



音声から社会的感情を自動検出する手法の検討

Examination of Automatic Detection of Social Emotions from Utterance Speech

麻田 千尋¹⁾, 田村 祐一²⁾, 佐藤 克成³⁾

Chihiro ASADA, Yuichi TAMURA, and Katsunari SATO

- 1) 奈良女子大学 人間文化研究科 (〒630-8506 奈良県奈良市北魚屋東町, uac_asada@cc.nara-wu.ac.jp)
 2) 甲南大学 知能情報学部 (〒658-8501 兵庫県神戸市東灘区岡本 8-9-1, tamura@konau-u.ac.jp)
 3) 奈良女子大学 人間文化研究科 (〒630-8506 奈良県奈良市北魚屋東町, katsu-sato@cc.nara-wu.ac.jp)

概要: 人は相手や自分の行動に変化を与える社会的感情を表出させ、円滑なコミュニケーションを行う。社会的感情を表出することはソーシャルスキルとして重要であるが、その言い回しを理解するには経験や学習が必要である。本研究では音響的特徴、間合いから機械学習を用いて社会的感情を自動検出する手法を検討する。本提案手法を応用することで、対人コミュニケーションでの発話訓練ツールへの応用が期待される。

キーワード: 社会的感情, ラベリング, コミュニケーション

1. はじめに

人は、婉曲的な相手の発言から「話の意図」という曖昧な情報を読み取り、その意図に応じた反応を返し、円滑なコミュニケーションを行っている。この時、話の意図は言語化されておらず、コミュニケーションを通して相手の表情の変化や間合いの変化といった非言語情報を通して理解していると考えられる。意図を感じ取ることは最低限の社会的スキルとして重要であるが、その言い回しを理解するには過去のコミュニケーションで培われた経験的習慣が必要である。

本研究では、音響的特徴、間合いから相手や自分の行動に変化を与える意図を含んだ社会的感情を、機械学習で自動検出する。音声から社会的感情の自動検出を可能とすることで、オンラインコミュニケーション上で話し手が発話に込めた意図と相手を受け取る意図とのギャップを振り返ることを可能とする。さらに、対面コミュニケーション時のウェアラブルシステムとして実装することで、社会的スキルに課題を持つユーザが自分の意図を正確に伝えるために、対人コミュニケーションでの発話方法を訓練するシステムへの応用が期待できる。本稿では、社会的感情の検出手法を検討するために、社会的感情の定義、ラベルの定義、最適なラベリング方法の検討を行った。

1.1 先行研究

先行研究[1]では音響的特徴、間合いから相手の発話意図を、機械学習の分類手法である SVM (Support Vector Machine) を用いて自動検出する方法を検討した。しかし、意図を含んだ発言を分類することは困難であった。先行研究[1]で精度が上がらなかった原因として、意図を含む発言

の定義が曖昧でラベルが分散した、学習データ数が少ない、時間遷移の要素として追加した間合いが効果を示さなかった、SVM のみで学習した、などがあげられる。先行研究の結果を踏まえて、社会的感情の定義、ラベリング手法を再検討し、社会的感情を含む発話の自動検出する手法の構築を行う。

2. 社会的感情の定義

怒りの表出には、大きく分けて「相手の反応を求めない一方的な怒り」と「相手の行動変化を必要とする怒り」がある。「あの態度はおかしいだろ!」といった発話例では、一方的な感情をぶつけること、つまり自己の感情を外に発散させることを目的としている。これは、相手の行動や反応を必要とせず、自己発散する感情表出であると評価できる。本研究では、相手の行動変化を必要としない感情を基本感情や原始情動のような情動的感情とする。

一方、「コンビニで買ったのか?」[2]では、受け手から見て怒っているように捉えられることを意識した“意図的な怒り”を表出させる。受け手が意図を読み取り、行動の改善を誘導させるような発話である。このように、相手の行動変化を意識した、感情の収束に自分以外の他人を必要とする社会的感情が存在する。本研究では、福田の感情階層説[3]を参考に、感情の収束に相手の反応を必要とする怒りなど、相互作用的な基本情動を社会的感情に含むことで、一方向的な情動的感情と区別する。また、後藤は「社会的感情は他者との関係を考慮して行われる行動の起動因となり得る」と述べており[4]、行動と社会的感情に関連性があると考えられる。以上のような関連研究を参考に、本研究における社会的感情を、「相手や自分を含んだ集団の行動に



Positive や Negative な変化を与えるため、意図的に表出させる感情」として定義した。Positive な変化とは、相手の行動/自分の行動を促進/発展させる、Negative な変化とは、相手の行動/自分の行動を抑制/回避させるものである。

3. システム構成

図 1 に本研究で構築を目指すシステムの処理の流れを示す。

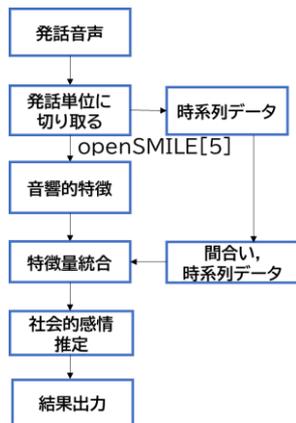


図 1 処理の流れ

本研究では、音声情報から自分の発話に含まれる社会的感情を機械学習でリアルタイムに検出し、意図が伝わるか、客観的に社会的感情を評価、話し手にフィードバックするシステムの構築を行う。音響的特徴の抽出には、openSMILE[5]を使用する。また、社会的感情の推定に使用する学習方法には、ニューラルネットワーク、SVM を検討している。

4. 社会的感情のラベル

社会的感情の定義を踏まえ、音源のラベリングに使用する社会的感情のラベルを定めた。使用するラベルには Positive な社会的感情を含む発言、Negative な社会的感情を含む発言、社会的感情を含まない発言を使用する。図 2 に Positive な社会的感情を含む発言の詳細を示す。

Positiveな社会的感情

- ①相手の行動に対して理解を示して、相手（集団）の行動を受動的（相手の会話の上で）に促す発言
- ②相手の行動に対して理解を示して、相手（集団）の行動を主体的に（自分から）促す発言

図 2 Positive な社会的感情

①は、コミュニケーションの相手に対して、意図が伝わったことを示す発言や維持、補足、発話の継続を促すなど相手の行動を促進する社会的感情である。

②は、相手の行動に対して、相手を強く肯定する、または相互的な提案を行うことで、主体的に相手や集団の行動を促す発言を指す。

これらの他にも人によって様々な社会的感情の表出の仕方が考えられるが、本研究では、大きく①②を Positive な社会的感情とした。

次に、図 3 に示す Negative な社会的感情について説明する。

Negativeな社会的感情

- ①相手（集団）の行動に対して理解ができないことを示し、相手（集団）の行動を受動的に抑制する発言
- ②相手（集団）の行動に対して理解ができないことを示し、相手（集団）の行動を主体的に抑制する発言

図 3 Negative な社会的感情

①は、相手の行動に対しての再説明を促し、相手の意図する行動が発展することを回避するため、相手や集団の行動を受動的に抑制する発言を指す。

②は、主体的に相手の発言を否定、注意することで、相手の行動を主体的に抑制する発言を指す。

これらの他にも人によって様々な表出の仕方が考えられるが、大きく①②を感じた場合を Negative な社会的感情とした。

5. ラベリング

ラベリングには音声コーパスである「宇都宮大学 パラ言語情報研究向け音声対話データベース(UUDB)」[6]の音源を使用した。UUDB には、大学生 7 ペアの話者同士が、4 コマまんがの並び替えタスク[7]に音声のみの非対面で行った際の対話が収録されている。全ての話者が積極的に関与を続けており、声の大きさや高さのバリエーションに富んでいる[6]。関連研究[8]から、パラ言語メッセージと社会的感情が類似しており、使用する音源として適切であると判断した。UUDB の音源に対して、社会的感情をラベリングする。ラベルには、社会的感情の定義を参考に、Positive な社会的感情、Negative な社会的感情、社会的感情を感じないの 3 種類のラベルを使用する。

社会的感情を一般の人に理解してもらい、妥当なラベリングを行うため、また本研究で定義した社会的感情を人が認識できることを確認するために、2 種類の予備実験を行った。

5.1 予備実験 1

予備実験 1 では、UUDB の 4 種類の音源を話者ごとに分割した 8 種類の音源を使用した。連続した一話者の音源を聞いて社会的感情を感じたら、ソフトのボタンを押すことでラベルの付与を行う。この手法では、音源の聞き直しは不可として、直感的に社会的感情を判断するラベリングを行ってもらった。その後、UUDB 記載の発話単位に分割した音声に、発話区間に該当するラベルを付けた。被験者は 20 代・50 代の男女 6 名である。実験終了後 20 代の男女 5 名にはアンケート回答を行い、社会的感情への理解度、社会的感情の感じ方を評価した。

予備実験の結果、各音源にラベリングされた回数を図 4 に示す。

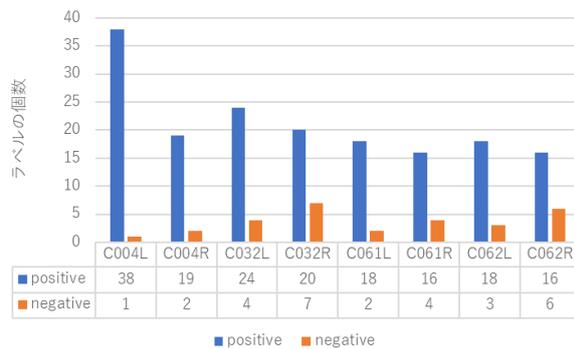


図 4 ラベルの分布

C004L, C004R (3分49秒), C032L, C032R (4分) は女性の音源, C061L, C061R (4分24秒), C062, C062R (4分9秒), は男性の発話音源である。C061Lの音源ではソフトの不具合により, 被験者1名のラベルデータが不足していた。

ラベルの付与された発話のうち, C004Lの「じゃあさ, これはさ眼鏡を外したら超かっこ良かったみたいな感じなんだね」は, 被験者全員が Positive な社会的感情であると回答した。一方, C061Rの「DからC, CからDじゃないのあら, あれ」の発話では被験者3名が Negative な社会的感情であると回答した。また, C062Rの「なにこいつみたい顔してる」には Positive 1つ, Negative 1つのラベルが付与された。

次にアンケートの結果を示す。「ラベリングに対する説明を聞いて(読んで)理解できましたか」の回答には、「分かった」が4名, 「分からなかったが」1名, 「ラベリングを直観的に行えましたか」の回答では, 「直観的に行えた」が1名, 「少し考えてからラベリングをした」が4名, 「音声を聞いて社会的感情を感じ取れましたか」の回答では, 「感じ取れた」が3名, 「感じ取れなかった」が2名, また, 「感じ取れた」と回答した人に「社会的感情を感じ取れた人は, 特定の一つの発話に対して」社会的感情を感じましたか?それとも, 「連続した発話」に対して社会的感情を感じましたか」の回答では, 「連続的な発話」に対して社会的感情を感じた」が2名, 「ある特定の一つの発話」に対して社会的感情を感じた」が1名という結果になった。

5.1.1 考察

Positive ラベルの集中した C004L の発話では, 明るいトーンの「じゃあさ」で相手の意図への理解を示し, 「これはさ眼鏡を外したら超かっこ良かったみたいな感じなんだね」の部分で, C004L が解釈した相手の意図を具体的に説明している。C004L は, 話し手に受け手が感じた意図の解釈が正しいかの回答を求め, 相互的に今の会話を発展させようとする Positive な社会的感情の②に該当する。

C061R の発話は, 自分の考えと相手の考えが違う意図を

表出させ, 相手が望む, コミュニケーションの方針や行動を主体的に抑制する②の Negative な社会的感情である。音源とラベルの分布から, ②に該当する主体的な社会的感情は, 強く意図が表出するため, ①の受動的な感情表出よりもラベルが集中する傾向がある。

C062R の発話は, 連続する発話の一部である。直前の発話である「がみがみもしてないなんかね」で相手の発話に対して応答した後, ラベルの分散した「なにこいつみたいな顔してる」の「何こいつ」の部分でマンガに出てくる登場人物のセリフを意識して, 嫌悪感を意図的に表出させていた。そのため, 発話に表出した複雑な感情を見極める時間の欠如や, コーパスの発話単位と被験者のつけたい発話位置が一致せず, 正しいラベルが付与されなかった。

アンケートの結果から, 社会的感情を理解するためには, 文字情報のみの資料では個人差があることが判明した。習慣化された社会的感情を感じるためには, テキスト以外の非言語情報を駆使して理解しているためと考える。

予備実験1の結果から, 社会的感情の認識には, 一定の時間や訓練が必要で直感的ではないこと, 社会的感情を認識する発話単位とコーパスで設定された発話単位が異なることが示唆された。すなわち, 直感的なラベリングは, ラベルの判断ミスや誤操作を招くため, 適していない。予備実験1の結果を踏まえて, 予備実験2を行った。

5.2 予備実験2

予備実験2では, Audacity という音源解析ソフトを使用して, ラベリングを行った。Audacity は, 音声を波形として可視化すること, 被験者が任意の発話区間を指定可能, といった特徴がある。

予備実験2では, 予備実験1とは異なる UADB の音源を話者ごとに分割した4種類の音源を使用した。Audacity のラベル機能を用いて, 社会的感情を感じた発話区間すべてにラベリングを行った。この手法では, 音源に対してどんな社会的感情を感じるかをはっきりと認知できるまで聞き直し可能であるとして, 実験を行ってもらった。被験者は20代の男女4名である。実験終了後アンケート回答を行い, 社会的感情への理解度, 社会的感情の感じ方を評価した。また, 予備実験1の結果から, 社会的感情の音源例と, ある一人(研究者)のラベリングの回答例を備えた練習問題を作成した。練習問題を通して, 文字以外の音源情報から社会的感情がどのようなものか理解させる目的がある。

各音源でラベリングされた回数を図5に示す。

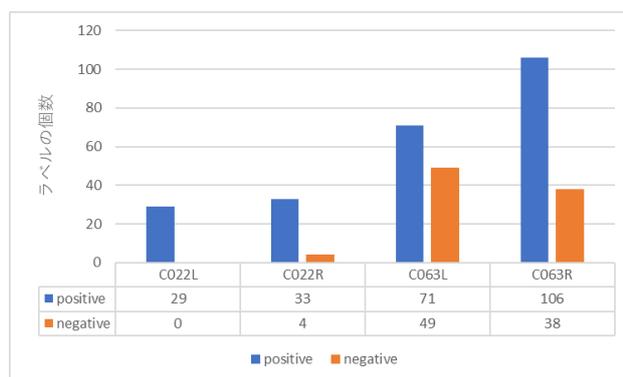


図 5 ラベルの分布

C022L, C022R (1 分 51 秒) は女性の音源, C063L, C063R (8 分 20 秒) は男性の発話音源である。発話ごとのラベリングから発話区間のラベリングに変更したことで、一つの発話に対するラベル量が増加した。C022R の「投げんのはどこだろう」の発話は、被験者 4 名全員が Positive な社会的感情と回答した。一方で C063L の「え、だから説明してよ」は 4 名全員が Negative な社会的感情と回答した。

アンケートの結果では、「前回のラベリング作業よりも今回の方がやりやすかったですか」の項目で、「やりやすかった」が被験者全員の 4 人「ラベリングを前回よりも間違いなく正確に行えましたか」の回答には、「行えた」が被験者全員の 4 人であった。

5.2.1 考察

C022R の「投げんのはどこだろう」の発話は、相手の行動に対して、相互的な提案を行う②の Positive な社会的感情である。C063L の「え、だから説明してよ」は、わからないという意図をはっきり伝えて主体的に相手の行動を変化させる②の Negative な社会的感情である。

予備実験 2 の結果から、予備実験 1 の手法よりも被験者が発話区間を任意に定めて、ラベリングを行う手法が社会的感情を適切に判別できたと考える。一方で、社会的感情の音源例のついた説明資料であっても認識に個人差がみられるため、さらなる工夫が必要である。

6. 予備実験 1, 2 に関する考察

今回使用した音源に対するラベルの分布から、性別を問わず人は、Positive な社会的感情を頻繁に使用することで意図を伝えることがわかった。特に女性は、Positive な社会的感情を駆使して、発話に Negative な意図が含まれていたとしても、Positive な社会的感情として認識されるようなコミュニケーションを行っていた。そのため、本研究で取り扱う情動的感情に近い Negative な社会的感情を被験者が検出することが困難であった。一方、男性話者は意図を婉曲的に表出させず相手に伝わるように社会的感情の表出を行っていたため、Positive, Negative ラベルが比較的均一にラベリングされた。

以上の予備実験から、最終的なラベリング手法を定めた。

ラベリングには、audacity を用いて、被験者が理解するまで音源の発話を聞きなおし、被験者が任意の発話区間に対してラベリングを行う手法を用いる。

7. 終わりに

本研究では、社会的感情を機械学習で自動検出する手法を検討するために、社会的感情の定義、ラベルの定義、最適なラベリング方法の検討を行った。今後はラベリング作業を行い、社会的感情を機械学習で分類可能かを検証する予定である。

8. 謝辞

本研究にあたり、助言や議論を重ねてくださった甲南大学の永田先生に心から感謝します。本当にありがとうございました。

参考文献

- [1] 麻田千尋, 田村祐一, 永田亮, “音声から相手の発話意図を自動検出する手法の提案”, 2020 年電子情報通信学会総合大会論文集, 2020.
- [2] “きのう何食べた?”, 第一話, テレビ東京, 2019.
- [3] 福田 正治, “感情の階層性と脳の進化”, 感情心理学研究, 16 巻, 1 号, p. 25-35, 2008.
- [4] 後藤品, “社会的感情の機能とその進化に関する一考察 -社会的行動に対する二重過程理論的アプローチ”, 情報コミュニケーション研究論文集, 第 1 号, 2010.
- [5] Eyben, Florian, Martin Wöllmer, Björn Schuller. "Opensmile: the munich versatile and fast open-source audio feature extractor. ", Proceedings of the 18th ACM international conference on Multimedia, pp.1459-1462, 2010.
- [6] 宇都宮大学パラ言語情報研究向け音声対話データベース, NII 音声資源コンソーシアム, <http://uudb.speechlab.org/>.
- [7] H. Mori, H. Kasuya, M. Nakamura and M. Amanuma, “Some considerations for designing spoken dialogue database from the viewpoint of paralinguistic information”, Acoustical Science and Technology, vol.24, no.6, pp.376-378, 2003.
- [8] 森, 大毅, 前川 喜久雄, 粕谷 英樹, “音声は何を伝えているか: 感情・パラ言語情報・個人性の音声科学”, コロナ社, 2014.