

表出感情と経験感情をタグ付けした対話コーパスの構築

井手竜也 河原大輔
早稲田大学理工学術院
{t-ide@toki.,dkw@}waseda.jp

概要

人間は、対話において相手の感情を認識し、共感や慰めなどを込めて応答する。本論文では、このような能力をもつ対話システムの開発に向けて、感情タグ付き対話コーパスを構築する。ある発話を話した人が表した感情（表出感情）と、それを聞いた人が感じた感情（経験感情）を、クラウドソーシングによってタグ付けする。日本語の感情タグ付き対話コーパスを構築し、それを統計的に分析した結果、対話における表出・経験感情の特徴および違いが明らかになった。また表出・経験感情の認識実験を行い、経験感情の認識がより難しいことや、2種類の感情認識が互いの性能を向上させることを示した。

1 はじめに

社会のオンライン化に伴って、テキストベースのコミュニケーションは不可欠なものとなっている。Transformer [1] の登場や転移学習 [2, 3] の成功を契機に、対話システムの性能も急速に向上している。最近では、人間らしい応答ができるオープンドメインのチャットボットを目指した研究 [4, 5, 6] もある。

人間は相手をもつ感情を認識し、それにふさわしい感情（共感や慰めなど）を込めて応答を返すと考えられる。対話システムを人間らしくするための鍵として、相手の感情をふまえて応答を生成すること [7] が挙げられる。感情タグ付き対話コーパスはいくつか存在するが、発話ごとに話者の感情のみをタグ付けしているもの [8, 9] が多く、不十分である。

本研究では、相手の感情を認識した上で適切な感情を込めた応答を生成できる対話システムを実現するための感情タグ付き対話コーパス¹⁾を構築する。話し手が発話に込めた感情（表出感情）と発話の聞き手が受けた感情（経験感情）を発話ごとにタグ付けする。なお本研究ではこれを日本語で構築する

1) <https://github.com/nlp-waseda/expr-exper-emo>にて公開する予定である。

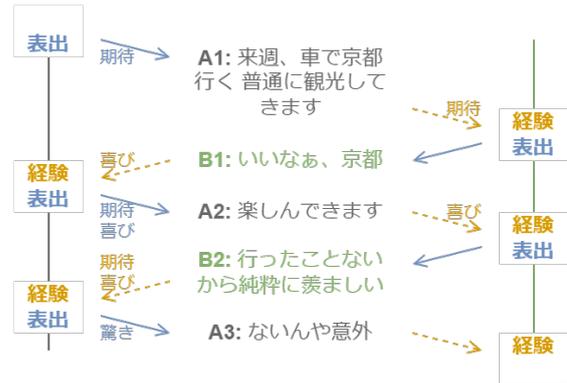


図 1: 表出感情と経験感情のタグ付け

が、手法はどの言語にも適用できる。

手法としては、Twitter から 2 人が交互に話すような対話を収集し、対話コーパスとする。Plutchik の感情の輪 [10] をラベルに採用し、クラウドソーシングによって各発話における表出感情と経験感情をタグ付けする。複数のラベルがタグ付けされることを許し、投票数に応じて各ラベルに強弱を設ける。

上記の手法によって、3,828 対話 (13,806 発話) からなる対話コーパスを構築した。統計的な分析から、表出感情と経験感情の関係が明らかになった。さらに表出感情と経験感情を BERT [3] で認識する実験を行った。各ラベルの強度を回帰させ、相関係数を評価指標として性能を評価した。実験の結果から、表出感情よりも経験感情を認識する方が難しいことや、それらのマルチタスク学習が認識の性能を向上させることがわかった。本研究で構築したコーパスによって、対話における感情の認識および感情に基づく応答生成の研究が促進されると期待する。

2 関連研究

EmoBank [11] はある文を書いた人とそれを読んだ人の感情を、クラウドソーシングによってタグ付けしたコーパスである。WRIME [12] は EmoBank と同様、主観的と客観的な感情をタグ付けした日本語のコーパスである。EmoInt [13] は怒り・恐れ・喜

表 1: 表出感情と経験感情がタグ付けされた対話の例

発話	表出		経験	
	強	弱	強	弱
A1: 来週、車で京都行く 普通に観光してきます	{期待}	{期待, 喜び}	{期待}	{期待}
B1: いいなあ、京都	{}	{期待}	{喜び}	{期待, 喜び}
A2: 楽しんできます	{期待, 喜び}	{期待, 喜び}	{喜び}	{期待, 喜び}
B2: 行ったことないから純粋に羨ましい	{}	{期待}	{期待, 喜び}	{期待, 喜び}
A3: ないんや意外	{驚き}	{驚き}	{}	{喜び, 驚き}

び・悲しみの4感情について、その強度をタグ付けしたコーパスである。StoryCommonsense [14] は短い物語をなす一連の文に、登場人物の Motivation と Emotional Reaction をタグ付けしたコーパスである。しかし、上記のコーパスはどれも対話に関するものではない。StoryCommonsense は本研究に似ているが、対象は物語である。

EmpatheticDialogues [7] は2人の話者を Speaker と Listener に割り当て、対話ごとに Speaker の感情とその状況をタグ付けした対話コーパスであり、日本語版 [15] も存在する。EmpatheticDialogues は対話ごとに感情をタグ付けしているため、発話ごとの感情を理解することには向かない。DailyDialog [8] は発話ごとに感情と意図をタグ付けした対話コーパスである。EmotionLines [9] は発話ごとに感情をタグ付けした対話コーパスで、話者が複数人である。本研究では DailyDialog や EmotionLines と同じく、発話ごとに感情をタグ付けする。これらの対話コーパスは発話に込められた感情のみをタグ付けしているが、本研究ではその発話を聞いた相手が抱く感情もタグ付けする。さらに本研究では複数の感情を許容し、それぞれの感情に強弱を設ける。

3 コーパスの構築

3.1 対話の収集

対話は Twitter API によって収集する。2人のユーザによるツイートとリプライのかけあいを対話とみなし、それを抽出する。ただしハッシュタグや URL、画像を含む対話はすべて除外する。対話あたりの発話数は2から9までに絞り、いくつかのフィルタを施す。まず特殊な記号や絵文字を含む対話はすべて除外する。ほかにも4回以上繰り返される文字や単語を含む場合や、発話が4文字未満の場合も除外する。

上記の手法によって、3,828 対話 (13,806 発話) を獲得した。ただしターン数が多い対話ほど、そのサンプル数は減る傾向にある。

表 2: ラベルごとの発話数

ラベル	表出		経験	
	強	弱	強	弱
怒り	430	1,349	124	870
期待	1,906	4,229	1,215	4,068
喜び	1,629	3,672	1,553	4,549
信頼	247	1,732	520	3,455
恐れ	252	942	123	846
驚き	602	2,018	434	2,798
悲しみ	1,227	2,936	889	3,037
嫌悪	476	1,979	186	1,535
どれか	6,371	12,215	4,705	12,515

3.2 表出感情と経験感情のタグ付け

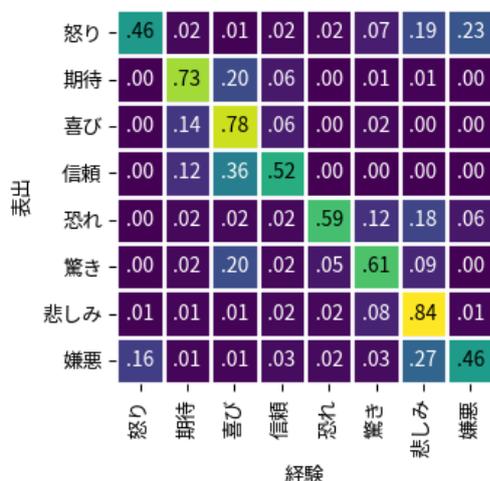
本研究では Plutchik の感情の輪 [10] をラベルに採用する。²⁾具体的には{怒り, 期待, 喜び, 信頼, 恐れ, 驚き, 悲しみ, 嫌悪}の8感情である。発話ごとにそれを話した人が抱いていた感情 (表出感情) とそれを聞いた人が抱いた感情 (経験感情) をタグ付けする。つまり発話ごとに主観的な感情と客観的な感情 [11, 12] をタグ付けすることになる。表出感情と経験感情を分けてタグ付けすることで、発話間や話者間における感情の変化を捉える。

感情のタグ付けはクラウドソーシングによって行う。プラットフォームとしては Yahoo!クラウドソーシング³⁾を用い、クラウドワーカーに発話とその履歴を与え、各感情の有無を問う。複数の感情を選択することや、どの感情も選択しないことを許容する。発話ごとに7人のクラウドワーカーを雇い、半数 (4人) 以上が選んだ感情を強ラベル、4分の1 (2人) 以上が選んだ感情を弱ラベルとする。このタグ付けを表出感情と経験感情のそれぞれについて独立に行う。タグ付けされた対話の例を表 1 に示す。

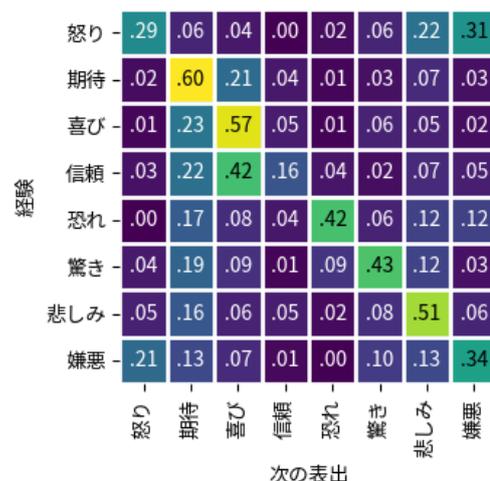
ラベルごとの発話数を表 2 に示す。およそ9割の発話がいずれかの弱ラベルを伴っており、その他の

2) 感情タグ付きコーパスでは、Ekman の6感情 [16] や Plutchik の感情の輪 [10] がよく用いられる。クラウドソーシングを用いた予備実験の結果、後者の方がより感情タグとして適切であることが明らかになった。したがって本研究ではそれを採用する。

3) <https://crowdsourcing.yahoo.co.jp/>



(a) 表出感情と経験感情（特定の発話について）



(b) 経験感情と次の表出感情（特定の話者について）

図 2: 表出感情と経験感情の関係

感情や無感情のラベルが大部分を占める DailyDialog [8] や EmotionLines [9] よりも有効である。

4 コーパスの分析

構築した対話コーパスにおける，表出感情と経験感情の関係を分析した．具体的には，次のような関係が考えられる。

1. 同一の発話に対する表出感情と経験感情（異なる話者）
2. ある発話に対する経験感情と次の発話に対する表出感情（同一の話者）

強ラベルに関する，それぞれの混同行列を図 2 に示す．なお各要素の値は行方向に正規化してある．対角成分における値が大きくなっていることから，同一の発話および同一の話者についての関係としては，同じ感情が起りやすいことが言える．異なる感情間の関係としては，図 2a から，期待・信頼・驚きの発話に対して喜び，怒りと嫌悪にはそれぞれ嫌悪と悲しみを抱きやすいことがわかる．また図 2b から，信頼と怒りを抱いたあとはそれぞれ喜びと嫌悪を表しやすいことがわかる．図 2a と図 2b を比較すると，特に悲しみに対する反応が異なり，特定の話者においては話者が悲しみを受けたとき次に期待を表すことがある．これは相手の発話に対して悲しみを抱いた話者が，相手を慰めるために期待感のある発話を返していると考えられる。

5 感情認識の実験

表 3: モデルごとの相関係数

モデル	表出		経験	
	Pearson	Spearman	Pearson	Spearman
京大 (BASE)	58.84	44.33	53.60	41.84
京大 (LARGE)	60.85	45.16	55.09	42.94
NICT	61.50	46.05	56.23	43.88

表 4: ラベルごとの相関係数 (NICT)

ラベル	表出		経験	
	Pearson	Spearman	Pearson	Spearman
怒り	50.21	33.80	38.11	23.80
期待	62.76	55.55	57.46	51.22
喜び	67.25	55.22	61.92	54.47
信頼	41.15	36.69	43.91	40.48
恐れ	59.09	31.47	49.60	24.90
驚き	49.86	39.58	40.58	33.86
悲しみ	63.70	51.50	55.48	43.88
嫌悪	47.76	38.18	37.32	28.13

5.1 モデル

表出感情と経験感情について，それらの認識を実験する．対象の発話とそれまでの履歴から，各ラベルの強度を回帰することを考える．ラベルがない場合に 0，弱ラベルと強ラベルにそれぞれ 1 と 2 を充てる．モデルは平均二乗誤差で学習する．

本研究ではモデルとして BERT [3] を利用する．事前学習済み BERT として，京都大学 [17] の WWM 版と NICT⁴⁾ の BPE 版を採用する．入力は Juman++ [18] によって単語に分割し，さらに BPE [19] によってサブワードに分割する．履歴と対象の発話を [SEP] で結合し，[CLS] に対するベクトルを全結合

4) <https://alaginrc.nict.go.jp/nict-bert/index.html>

表 5: 予測結果の例 (NICT, 最後の発話に対する表出感情)

対話	ラベル	
	予測	正解
A1: ゲームの検証してる人が検証してほしいことあれば言ってください的なこと 言ってたから依頼したら無視されて悲しくなったのはいい思い出 B1: それは悲しいね	強い悲しみ	強い悲しみ
A1: youtube でバーのマスターが氷砕いてる動画見てポーッとしてる B1: なんかしてよ, そのうちこういうときにツイキャスしようかなと思っておる A2: 天才の発想 スマホでも見やすいから助かる	弱い期待 弱い喜び	強い喜び 弱い信頼
A1: 今、部活終わって帰るところやけど雨やばいしかっぱ持ってきてないし最悪 B1: わたしも学校出た瞬間大雨降ってきた	強い驚き	強い悲しみ

表 6: 特定の発話に関するマルチタスク学習

訓練\テスト	表出		経験	
	Pearson	Spearman	Pearson	Spearman
表出	61.50	46.05	52.89	40.91
経験	55.49	43.34	56.23	43.88
マルチタスク	62.20	46.63	57.35	45.01

表 7: 特定の話者に関するマルチタスク学習

訓練\テスト	経験		次の表出	
	Pearson	Spearman	Pearson	Spearman
経験	54.62	43.47	29.53	25.46
次の表出	43.32	35.27	33.91	28.31
マルチタスク	55.75	49.50	35.17	30.49

層に与える。全結合層から得た 8 次元のベクトルに関して、各要素を 8 種類のラベルそれぞれの強度に回帰する。

対話コーパスを 8:1:1 に分割し、それぞれ学習データ・検証データ・テストデータに充てる。学習は 3 エポックと定め、テストデータに対して評価を行う。評価指標には Pearson と Spearman の相関係数を採用する。

5.2 結果

モデルごとの相関係数を表 3 に示す。モデルとしては、どの値も NICT がもっとも高い。表出感情と経験感情を比較すると、どのモデルも経験感情の値が低い。つまりある発話を話した人が抱いていた感情よりも、それを聞いた人が抱いた感情を予測することの方が難しい。NICT のモデルに関する、ラベルごとの相関係数を表 4 に示す。表 4 と表 2 を比較すると、ラベルの発話数が多いほど相関係数の値も高くなっていることがわかる。モデルによる予測結果の例を表 5 に示す。

5.3 マルチタスク学習

単一のモデルに表出感情と経験感情の両方を認識させることを考える。表出感情と経験感情には相関があるため、このことはそれぞれの認識に良い影響

があると考えられる。タスクごとに全結合層を用意し、それらを同時に訓練する。ロスそれぞれに対して計算し、その算術平均を全体のロスとしてパラメータを最適化する。

4 章の関係を踏まえ、特定の発話と特定の話者に関する表出感情と経験感情をマルチタスク学習 [20] する。また訓練データとテストデータが異なる場合も実験する。マルチタスク学習の相関係数を表 6 と表 7 に示す。まず訓練とテストでデータが異なる場合、それらが同じ場合よりも値が低い。これは表出感情と経験感情を分けてタグ付けすることに意味があることを示している。マルチタスク学習としては、表出感情と経験感情のどちらもそれらを同時に学習させたときの方が値は高い。つまり表出・経験・次の表出感情は互いに助け合うと言える。

6 おわりに

対話システムが話者の感情やその変化を理解するための、感情タグ付き対話コーパスの構築手法を提案した。クラウドソーシングによって、2 人の話者による対話の発話ごとに、その発話を話した人が抱いていた感情（表出感情）とそれを聞いた人が抱いた感情（経験感情）をタグ付けする。

上記の感情タグ付き対話コーパスを日本語で構築し、それを統計的に分析した。感情ごとに頻出な単語の傾向や、特定の発話および話者における表出感情と経験感情の関係が明らかになった。また日本語の事前学習済み BERT [17] を用いて、表出感情と経験感情の認識を実験した。経験感情の認識がより難しいことや、表出感情と経験感情のマルチタスク学習が互いの認識に対する性能を向上させることが示された。

謝辞

本研究は LINE 株式会社との共同研究の助成を受けて行った。

参考文献

- [1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, **Advances in Neural Information Processing Systems**, Vol. 30. Curran Associates, Inc., 2017.
- [2] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. 2018.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [4] Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. The design and implementation of XiaoIce, an empathetic social chatbot. **Computational Linguistics**, Vol. 46, No. 1, pp. 53–93, March 2020.
- [5] Daniel Adiwardana, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, and Quoc V. Le. Towards a human-like open-domain chatbot, 2020.
- [6] Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. Recipes for building an open-domain chatbot. In **Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume**, pp. 300–325, Online, April 2021. Association for Computational Linguistics.
- [7] Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. Towards empathetic open-domain conversation models: A new benchmark and dataset. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 5370–5381, Florence, Italy, July 2019. Association for Computational Linguistics.
- [8] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. DailyDialog: A manually labelled multi-turn dialogue dataset. In **Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)**, pp. 986–995, Taipei, Taiwan, November 2017. Asian Federation of Natural Language Processing.
- [9] Chao-Chun Hsu, Sheng-Yeh Chen, Chuan-Chun Kuo, Ting-Hao Huang, and Lun-Wei Ku. EmotionLines: An emotion corpus of multi-party conversations. In **Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)**, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA).
- [10] Robert Plutchik. A general psychoevolutionary theory of emotion. In **Theories of emotion**, pp. 3–33. Elsevier, 1980.
- [11] Sven Buechel and Udo Hahn. EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In **Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers**, pp. 578–585, Valencia, Spain, April 2017. Association for Computational Linguistics.
- [12] Tomoyuki Kajiwara, Chenhui Chu, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. WRIME: A new dataset for emotional intensity estimation with subjective and objective annotations. In **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 2095–2104, Online, June 2021. Association for Computational Linguistics.
- [13] Saif Mohammad and Felipe Bravo-Marquez. Emotion intensities in tweets. In **Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (*SEM 2017)**, pp. 65–77, Vancouver, Canada, August 2017. Association for Computational Linguistics.
- [14] Hannah Rashkin, Antoine Bosselut, Maarten Sap, Kevin Knight, and Yejin Choi. Modeling naive psychology of characters in simple commonsense stories. In **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 2289–2299, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [15] Hiroaki Sugiyama, Masahiro Mizukami, Tsunehiro Arimoto, Hiromi Narimatsu, Yuya Chiba, Hideharu Nakajima, and Toyomi Meguro. Empirical analysis of training strategies of transformer-based japanese chat systems, 2021.
- [16] Ekman Paul. An argument for basic emotions. In **Cognition and Emotion**, Vol. 6(3/4), pp. 169–200, 1992.
- [17] 柴田知秀, 河原大輔, 黒橋禎夫. Bert による日本語構文解析の精度向上. 言語処理学会 第 25 回年次大会, 2019.
- [18] Arseny Tolmachev, Daisuke Kawahara, and Sadao Kurohashi. Juman++: A morphological analysis toolkit for scriptio continua. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 54–59, Brussels, Belgium, November 2018. Association for Computational Linguistics.
- [19] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1715–1725, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [20] Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. Multi-task deep neural networks for natural language understanding. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 4487–4496, Florence, Italy, July 2019. Association for Computational Linguistics.

「B2」を発言した人の感情として適切なものをチェックしてください（複数選択可）。

対話	A1: アコギ見ると欲しくなっちゃうね。性だね B1: 弾けるの A2: ここ数年弾いてないから鈍りまくってそうだけど一応弾ける B2: かつこいい
<input type="checkbox"/> 怒り	
<input type="checkbox"/> 期待	
<input type="checkbox"/> 喜び	
<input type="checkbox"/> 信頼	
<input type="checkbox"/> 恐れ	
<input type="checkbox"/> 驚き	
<input type="checkbox"/> 悲しみ	
<input type="checkbox"/> 嫌悪	
<input type="checkbox"/> どれでもない	

図 3: クラウドソーシングの例

表 8: ラベルごとに頻出な単語の Top-3

ラベル	表出		経験	
	動詞	形容詞	動詞	形容詞
怒り	止める, 許す, 違う	マジだ, うるさい, 馬鹿だ	話す, 広げる, 居る	うるさい, マジだ, 悪い
期待	教える, 願う, 待つ	一緒だ, 楽しみだ, 面白い	待つ, 教える, 上げる	楽しみだ, 強い, 一緒だ
喜び	笑う, 交じる, 孵る	楽しい, 嬉しい, おもしろい	笑う, 交じる, 孵る	楽しい, 嬉しい, おもしろい
信頼	飲む, 願う, らっしゃる	大丈夫だ, 優しい, 一緒だ	教える, 付ける, 待つ	大事だ, 大丈夫だ, 同じだ
恐れ	付ける, 知れる, 助ける	怖い, やばい, 危険だ	知れる, 付ける, 入る	怖い, やばい, 危険だ
驚き	交じる, 気付く, 居る	やばい, 怖い, 強い	居る, 気付く, 知る	怖い, 痛い, やばい
悲しみ	泣く, 生きる, 帰る	痛い, 悲しい, 辛い	泣く, 生きる, 消える	辛い, 痛い, 悲しい
嫌悪	知る, 止める, 交じる	悪い, 嫌いだ, 嫌だ	困る, 上げる, 働く	悪い, 嫌だ, 嫌いだ

A クラウドソーシング

表出感情をタグ付けするクラウドソーシングの例を図 3 に示す。チェックボックス式のフォーマットを用いることによって、複数の感情が選ばれることを許容している。

B 感情ごとに頻出な単語

感情ごとの発話をもつ特性を知るため、感情ごとに単語の頻度を求める。本研究では強ラベルのみを対象とし、品詞として動詞と形容詞を採用する。品詞の解析は Juman++ [18] によって行い、代表表記を抽出する。どの感情にも現れる普遍的な単語を除くため、IDF によるフィルタを施す。具体的には IDF が最大値の半分を下回った単語を無視する。

フィルタを施した、感情ごとに頻出な単語の Top-3 を表 8 に示す。喜びの動詞や喜び・恐れ・嫌悪の形容詞は表出感情と経験感情で等しい。一方で、怒り・嫌悪の動詞や信頼の形容詞は表出感情と経験感情で異なる。喜びや驚きの動詞に「交じる」があるが、これは形容詞の「マジだ」が誤って識別されたものである。同様に、喜びの形容詞にある「孵る」は「帰る」に対する誤りである。