

Relational Graph Attention Networks を利用した感情認識の検討

Study of Emotion Recognition using Relational Graph Attention Networks

石渡 太智[†] 安田 有希[†] 宮崎 太郎[†] 後藤 淳[†]
Taichi Ishiwatari Yuki Yasuda Taro Miyazaki Jun Goto

1. はじめに

放送局における番組制作過程では、SNS 上で話題となっている出来事を利用することが多く見受けられる。話題の調査を目的に、ツイートに含まれる頻出単語をもとに、社会的に関心の高いキーワードを取得している。このように得られたキーワードには企業広告などが多く含まれ、番組制作に不要なものを提示してしまう課題がある。そこで、リツイート等の反響件数だけでなく、投稿が示す感情極性を利用し、番組制作に有用な投稿を取得する研究を進めている[1]。社会的関心が高いテーマは、感情と関わりが強い内容を多く含むため、話題を抽出する際の有効な指標と考えられている[2]。しかし、SNS 上の投稿は、一般的に感情を示すラベルが付与されない。したがって、投稿にどのような感情が現れているか推定する必要がある。そこで本稿では、投稿される発話文からその感情ラベルを推測するとともに、リプライ等の対話文(図 1)も利用することで高精度な感情ラベルの推定を行う。一般に、発話に現れる感情は、本人の発話内容だけでなく、他者の感情や発話にも影響を受けることが知られている[3]。したがって、他者との関係性も考慮することで、認識精度の向上を期待できる。

近年の対話文における感情認識技術の多くは、発話内容の理解に加えて発話間の関係性の取得に着目している。DialogueRNN[4]は、発話内容と発話間の関係、そして感情の移り変わりをそれぞれ GRU でモデリングした手法である。DialogueGCN[5]は、発話内容の抽出に GRU を用い、発話関係の抽出に 2 層の Graph Neural Networks (GNN) を用いた手法である。DialogueGCN は発話間の関係性取得に、初めて GNN を用いたことで、State-of-the-art を達成した。

本稿では、DialogueGCN モデルを参考にして、対話文における発話間の関係性を、GNN の 1 つである Relational Graph Attention Networks (RGAT) を用いてモデリングした高精度な感情認識技術を提案する。提案手法では、発話内容を抽出するネットワークと発話間の関係性を抽出するネットワークを組み合わせることで、認識精度の向上を図る。感情認識分野で広く用いられている IEMOCAP[6] ベンチマークデータセットを用いて提案手法の有効性を検証したところ、Weighted F1 score で 65.09% となり、従来の State-of-the-art 手法である DialogueGCN の認識精度 64.18% よりも精度を向上させることができた。

2. 提案手法

図 2 に提案手法の概要を示す。提案手法は発話内容抽出部と発話関係抽出部、認識部からなる。入力是对話文であり、出力は各ラベルの確率値である。出力された一番高い確率値を予測ラベルとして判定する。

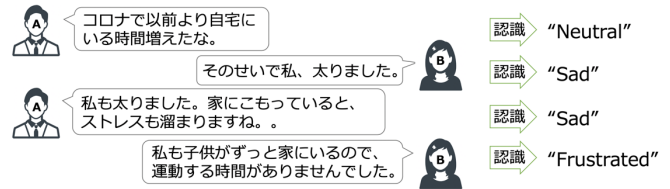


図 1 SNS 上の対話文における感情認識

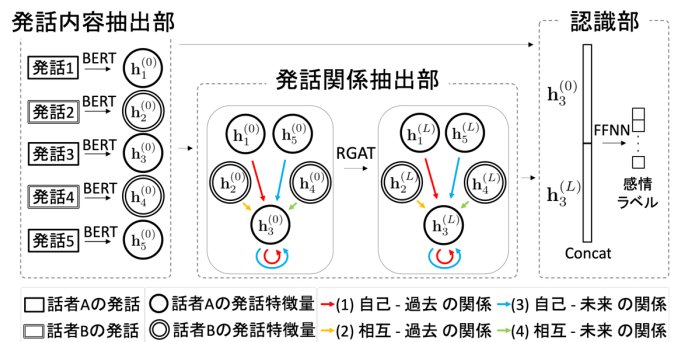


図 2 提案手法の概要

2.1 発話内容抽出部

発話内容抽出部では発話ごとに内容に応じた特徴量を生成する。発話内容の抽出には、近年数々の NLP タスクで著しい成功をおさめた事前学習モデルの BERT[7] を用いる。初めに、各発話を BPE tokenizer を用いて token ごとに分割する。次に、各 token の特徴量を入力とし、事前学習した uncased BERT-Base[‡] モデルを用いて、発話内容を加味した特徴量に修正する。最後に、Max pooling を用いて、token ごとの特徴量から各発話の特徴量を得る。得られた発話特徴量は、発話全体の意味を示すベクトルとなる。発話 i の発話特徴量を $h_i^{(0)}$ とする。BERT モデルのパラメータは学習時に再学習(ファインチューニング)する。

2.2 発話関係抽出部

発話関係抽出部では、発話間の関係性を考慮する。対話文における人の感情は、他の発話に依存する。なかでも自己依存と相互依存に影響を受けることが報告されている[3]。自己依存と相互依存の重要性を図 1 の例を用いて説明する。図 1 は“コロナウィルス”という社会的関心の高いテーマについて、SNS 上で投稿された対話文である。女性は一連の会話を通じて、他者の発言に関係なく常に negative な感情を抱いている。このように、自身の状態が自身の感情に影響を及ぼす、自己依存の重要性がわかる。一方で、男性は 3 番目の発話において、直前の女性の negative な発言に共感し、自身の感情も negative に転じている。このように、他者の状態が自身の感情に影響

[†] NHK 放送技術研究所

[‡] <https://github.com/google-research/bert>

響を及ぼす、相互依存の重要性がわかる。また、自身の発言よりも前に発言されたか、後に発言されたかを示す時系列情報も、対話文における人の感情を決定する要因となる。まとめると、発話関係と時系列情報の両方を考慮し、(1) 自己-過去 (2) 相互-過去 (3) 自己-未来 (4) 相互-未来 の計 4 つに依存関係は区別される。

これらの関係性を取得するために、本稿では RGAT を用いる。RGAT は RGCN [8] を基にした GNN の 1 つで、関係ごとに別々のネットワークを形成するため、発話間の関係性を別々にモデリングすることができる。発話関係抽出部では、発話特徴量を入力とし、RGAT を用いて、発話間の関係性を考慮した特徴量に修正する。L 回修正して得られた関係特徴量を $h_i^{(L)}$ とする。

2.3 認識部

認識部では、得られた発話特徴量 $h_i^{(0)}$ と関係特徴量 $h_i^{(L)}$ を連結し、2 層の Feed Forward Neural Networks (FFNN) を用いて各感情ラベルの確率値を得る。

3. 評価実験

提案手法の有効性を示すために、1 対 1 の会話をテキスト化したベンチマークデータセット IEMOCAP[6]を用いて、従来手法との比較、および発話関係の効果を評価する実験を実施した。IEMOCAP は、それぞれの発話に計 6 ラベル (“happy”, “sad”, “neutral”, “angry”, “excited”, “frustrated”) のいずれかが付与されており、120 対話文を含む train セットと 31 対話文を含む test セットで構成されている。実験では、train セットの内 10% を検証用、残りを訓練用として用いた。また、評価指標として weighted F1 score を用い、すべての実験は 5 回行い、平均値を実験結果として使用した。

実験では、全ネットワークを、交差エントロピー誤差を最小化するよう Adam 法により最適化した。BERT と RGAT の学習率はそれぞれ $4e-5$ と $2e-3$ に設定し、バッチサイズは 1 とした。発話抽出部と発話関係抽出部の特徴量は 768 次元で、FFNN の中間層は 384 次元に設定した。RGAT のレイヤ数は 1, 2, 3 のいずれかを選択し、FFNN の連結部は合計または連結を選択し、検証データで各種ハイパーパラメータを決定した。

従来手法との比較

提案手法の性能評価のため、従来手法である BERT[7] と DialogueRNN[4], DialogueGCN[5] との比較を行った。BERT は文脈情報抽出部のみを用いた手法である。DialogueGCN は 2 層のグラフ構造の内、1 層目に RGAT、2 層目に RGCN を用いているため、全層 RGAT を用いる提案手法とグラフ構造の構成が異なる。また DialogueGCN は、8 つの依存関係*を考慮するため、依存関係の種類も異なる。

評価実験の結果を表 1 に示す。提案手法の精度は 65.09% となり、全手法の中で最も高い精度となった。文脈情報抽出部に BERT を使用したこと、発話関係抽出部に従来手法と異なるモデルを用いたことが認識精度向上に繋がったと考えられる。

* 話者数に応じて、4 つの関係を考慮する。したがって、話者が 2 人の場合は $2 \times 4 = 8$ 通りの依存関係を持つ。

	Weighted average F1 score
BERT [7]	53.31
DialogueRNN [4]	62.75
DialogueGCN [5]	64.18
提案手法	65.09

表 1 他手法との精度比較

自己-過去	相互-過去	自己-未来	相互-未来	Score
-	○	○	○	63.52
○	-	○	○	62.60
○	○	-	○	63.51
○	○	○	-	64.80

表 2 発話関係の効果

発話関係の効果

発話関係の効果を検証するため、発話関係抽出部における(1) 自己-過去 (2) 相互-過去 (3) 自己-未来 (4) 相互-未来 の計 4 つの関係の内、それぞれ 1 つを使用せず実験を行った。表 2 に実験結果を示す。“○”は使用する関係を示し、“-”は使用しない関係を示す。“Score”は weighted average F1 score を示す。実験の結果、(2) 相互-過去の関係を除き、残り 3 つを用いた場合、認識精度は 62.60% となり、4 つの関係の中で最も低い精度となった。これより、過去に発言された他者の発話が感情認識に最も影響を与えていることが分かった。

4. おわりに

本稿では、Relational Graph Attention Networks を用いた感情認識手法を提案した。提案手法では、発話内容を抽出するネットワークと発話間の関係性を抽出するネットワークを組み合わせ、認識精度の向上を図った。IEMOCAP データセットを用いた実験では、従来手法の認識精度よりも高い精度を達成できた。今後は、依存関係の数を調整し、対話文における最適な依存関係の種類を決定したい。

参考文献

- [1] 武井友香, 宮崎太郎, 後藤淳, “感情極性に着目したトレンド情報抽出手法の検討”, 信学技報, vol. 119, no. 98, NLC2019-4, pp. 23-28 (2019).
- [2] Stieglitz, Stefan, and Linh Dang-Xuan. "Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior." *Journal of management information systems* 29.4, 217-248 (2013).
- [3] Poria, Soujanya, et al. "Emotion recognition in conversation: Research challenges, datasets, and recent advances." *IEEE Access* 7, 100943-100953 (2019).
- [4] Majumder, Navonil, et al. "Dialoguernn: An attentive rnn for emotion detection in conversations." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. (2019).
- [5] Ghosal, Deepanway, et al. "Dialoguegcnn: A graph convolutional neural network for emotion recognition in conversation." *arXiv preprint arXiv:1908.11540* (2019).
- [6] Busso, Carlos, et al. "IEMOCAP: Interactive emotional dyadic motion capture database." *Language resources and evaluation* 42.4, 335 (2008).
- [7] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
- [8] Schlichtkrull, Michael, et al. "Modeling relational data with graph convolutional networks." *European Semantic Web Conference*. Springer, Cham (2018).