

感情推定における感情カテゴリに関する先験的知識の利用

田辺 ひかり

小川 哲司

小林 哲則

林 良彦

早稲田大学理工学術院

tanabe@pcl.cs.waseda.ac.jp

1 はじめに

物語テキストからの感情推定はテキスト理解の重要な要素であるが、登場人物の感情は陽には表現されないことが多く、商品レビュー分類などとは異なる難しさがある。また、一般に複数の感情ラベルが付与されるマルチラベル分類問題となることから、これに伴う課題も解決する必要がある。本論文では、物語テキストにおける登場人物の感情を推定する問題において、背景にある感情カテゴリ体系に関する先験的知識を利用することの有効性について議論する。より具体的には、Plutchikの8感情カテゴリ [4] によるアノテーションが付与されたデータセット [5] を用い、(1) Plutchikの8感情の環の構造、(2) 感情極性と同時学習を分類における損失関数に組み込み、これらを実験的に比較した。その結果、感情極性を組み込んだモデルがより良い精度を達成した。また、文のエンコーディングにBERTを用いることで、従来研究を大幅に上回るベースライン精度が得られることも確認できた。

2 関連研究

2.1 テキストからの感情分析 [6]

テキストからの感情分析は、その応用の幅広さから古くから研究されてきた。感情極性を分類する二値分類タスクが代表的であるが、観点 (aspect) ごとの極性を分類する派生的なタスクも生まれている。手法としては、感情辞書などの言語資源に依存する方法からニューラルネットを用いた学習ベースの手法に主流が移ってきている。感情分析の対象としては、商品レビューに代表される感情を陽に表現した文章が主であったが、最近では物語テキストを対象とする試みが注目されている。物語テキストでは感情が必ずしも陽に表現されないため、いわゆる行間を読む機能が必要となることが多く、人間と機械の推定精度には依然ギャップがある。



図 1: Plutchik の 8 感情 ([5] より転載)

また、テキストに込められた感情は一つとは限らないため、感情推定の問題は一般にはマルチラベル分類問題となる。感情ラベルの基本となる感情カテゴリ体系は心理学的理論に基づくものが多く、Ekmanの基本6感情やPlutchikの8感情、Russellの感情円環モデルなどが多く用いられている。本研究で扱うStory commonsense dataset [5] では、Plutchikの8感情によるラベル付けがされている。

2.2 Plutchikの環における先験的知識

図1に示すように、このPlutchikの感情カテゴリ体系では、8つの感情が円環上に位置している。対極にある感情は反対の性質を持ち、隣接する感情は何らかの意味で類似しているという考えを反映している。これを先験的知識として感情推定タスクに利用した研究として [2, 3] がある。Deyuら [2] は、emotion distribution learningにおいて、Plutchikの環の角度の大小から感情カテゴリ対の重みを先見的に定め、学習時のロスに組み込んだ。He [3] らはDeyuらの方針をマルチラベル文書分類タスクに適用した。Heらの損失関数を以下に示す。

$$L = - \sum_{c=1}^C (y_c \log p_c + (1 - y_c) \log(1 - p_c)) + \lambda_1 \sum_{s,t} w_{s,t} (p_s - p_t)^2 + \lambda_2 \|\theta\|^2 \quad (1)$$

第1項で表される各カテゴリに対するバイナリクロスエントロピー (BCE) に加え, 第2項に感情 s, t のペアに対する間違いのペナルティー重み $w_{s,t}$ を導入している. また, 第3項で学習パラメータ θ を用いた L_2 正則化を行う. ここで $w_{s,t}$ は感情 s, t の円環上の相対位置によって以下のように事前に固定されている.

- -1: s, t が対極の位置にある場合
- 0.5: s, t が隣り合っている場合
- 0: s, t の成す角が 90° の場合
- -0.5: s, t の成す角が 135° の場合

3 感情推定モデル

3.1 ベースラインモデル

本研究では事前学習済みの BERT [1] を利用する. BERT の最終層から "[CLS]" の 1024 次元の分散表現を取得し, 全結合層に入力することで 8 カテゴリそれぞれに対応した logits を得る. BCE を用いた本モデルの損失関数を以下に示す. C はカテゴリ数, y_c, p_c はそれぞれ c 番目のカテゴリの正解ラベル, logits を表す.

$$L = - \sum_{c=1}^C ([y_c \log p_c + (1 - y_c) \log(1 - p_c)]) \quad (2)$$

カテゴリ間の関係性を仮定しないこのベースラインモデルを以下では **BL** と呼ぶ.

3.2 カテゴリ間の関係性を用いるモデル

マルチラベル分類においては, カテゴリ間の潜在的な依存関係を適切に利用することが望まれる. 本研究では, BL に加え, 以下のカテゴリ間の関係性を損失関数に取り入れた推定モデルを比較する.

3.2.1 データ分布の利用

図2に示すようにカテゴリごとに共起しやすい/しづらいという偏りがある. そこで, カテゴリ間の距離は学習データにおけるカテゴリ間の共起率であると仮定し, 式(1)の重み $w_{s,t}$ を設定する. 共起率は(感情 s が 1 である場合に感情 t も 1 である回数) ÷ (感情 s が 1 である回数) で求める. 以下, 本モデルを **data_proportion** と呼ぶ.

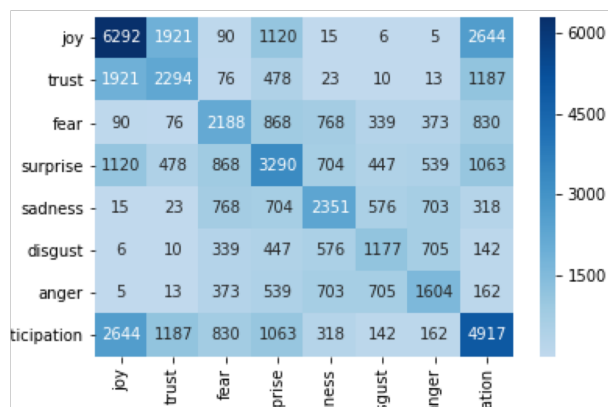


図2: ラベル間の共起回数

3.2.2 Plutchik の環の利用

2.2 節で述べた He らの損失関数を用いるモデルを **pl_He** と呼ぶ. He らはカテゴリ間のペナルティー重み $w_{s,t}$ を予め決めていたが, Plutchik の環において感情は等間隔に円環状に配置されていると仮定し, カテゴリ間の角度を 45° として式(1)の重み $w_{s,t}$ を設定するモデル (**pl_cos**) を提案する. より具体的には, n を整数としてカテゴリ s, t 間角度が $45n^\circ$ のとき, $w_{s,t} = \cos(45n^\circ)$ とする.

3.2.3 感情極性の利用

Plutchik の 8 感情は, ポジティブなカテゴリ (*joy, trust*) とネガティブなカテゴリ (*fear, sadness, disgust, anger*) に大別でき, 図2に示すように, 同じ極性のカテゴリ同士の共起が多い. そこで, 感情極性の二値分類と 8 感情の分類でマルチタスク学習を行うモデルを提案する. 以下に学習に用いた損失関数を示す. 第1項で 8 カテゴリ分類, 第2,3 項目で二値分類の損失を計算し, 足し合わせたものを最終的な損失とする.

$$L = - \left[\sum_{c=1}^C ([y_c \log p_c + (1 - y_c) \log(1 - p_c)]) + (\{y^+ \log p^+ + (1 - y^+ \log(1 - p^+))\}) + (\{y^- \log p^- + (1 - y^- \log(1 - p^-))\}) \right] \quad (3)$$

なお *anticipation, surprise* については多くのカテゴリと共起しやすいことが見受けられたため, ポジティブでもネガティブでもないとした. 実験においては, この損失関数を 2 種類のニューラルモデルにより実装した. 1 つは BL モデルと同じ構造で式(3)を使用したもの (**mt_same_layer**) であり, もう 1 つは BERT

表 1: 実験に使用した各感情ラベルデータの総数

	joy	trust	fear	surprise	sadness	disgust	anger	anticipation	total
train	6292	2294	2188	3290	2351	1177	1604	4917	13004
tecest	5829	1792	1636	2537	2104	944	1356	4196	11859

表 2: 各モデルによる Plutchik ラベル分類精度. * は Rashkin ら [5] の実験結果を引用.

Model	P	R	F1
BiLSTM*	20.27	30.37	24.31
CNN*	21.15	23.36	22.20
REN*	26.22	33.26	29.32
NPN*	21.98	37.31	27.66
BL	58.4	58.8	58.6

の最終層を 8 感情分類と感情極性の二値分類で分けたもの (`mt_diff_layer`) である.

4 実験

4.1 データセットと正解ラベルの設定

短いテキストにおける登場人物の心的状態 (基本的欲求および感情状態) の推定・理解を目的として作成された Story commonsense dataset [5] を利用した. Story Cloze test dataset からランダムに選択された, 5 文からなる物語文に対して, クラウドソーシングによるラベル付けが行われており, 感情状態のラベルは, Plutchik [4] の 8 感情を用いて下記の例のように, 一般に複数が付与されている. ここで, 1 から 3 の数字は各カテゴリの強さを表す感情スコアである.

Danielle was really short on money.

emotions=["fear:3", "disgust:2", "anticipation:2"]

本研究では各データごとに感情スコアの平均値を求め, 平均スコアが 2 以上のカテゴリを正例とした. なお, 全感情ラベルを感情スコア 2 または 3 としたデータは spam とみなし削除した.

4.2 実験設定

公開されているデータセットのうち感情ラベルが付与されているデータが dev, test セットであったため, それぞれを訓練, テストデータとして利用した. 表 1 に使用したデータの総数を示す. 各感情推定モデルには登場人物と物語文 1 文を '#' で連結し入力した. また評価指標には micro-averaged Precision(P), Recall(R), F1 値を使用した.

4.3 実験結果

ベースラインモデルと先行研究との比較: 表 2 に本研究で用いたモデルと先行研究による感情推定の精度を示す. BERT を単純に用いたベースラインモデル BL が先行研究を大幅に上回る精度を達成した. 先行研究では表 2 の Model の列に示す文エンコーダの後段にロジスティック回帰を組み合わせたモデルにより推定を行っている. なお表中で, REN は Recurrent Entity Network (EntNet), NPN は Neural Process Network を指す. 損失関数は本研究モデルと同様に BCE を使用しているが, カテゴリ別の重みをかけ合わせている.

カテゴリ間の関係性を用いるモデルの比較: 表 3 にラベル間の関係性を組み込んだモデルによる実験結果を示す. 数値は順に Precision/Recall/F1 を示しており, 同感情のうち最高の値を評価指標ごとに下二重線, 下線, 太字で示してある. 総合指標である F1 値に着目すると, `mt_diff_layer` が僅差で最高値であり, 次いで BL, `mt_same_layer`, `data_proportion` であった. カテゴリ別の推定精度に注目すると, カテゴリ間の関係性をモデルに組み込むことで *disgust*, *anger* で多くのカテゴリの Precision, Recall が向上したことがトータルの精度向上の一要因であることがわかる. また, `mt_same_layer`, `data_proportion`, `mt_diff_layer` では, それぞれ 6 カテゴリで BL を上回る Precision を達成した. これらの精度向上は共起しづらいカテゴリ間での推定間違いが減ったためと考えられる. BL とその他のモデルの Recall を比較してみると *disgust* においては全てのモデルが BL を上回っているもの多くのカテゴリで BL が他モデルよりも高い値となった. `data_proportion` モデルについては, *surprise*, *sadness*, *disgust*, *anger* カテゴリで BL を上回る値を達成した. これらのカテゴリは互いに共起しやすいカテゴリであったことから, 共起のしやすさを損失関数に反映したことで正解ラベルの取りこぼしが減少したと考えられる. `pl_He` と `pl_cos` を比較すると, 総合的な精度には大きな差が出なかったことから, Plutchik の環のモデル化方法について, `pl_he` と `pl_cos` で明確な優劣はつけられないことがわかった.

表 3: 損失関数の異なる各モデルのカテゴリ別 mirco-averaged Precision/Recall/F1.

model	total	joy	trust	anticipation	surprise
BL	58.4/58.8/58.6	81.5/83.2/82.3	32.5/32.6/32.6	55.5/58.8/57.1	43.3/47.1/45.1
data_proportion	58.9/58.1/58.5	81.8/82.5/82.1	33.0/28.8/30.7	56.6/57.8/57.2	44.8/47.2/46.0
pl_He	58.4/58.1/58.2	81.5/82.9/82.2	32.2/29.6/30.8	56.0/57.7/56.8	44.3/45.0/44.6
pl_cos	58.5/57.9/58.2	81.3/83.2/82.2	31.2/30.2/30.7	56.1/57.3/56.7	44.3/43.8/44.1
mt_same_layer	59.0/58.0/58.5	81.3/83.6/82.4	32.7/28.5/30.4	56.0/57.8/56.9	43.9/46.9/45.4
mt_diff_layer	59.4/57.9/58.7	82.0/82.6/82.3	31.8/28.3/30.0	55.3/58.1/56.7	45.6/46.5/46.0

model	fear	sadness	disgust	anger
BL	49.6/52.3/50.9	61.6/58.3/59.9	35.4/23.5/28.3	48.2/43.1/45.5
data_proportion	47.8/50.1/48.9	59.8/58.4/59.1	36.9/27.3/31.4	48.3/44.5/46.3
pl_He	48.2/53.2/50.6	59.5/57.0/58.2	33.7/25.2/28.8	48.4/44.8/46.5
pl_cos	49.3/52.2/50.7	59.4/57.8/58.6	34.6/25.2/29.2	48.9/43.4/46.0
mt_same_layer	47.8/51.5/49.6	62.4/55.5/58.7	37.9/26.0/30.8	49.2/41.9/45.3
mt_diff_layer	50.2/50.7/50.5	61.7/58.7/60.2	37.0/24.4/29.4	50.4/42.9/46.4

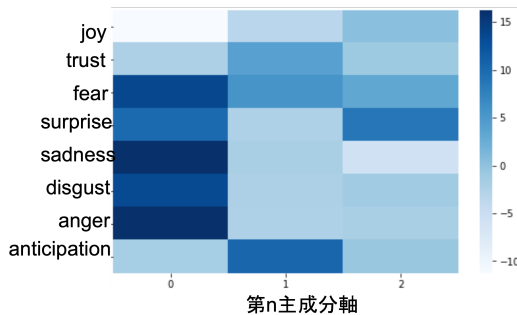


図 3: 主成分分析により 3 次元に圧縮した各感情カテゴリの分散表現の平均.

4.4 再考: Plutchik の環

図 3 に BL で学習した後の “[CLS]” の分散表現を主成分分析により 3 次元まで圧縮した結果を示す。主成分分析の対象は、簡単のため正解ラベルが単一のもののみを抽出し、カテゴリごとに平均をとった分散表現とした。横軸の主成分軸ごとに各感情を見ると、第 0 軸がその極性を分ける軸となっており、*anticipation* はポジティブ、*surprise* はネガティブ側に分かれている。このように恣意的に先験的知識を損失関数に組み込まない BL モデルにおいても第 0 主成分が感情極性に対応することから、感情極性は多カテゴリの感情分類においても重要な要因であることがわかる。これは、実験において極性の二値分類とマルチタスク学習を行ったモデルがより高い精度を示したことの裏付けともなりうる。また第 2 軸では、ネガティブな感情の中でも *sadness* と (*surprise*, *fear*) が対極関係にあることから、後者の 2 感情はネガティブな感情のなかでも性質が異なると考えられる。

これらの結果から、本研究によって推測される

Plutchik の環の構造は、図 1 に示される均一な構造とは異なり、感情極性によりポジティブ、ネガティブの 2 つのグループに大別され、ネガティブのグループ内では (*sadness*, *disgust*, *anger*) と (*surprise*, *fear*) のグループに分かれる形であると考えられる。

5 おわりに

物語文から登場人物の感情推定を行うマルチラベル分類タスクにおいて、BERT による単純なベースラインモデルが高い精度を示すこと、この上に組み込む感情カテゴリの先験的知識については、Plutchik の 8 感情の構造をモデル化したものよりも感情極性分類とのマルチタスク学習を行うモデルのほうが優れた性能を示すことを実験的に確認した。今後は、ある感情状態を引き起こす行動との因果関係分析へと展開し、物語理解の基盤技術として確立を目指したい。

参考文献

- [1] Jacob Devlin, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL 2019*, pp. 4171–4186, 2019.
- [2] Zhou Deyu, et al. Emotion distribution learning from texts. In *EMNLP 2016*, pp. 638–647, 2016.
- [3] Huihui He, et al. Joint binary neural network for multi-label learning with applications to emotion classification. In *NLPCC 2018*, pp. 250–259, 2018.
- [4] Robert Plutchik. A general psychoevolutionary theory of emotion. In *Theories of emotion*, pp. 3–33. Elsevier, 1980.
- [5] Hannah Rashkin, et al. Modeling naive psychology of characters in simple commonsense stories. In *ACL 2018*, pp. 2289–2299, 2018.
- [6] Lei Zhang, et al. Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, Vol. 8, No. 4, p. e1253, 2018.