併用して画像の感情分析を行った際に，90%よりも高い精度が出ている．これは，画像と文章を同時に入力として感情分析する場合，文章の分析精度の方が画像のそれよりも全体の分析精度への関与が大きいことを示唆していると解釈でき，今回の実験結果もこれと傾向が一致することが確かめられた．

6　まとめ

提案手法による精度の向上は見られなかった．しかし，キャプション生成モジュールでの positive と negative の推定精度の偏りや，画像と文章のスコアを統合する際の寄与の偏りをなくすことで，改善できると思われる．今後は，より高精度な極性辞書の利用や，画像とキャプションのスコアの統合手法の改善が課題である．

今回，neutral と判断されてしまうような，感情極性分析に寄与しづらい文章も多く生成されてしまっていたため，精度向上のためにも，感情表現を多く含む文章を生成するための工夫を加えていく予定である。

参考文献

[1]桂井 麻里衣，佐藤 真一，“画像・テキスト・感情語の潜在的な相関に基づく画像の感情分類，”日本データベース学会論文誌， vol. 15-J, No. 10, 2017.

[2] V. Lopes, A. Gaspar, L. A. Alexandre, J. Cordeiro, “An AutoML-based Approach to Multimodal Image Sentiment Analysis,” International Joint Conference on Neural Networks(IJCNN), 2021.

[3] Z. Li, Q. Sun, Q. Guo, H. Wu, L. Deng, Q. Zhang, J. Zhang, H. Zhang, Y. Chen, “Visual sentiment analysis based on image caption and adjective–noun–pair description,” Springer Open Access Journal, Soft Computing, November 2021.

[4] S. Baccianella, A. Esuli, F.Sebastiani, “SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining,” Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), 2010.

[5] L. Young, S. Soroka, “Affective News: The Automated Coding of Sentiment in Political Texts,” *Political Communication*, vol. 29(2), 2012.

[6]https://data.world/crowdflower/image-sentiment-polarity