

話者の性格特性を考慮した対話音声からの対話雰囲気推定

Dialogue Mood Estimation from Dialogue Voice Considering Personality Trait of Speaker

長谷川達也 *1 加藤昇平 *1*2

Tatsuya Hasegawa

Shohei Kato

*1名古屋工業大学 大学院工学研究科情報工学専攻

Dept. of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology

*2名古屋工業大学 情報科学フロンティア研究院

Frontier Research Institute for Information Science, Nagoya Institute of Technology

For natural conversation with people, dialogue system has to return an appropriate reaction depending on dialogue mood. In this study, we aim to develop a method of estimating dialogue mood for that. We conducted the experiment to collect dialogue speech data, dialogue mood evaluation, and personality traits of subjects. Dialogue speech data and dialogue mood evaluation was used for calculating intervals of utterance state. Our method estimated dialogue mood by these feature values. In this paper, we verify effectiveness of personality traits added this time and the method of dialogue mood estimation. As a result, personality traits are effective on dialogue mood estimation in a part of dialogue mood labels, and the result suggests relevance of personality features with dialogue mood.

1. はじめに

近年、人と対話システムによるコミュニケーションを目的とした研究が盛んに行われている。具体的には、コミュニケーションにおけるシステムに対するユーザの親しみやすさの向上を目的とした研究 [福田 18] や、システムとユーザとの対話がより自然に行われることを目指した研究 [小林 15][高杉 10] などがあげられる。これらの研究は、人とシステムとの一対一のコミュニケーションを目的としており、人とシステムとが対等な立場で対話することを目指しているといえる。

人とシステムが対等な立場で対話するためには、システムに対する人側からの配慮を減らす必要がある。そのための方法の一つとして、対話の雰囲気に応じた適切な反応を返すことが考えられる。適切な応答を返すには対話雰囲気を推定する機能が不可欠である。そこで、本研究では対話の雰囲気を推定するシステムの構築を目指す。

そのためにまず、人とシステムとの対話の雰囲気を推定する要素研究として、1対1対話に焦点を当てた対話雰囲気の推定を試みた。本稿では、雰囲気ラベル付き音声データセットの収集実験と、それに基づく推定手法および特徴量を検討する。

2. 関連研究

人同士のコミュニケーション時の対話雰囲気を扱う研究として、Wrede ら、徳久ら、稲葉ら、豊田らなどの研究があげられる。Wrede らは、議論の盛り上がりに応じて、対話音声を4クラスに分類し、各クラスごとに韻律特徴と発話の種類を表す Dialogue Act(DA) タグによる分析を行った [Wrede1 03][Wrede2 03]。徳久らは、DA タグと修辭構造を表す Fhetorical Relation タグを用いて、それぞれのタグと対話の盛り上がりの関係性について分析した [徳久 06]。稲葉らは、テキスト対話の単語の共起情報から対話の雰囲気推定を行った [稲葉 11]。これらの研究では、タグやテキストなどの言語情報を利用しているが、

実際の対話中にこういった情報を正確に抽出するのは難しく、本研究の目的である人とのコミュニケーションに活かす対話雰囲気推定システムの構築には利用できないと考えた。また、豊田らは、対話時の発話時間特徴に着目して、第三者視点からの対話雰囲気推定を行った [豊田 12]。発話時間特徴は言語情報や韻律情報などとは違い、抽出において高い計算コストを必要としない。また、マイクの距離や発話音量の個人差などの影響を受けない。そのため、実際の対話中に情報を抽出することを考えても、対話雰囲気推定システムを構築するうえで扱いやすい特徴量であるといえる。だが、一対一のコミュニケーションにおいてこのシステムを使用することを考えると、推定の視点が第三者であるため、システム自身が対話するうえで活用するのが困難ではないかと考えられる。

非言語情報と対話雰囲気の関連を調査した研究として Gatica-Perez ら、伊藤ら、森らの研究がある。Gatica-Perez らは、会議中の動画と音声を用いて議論の盛り上がり部分の推定を行っている [Gatica-Perez 04]。伊藤らは、2者による自由対話を「盛り上がっている」、「盛り上がっていない」の2クラスに分類し、各クラスごとに話者の発話および動作との関係性の分析を行っている [伊藤 02]。森らは、対話を発話時と非発話時の2つの状態に場合分けを行い、状態毎の各話者の顔表情に基づいて対話雰囲気の推定を行っている [森 10]。しかし、これらの研究に用いられる動作認識や顔表情認識は高い計算コストを必要とし、人とのリアルタイムなコミュニケーションには向いていないと考えられる。

本研究では、豊田らの研究でも用いられた発話時間特徴を対話雰囲気の推定に利用する。ただ、対話への介入を想定しているわけではないため、本研究において使用する特徴量の有効性の検証および、より適切な特徴量の選定が必要である。

3. 対話雰囲気推定システム

図1に本研究で想定している対話雰囲気推定システムの概要図を示す。本システムは、話者自身の視点から見た対話雰囲気を推定するために、特徴量を入力する際に話者とその対話相手の

連絡先: 加藤昇平, 名古屋工業大学, 愛知県名古屋市昭和区御器所町, 052-735-5625, shohey@katolab.nitech.ac.jp

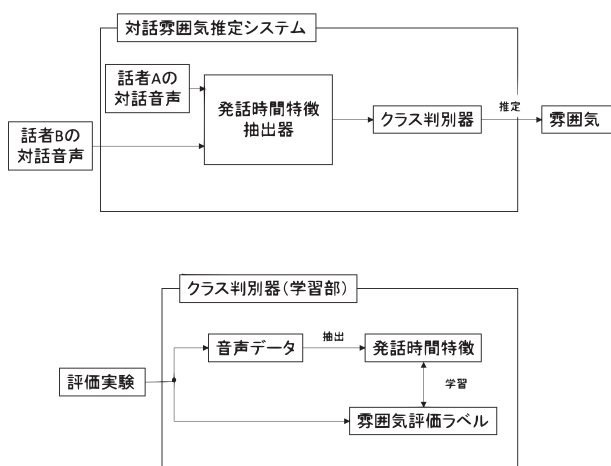


図 1: 対話雰囲気推定システム

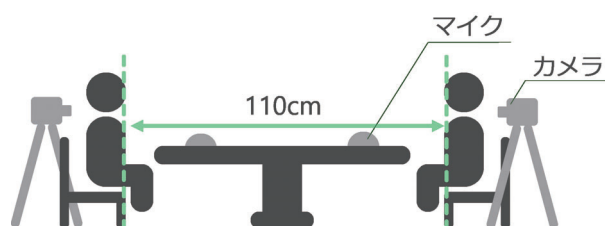


図 2: 対話実験の模式図

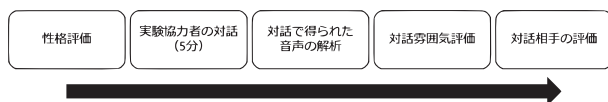


図 3: 対話実験の手順

音声をそれぞれ決まった形で特徴量抽出器に入力する。

具体的には、話者自身の音声（対話システムの場合は、システム発声情報）を話者 A 側、話者から見た対話の相手の音声を話者 B 側の対話音声としてそれぞれ入力する。入力後、今回使用する特徴量の 1 つである発話時間特徴を算出し、クラス判別器の入力として用いる。そして、クラス判別器から得られた出力を、入力とした対話音声における対話雰囲気の推定結果として出力する。

4. 対話雰囲気評価実験

4.1 実験概要

実験協力者は 18~24 歳の大学生 28 名（男性 13 名、女性 15 名）であり、全員がそれぞれ 1 回ずつ実験に参加した。実験の組み合わせとして、「男性 - 女性」5 組、「男性 - 男性」4 組、「女性 - 女性」5 組の計 14 組により実施した。図 2 に実験の概要図を示す。実験は、2 人 1 組で対面に着席させた状態で、対話時のそれぞれの様子を音声及び動画で計測する。図 3 に実験手順を示す。まず、推定システムの入力の 1 つである性格特徴を得るために、協力者に自信の性格評価に関するアンケートに回答させた。続いて、実験者は事前に用意した話題の中からランダムに 1 つ提供し、実験室から退室した。実験協力者には提供した話題になるべく沿う形で 5 分間の対話をさせた。対話終了後、計測した音声から音声区間抽出を行い、それぞれを実験協力者に順に聴取させる。実験協力者には区間ごとで対

非常に当てはまる どちらでもない 全く当てはまらない
話し好き ————— 話さない

図 4: 性格評価アンケートの例

当てはまる どちらでもない 当てはまる
盛り上がった ————— 停滞した
まじめな ————— くだけた
噛み合った ————— ぎこちない
明るい ————— 暗い
落ち着いた ————— 落ち着きのない
親密な ————— 立場の違う
対等な ————— 疎遠な

図 5: 対話雰囲気評価アンケートの構成

話雰囲気評価を行わせた。最後に、対話相手に関する評価をさせ、実験は終了した。

4.2 性格評価の方法

対話の前に、実験協力者には自身の性格評価をさせるためのアンケートに回答してもらった。アンケートは和田らの全 60 項目からなる日本語版 BigFive 尺度 [和田 96] を用いた。図 4 に、実験で用いたアンケートから 1 項目を例として示す。それぞれの項目が自分にどれだけ当てはまるかで 7 件法により回答してもらった。

4.3 音声区間抽出

音声区間の抽出は以下の手順で行った。まず、抽出する区間を選ぶ前に、計測した音声に対して発話区間検出をかける。発話区間検出は音声を有聲区間と無聲区間のそれぞれの系列に分割するために行う処理である。分割した区間のうち、無聲区間として検出されたものの開始時刻と終了時刻を記録し、それを基に評価対象の音声区間を抽出する。抽出する区間の選定基準としては、「区間の両端が 0.3 秒以上の無聲区間」、かつ「どちらの実験協力者も一度以上は発話している」とした。そして、全体の音声データからこの基準を満たす 20-40 秒の時間長をもつ区間を抽出した。また、抽出した区間に時間軸的な偏りがあると、類似したデータが集中してしまうと考えた。そのため、音声データ全体を 4 つの区間に分割し、それぞれに評価対象ができるだけ存在するように選定した。区間に選定基準を満たすものがない場合は、時間軸上でその区間のできるだけ近傍に存在できるように選定した。なお、本条件で候補区間が不足するという事象は全実験を通して観測されなかった。

4.4 対話雰囲気評価の方法

音声区間抽出後、実験開始時刻に近いものから順に、4 つの音声区間の対話雰囲気を実験協力者に評価をさせた。図 5 に、実験協力者に対話雰囲気の評価させる際に用いたアンケートの構成を示す。アンケートには対話雰囲気を表す形容詞対 7 種を用いる。左側の形容詞は対話雰囲気のラベルの意味をそのまま表すような、右側の形容詞は左側の形容詞を否定するような意味を持つような形容詞をそれぞれ配置した。対話実験の後、実験者が実験時の音声から抽出した評価対象の区間に聴取させる。実験協力者には、それぞれの区間ごとで、対話時にどのような印象を対話に対して持っていたかを思い出させながらアンケートに回答させた。対話直後に実験協力者に直接評価させるのは、作成した音声区間に対して、対話者視点での対話雰囲気ラベルを付与するためである。

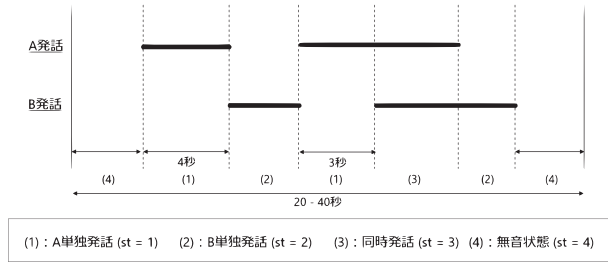


図 6: 発話状態の分割例

5. 特徴量

5.1 性格特徴

性格評価によって得られた評価データから、Big Five 尺度において個人の性格特性を表す「外向性因子」「神経症傾向因子」「開放性因子」「誠実性因子」「調和性因子」をそれぞれ算出する。本稿では、これらすべてを性格特徴として対話雰囲気推定で使用し、それぞれに特徴番号 001-005 を割り振る。個人の性格特徴を特徴量として用いることで、被験者属性を考慮した推定ができるため、対話雰囲気の捉え方における個人差を軽減できると考え、対話雰囲気推定の特徴量として採用した。

5.2 発話時間特徴

対話者 A, B をそれぞれの実験協力者とする。本研究では対話者 2 名の発話状態 st を次のように定義する。

$$st = \begin{cases} 1 & \text{(A 単独発話状態)} \\ 2 & \text{(B 単独発話状態)} \\ 3 & \text{(同時発話状態)} \\ 4 & \text{(無音状態)} \end{cases}$$

実験時に抽出した音声をそれぞれ上記の 4 つの状態の系列に分割し、それぞれ開始時刻と終了時刻の差をとって時間長を算出する。そして、それらを要素とする状態毎の多重集合 S_{st} を生成し、発話時間特徴の算出に利用する。図 6 に、対話 d における発話状態の分割例を示す。この場合、A 単独発話 ($st = 1$) は 2 秒, 3 秒があり、多重集合 $S_1^d = \{2, 3\}$ となる。

表 1 に発話状態集合から算出する計 95 種の発話時間特徴とその算出式例を示す。このうち、特徴番号 006-025 の計 20 種は発話時間統計特徴である。4.1 で定義した発話状態集合 S_{st}^d を基に以下に示す 5 種類の関数によって統計量を算出する。

$\text{mean}(S_{st}^d)$	: 発話状態集合 S_{st}^d の平均
$\text{var}(S_{st}^d)$	: 発話状態集合 S_{st}^d の分散
$\text{min}(S_{st}^d)$	: 発話状態集合 S_{st}^d の最小値
$\text{max}(S_{st}^d)$	: 発話状態集合 S_{st}^d の最大値
$\text{occupy}(S_{st}^d)$	: 発話状態 st の評価区間内における占有率

また、特徴番号 026-100 の計 75 種は話者間時間特徴であり、それぞれの発話時間統計特徴を st の各状態の発話時間統計特徴間で比較するための指標である。表 1 における関数 $\text{stat}(S_i^d)$ は先述した 6 種の関数のうちのいずれかを適用することを表す。発話状態間の比較は、対話における話者間の状態を明らかにできるため、今回用いる特徴量として有効であると考えた。

表 1: 発話時間特徴

特徴番号	特徴量	算出式例
006 - 010	A 発話に関する統計量	$\text{stat}(S_1^d)$
011 - 015	B 発話に関する統計量	$\text{stat}(S_2^d)$
016 - 020	同時発話に関する統計量	$\text{stat}(S_3^d)$
021 - 025	無音時間に関する統計量	$\text{stat}(S_4^d)$
026 - 045	A 発話と B 発話の比較	$\frac{\text{stat}(S_1^d)}{\text{stat}(S_1^d) + \text{stat}(S_2^d)}$
046 - 050	発音状態と無音状態の比較	$\frac{\text{stat}(S_4^d)}{\sum_i \text{stat}(S_i^d)}$
051 - 070	全状態における各状態の割合	$\frac{\text{stat}(S_1^d)}{\sum_i \text{stat}(S_i^d)}$
071 - 085	A 発話と同時発話の比較	$\frac{\text{stat}(S_3)}{\text{stat}(S_1) + \text{stat}(S_3)}$
086 - 100	B 発話と同時発話の比較	$\frac{\text{stat}(S_3)}{\text{stat}(S_2) + \text{stat}(S_3)}$

6. 対話雰囲気推定実験

6.1 実験概要

対話実験から得られた学習サンプル 100 個による Leave-Human-Out クロスバリデーションによる性能評価を行う。今回行った対話実験の中で、性格評価において未入力箇所があった実験協力者が 3 名いたため、残りの 12 個は学習サンプルとして今回使用しないこととする。Leave-Human-Out クロスバリデーションとは、テストデータと同一人物のデータを学習データに含めない検証方法である。

推定における目標出力として、評価実験で収集した対話雰囲気評価ラベルを用いる。クラス分けの方法については、どちらでもないより左側なら肯定、右側なら否定、真ん中を直接回答した場合のみ中間とした。

本稿では、以前我々[長谷川 18]が提案した発話時間特徴のみを特徴量とした Random Forest による推定手法と、性格特徴を特徴量として新たに追加した Random Forest による推定手法とを推定性能により比較し、性格特徴の対話雰囲気推定における有効性を検討する。

6.2 実験結果

表 2 に、比較したそれぞれの手法の推定結果から得られた F 値と正答率を示す。「盛り上がり」「まじめさ」「噛み合い」「明るさ」の 4 つラベルにおいて、先行研究と本実験のそれぞれの正答率に対して二項検定を行ったところ、有意水準 5% を満たした。このことから、これらのラベルについては、性格特徴が推定において有効であったと考えられる。

6.3 考察

今回、性格特徴を特徴量として追加したうえで作成した推定器をラベル別に見ると、有意差が確認されたラベルと確認されなかったラベルの両方が存在した。つまり、本稿で使用した性格特徴における被験者属性の構成要素に、推定において適切な要素があったラベルと、適切な要素がなかったラベルが存在したということになる。

表 3 に、有意差が確認されたそれぞれのラベルにおいて、重要度が割り当てられた性格特徴の特徴番号を重要度の高い順で示す。ラベルによって性格特徴の構成要素における優劣が異なる

表 2: 異なる特徴量による Random Forest の推定結果の比較

対話雰囲気 ラベル	特徴量 1		特徴量 2	
	F 値	正答率 (%)	F 値	正答率 (%)
盛り上がり	0.312	46.0	0.344	49.0
まじめさ	0.365	46.0	0.403	56.0
噛み合い	0.294	49.0	0.364	54.0
明るさ	0.285	44.0	0.364	56.0
落ち着き	0.355	46.0	0.356	46.0
親密さ	0.318	44.0	0.343	46.0
対等さ	0.279	38.0	0.318	45.0

特徴量 1: 発話時間特徴のみ

特徴量 2: 性格特徴 + 発話時間特徴

表 3: Random Forest で重要度を割り振られた性格特徴一覧

対話雰囲気ラベル	重要度が割り振られた性格特徴
盛り上がり	002, 001, 004, 005
まじめさ	001, 005, 003, 002, 004
噛み合い	001, 005, 002, 003, 004
明るさ	003, 001, 002, 005, 004

※割り振られた重要度の高い順

り、重要度が割り当てられていないものも存在していることがわかる。

これらを踏まえると、対話雰囲気ラベルごとに関係してくる被験者属性が違ってくのではないかと考えられる。ラベルごとにそれぞれ適した被験者属性による特徴量を構成することができれば、各ラベルに対する推定性能向上が期待できる。

また、被験者属性を特徴量として使用するのではなく、被験者属性によって実験協力者のクラスタリングを行い、クラスタごとで推定器を作成することで、被験者属性を考慮した推定をすることが可能である。ただし、この方法を用いる場合、各クラスタにデータが分散してしまうため、その分の各推定器に対する学習データ数の不足が課題となる。

7. おわりに

本稿では、対話雰囲気ラベル付き音声データセットの収集実験と対話雰囲気推定実験とそれを用いる特徴量についての考察を述べた。推定実験では、発話時間特徴のみを用いた Random Forest 推定器と、発話時間特徴と性格特徴を用いた Random Forest とを比較した。その結果、一部のラベルにおいて性能向上を確認し、今回における性格特徴の一部のラベルに対する対話雰囲気推定への有効性が示唆された。

また、今回使用した性格特徴は、一部のラベルにおいて有効でなかった。性格特性の特徴量としての使用方法を変更するか、他の被験者属性を導入する必要があると考えられる。他の被験者属性の具体的な例としては、実験協力者の年齢や性別などがあげられる。

今後は、各ラベルにおける有効な特徴量の検討および被験者属性を用いた学習データのクラスタリングを検討することで、より精度の高い対話雰囲気推定システムの構築を目指す。

参考文献

- [Gatica-Perez 04] Gatica-Perez, D., McCowan, I. A., Zhang, D., and Bengio, S.: Detecting Group Interest-level in Meetings, Idiap-RR-51-2004, IDIAP, Martigny, Switzerland, 2004
- [Wrede1 03] Wrede, B. and Shriberg, E: The Relationship between Dialogue Acts and Hot Spots in Meetings in PROC, IEEE Automatic Speech Recognition and Understanding Workshop(ASRU, Virgin Is-lands(2003)
- [Wrede2 03] Wrede, B. and Shriberg, E: Spotting “Hot Spots” in Meetings : Human Judgments and Prosodic Cues, in Proc. Eurospeech, pp.2805-2808(2003)
- [稲葉 11] 稲葉通将, 烏海不二夫, 石井健太郎: 語の共起情報を用いた対話における盛り上がりの自動判定, 電子情報通信学会論文誌.D, 情報・システム, Vol.94, No.1, pp59-67(2011)
- [伊藤 02] 伊藤秀樹, 重野真也, 西野卓也, 荒木雅弘, 新美康永: 対話における雰囲気分析, 情報処理学会研究報告. SLP, 音声情報処理, Vol.2002, No.10, pp.103-108(2002)
- [小林 15] 小林弘幸, 大村卓矢, 山本知仁: 音声対話システムにおける挨拶発話の適切なタイミング生成, 計測自動制御学会論文集, Vol.51, No.4, pp233-239(2015)
- [高杉 10] 高杉将司, 吉田祥平, 沖津健吾, 横山正典, 山本知仁, 三宅美博: コミュニケーションロボットとの対話における交代潜伏長と傾き先行時間長の影響評価, 計測自動制御学会論文集, Vol.46, No.1, pp.72-81(2010)
- [徳久 06] 徳久良子, 寺島立太: 雑談における発話のやりとりと盛り上がりの関連, 人工知能学会論文誌, Vol.21, pp.133-142(2006)
- [豊田 12] 豊田薫, 宮越喜浩, 山西良典, 加藤昇平: 発話時間長に着目した対話雰囲気推定, 人工知能学会論文誌 27 巻 2 号 SP-B, pp16-21(2012)
- [長谷川 18] 長谷川達也, 加藤昇平: Random Forest を用いた発話時間特徴からの対話雰囲気推定, 第 14 回日本感性工学会春季大会 (2019(to appear))
- [福田 18] 福田和彰, 加藤昇平: 親しみやすい感性対話エージェントの実現に向けた動的思考更新手法, 電気学会論文誌 C (電子・情報・システム部門誌), Vol.138, No.12, pp1485-1492(2018)
- [森 10] 森博章, 宮脇健三郎, 西口敏司, 佐野睦夫, 山下信行: グループ感情場モデルによる感情推定方式の検討, 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol.109, No.470, pp519-523(2010)
- [和田 96] 和田さゆり: 性格特性用語を用いた Big Five 尺度の作成, 心理学研究, 第 6 7 巻第 1 号