対面・遠隔対話における
称賛行為検出

令和4年度　卒業論文

日本大学　文理学部　情報科学科　宮田研究室

田原　陽平

概要

褒める行為は重要なコミュニケーション要素であり，相手を上手く褒めるためにはどのような振る舞いが重要であるか明らかにすることには重要な意義がある．我々はこれまでに，話者が相手を褒めているシーンに対して，褒め方の上手さを自動推定する試みを行ってきた．しかし，先行研究は分析対象シーンでは褒める行為が行われていることを前提としており，当該シーンで褒める行為が行われているか否かの判定は行っていない．そこで本稿では，対面，遠隔の両環境における対話シーンに対して，各シーンで話者が相手を褒めているのか否かを自動判定できるか検討を行う．加えて，対面，遠隔の環境の差が自動判定に及ぼす影響の有無と原因の調査結果も報告する．具体的には，対面および遠隔対話のコーパスを利用し，言語，非言語行動から褒める行為を検出する機械学習モデルの構築を行う．この結果，視覚的，韻律的，言語的特徴量を用いることで褒める行為を検出できることが明らかになった．

目 次

図 目 次

表 目 次

第1章　序論

1.1 研究の背景

褒める行為は，日常生活や社会活動において重要なコミュニケーションのひとつである [1]．グループや組織において，相手の良い側面を積極的に褒めることで相手の自信の向上や互いの信頼関係の構築が円滑に進むことが期待される．しかし，褒めることが苦手な人は，相手を褒めるためにどのように振舞うとよいのかわからないという問題がある．この問題を解決するために，我々は対話中の言語・非言語行動を利用して，相手を褒めるための振舞いを分析する取り組みを行う．この取り組みを行うにあたり，対話中の褒める行為を検出し，褒める際に重要な振舞いを明らかにする必要がある．先行研究では，褒め方の上手さに着目し，褒め方の上手さを推定する機械学習モデルの構築や相手を上手く褒めるための行動を分析してきた [2][3][4][5][6][7]．しかし，これらの研究では話者が褒めているシーンにおける褒め方の上手さを分析対象としており，褒める行為を検出する取り組みは行われていない．褒める行為を検出することで，褒める際に重要な振舞いや褒める行為特有の振舞いが明らかになると考えられる．加えて，COVID-19 の影響により，遠隔でコミュニケーションを行う機会が増えてきている．対面対話と遠隔対話では，言語，非言語行動の伝わる情報が異なることが知られている [8]．これより，褒める際に，対面対話と遠隔対話で伝わる情報や重要な振舞いが異なることが考えられる．対面，遠隔対話における褒める行為の検出を行い，褒める行為の振る舞いの違いを明らかにする．

1.2 研究の目的

1.1 節より，本稿では，対面および遠隔対話における褒める行為の検出を行う．具体的には，対面および遠隔対話のコーパスを利用し，言語・非言語行動から褒める行為を検出する機械学習モデルの構築を行う．これにより，対面および遠隔対話における褒めるときと褒めていないときの振舞いの違いや褒める行為特有の振舞いを定量的に判断できるようになると考えられる．

1.3 本論文の構成

本論文の構成は次のとおりである．

2 章では，対話中の振る舞いから行動・能力を分析する研究事例と褒める行為に関する研究事例について述べる．
3 章では，本論文における問題の定義と研究課題について述べる．
4 章では，対話コーパスの構築について述べる．
5 章では，特徴量抽出と機械学習モデルの構築について述べる．
6 章では，構築した機械学習モデルの結果・考察について述べる．
最後に 7 章にて，本論文の結論を述べる．

第2章　対話中の振る舞いから行動・能力を分析する研究事例

本章では，対話中の振る舞いから行動・能力を分析する研究事例と，褒める行為に関する研究事例について述べる．これらは，言語的・非言語的行動を利用して，対話中の行動・能力を分析するという点で本研究と関係している．2.1 節では，言語，非言語情報を用いた対話中のスキルや振舞いを推定する研究事例に関する研究事例について紹介する．2.2 節では，褒める行為に関する研究事例について紹介する．

2.1　言語，非言語情報を用いた対話中のスキルや振舞いを推定する研究事例

人間の言語的・非言語行動的に着目し，コミュニケーションスキル，プレゼンテーションスキル，話者の説得力，共感スキルなどの行動や能力を推定する研究は多数行われている．言語的，非言語的情報を用いた対話中のスキルや振舞いを推定する研究事例として [9]，[10]，[12]，[15]，[11]，[13]，[14]，[16]，[17]，[18] が挙げられる．Okada らは，グループ内の個々のコミュニケーションスキルに着目した．人事管理経験者の評価をもとに言語的，非言語的情報に着目して回帰モデルを構築した．コミュニケーションスキルのスコアを推定するモデルを構築した結果，単体の特徴量を用いるより，複数の特徴量を用いたモデルが最も性能の良いモデルとなった [9]．Wörtwein らは，プレゼンテーションスキル向上のための機械学習モデルの構築を行っている．非言語情報に着目し，コミュニケーションスキルの自動評価と人前で話す行動のマルチモーダルのモデルを構築した．特に，顔の表情．声の特徴，ジェスチャーがコミュニケーションスキルの向上に貢献することを明らかにした．また，単体のモダリティを用いることよりも視覚的，韻律的特徴量の複数のモダリティを用いたモデルの方が性能が高い結果になった．[10]．Yagi らは，プレゼンテーションスキルの評価を行っている．プレゼンテーションシーンから得られる映像，発話内容から言語的，非言語的特徴量を抽出した．抽出した特徴量をもとに．プレゼンテーション能力を推定するモデルの構築・評価を行った．結果，音響的特徴量がプレゼンテーションスキルの評価に重要であることが明らかになった．[12]．Chen らは，プレゼンテーションを評価するためのマルチモーダルスコアリングを構築している [15]. Oya らは，言語・非言語情報に着目し，遠隔会議と対面会議における対人印象の形成に及ぼす影響を分析している [11]．Ishii らは，複数人でのディスカッションでの共感スキルレベルの推定をするため，視線行動と対話行為に関する情報を用いて共感スキルレベルを推定するモデルを構築した [13]．Soleymani らは，人やエージェントへの自己開示に関連する言語行動と非言語行動を分析し，言語行動における感情的・認知的内容などの言語的特徴と，頭のジェスチャーなどの非言語行動が，親密な自己開示と関連することが明らかにした [14]. Judee らは，人前で話す際の非言語行動に着目し，話者の信頼性と説得力の関係について分析し，非言語行動と話者の信頼性・説得力には関連性が多数確認された [16]．Park らは，話者の説得力を予測するために，複数のモダリティに着目したアプローチを提案している．複数のモダリティを用いることや話者の感情の部分的な事前知識を持つことが，説得力のレベルをより良く予測することに寄与することを示している [17]．Leena らは，欺

瞞検出するために，人間の視覚的，韻律的，言語的，感情的の複数のモダリティに着目し，自動欺瞞検出を行うための機械学習モデルを構築した．マルチモーダル情報を用いることで欺瞞を検出することができることを明らかにし，視覚，発声，感情の 3 つのモダリティが欺瞞検出に貢献することを明らかにした [18].

2.2　褒める行為に関する研究事例

褒める行為に関する研究事例として文献 [1]，[19]，[20]，[21] が挙げられる．褒める行為に関する研究事例は数多く存在する．これらの研究事例では，褒める行為の定義や褒める行為による効果を明らかにしている．また，教育シーンにおける研究事例が数多く存在する．はじめに，褒める行為の定義に関する研究事例について述べる．Jenkins らは，褒める行為とは対象の行動や性格に向けられた称賛を示す言語的・非言語的な行為であると定義している．褒める行為はシンプルかつ効果的な方法であるが，教育者や立場の上の人間が褒める行為をする際には，事前に褒める行為を学ぶことが重要だということを提唱している．また，褒める行為は褒める人のモチベーション向上につながる可能性があることを示唆している．[1]．Kalis らは，褒める行為とは対象の行動の良い側面を示唆する言語的または身体的な行動であると定義している教師の褒める行為の回数を増やすためにをセルフモニタリングの有効性を調査している．セルフモニタリングとは，周囲の状況や他者の行動に基づいて、自己の行動や自己呈示が社会的に適切であるかを観察し，自己の行動をコントロールすることである．その結果，褒める行為の中でも非言語的行動が増えたことが確認された．[19]．Charles らは，教育の現場で子供が適切な行動を取った際に，その行動を褒めることが学級行動の改善に有効であることを明らかにしている．[20]．Brophy は，教育の現場で教師による褒める行為は，生徒のパフォーマンスの向上や強化に繋がることが明らかにしている．教師によって褒める行為は，(1）自然発生的な驚嘆や称賛の表現，(2）過去に行った指摘に対するバランスの維持，(3）代理的な強化（特定の生徒のみを褒め，他の生徒にもこうあるべきと伝える），(4）否定表現の回避（否定表現をせずに褒めることで指導する），(5）緊張を解す手段（褒めることで生徒の緊張を解す），(6）生徒からの誘発，(7）移行の儀式（タスクの終了と次のタスクへ移行する合図），(8）励ましの機能を 8 つに分類できると考えている [21].

第3章　研究課題

本章では，本研究における問題の定義と研究課題について述べる．

3.1　問題の定義

褒める行為は日常生活や社会生活で多くの人が行う重要なコミュニケーションの一つである．また，褒める行為とは対象の行動の良い側面を示唆する言語的または非言語的な行動であると考えられている [19]．対象の行動や性格に向けられた褒めることはシンプルかつ効果的な方法であり，褒める人のモチベーション向上につながる可能性があると考えられている [1]．このことから，褒める行為は日常的の多くのシーンで行っていることが考えられる．しかし，褒めることが苦手な人は，相手を褒めたいと思っていてもどのように振舞うと良いか分からない人が存在すると考えられる．褒める行為の検出を行うことで，褒める行為の振舞いを明らかにできると考えられる．

近年ではオンライン上のコミュニケーションが増えていることから，遠隔対話で褒める行為を行うことも考えられる．対面，遠隔対話での褒める行為の検出を行うことで，褒める行為の振舞いをより分析することができると考えられる．

3.2　研究課題の設定

対面対話と遠隔対話では，言語，非言語行動の伝わる情報が異なることが知られており [8]，褒める行為においても対面対話と遠隔対話で伝わる情報や重要な振舞いが異なることが考えられる．これより，対面，遠隔の両環境で褒める行為を検出できるか検討を行い，対面，遠隔の環境の差が褒める行為の検出に及ぼす影響の調査やそれぞれの環境での重要な振舞いの違いを明らかにする必要がある．

さらに，2.1 節より，話者の特定の行動を予測する際には複数のモダリティを用いることが有効であると考えられている．先行研究においても，複数のモダリティを用いることが褒める行為の検出には有用であることが確認された．このことから，遠隔対話においても複数のモダリティに着目することで褒める行為を検出することができると考えられる．

そこで本研究では，下記 2 点を研究課題とする．

- 言語，非言語行動に着目し，対面，遠隔対話において褒める行為を検出できるか明らかにする．
- 対面対話と遠隔対話における褒める行為の振舞いの違いを明らかにする．

第4章　対話コーパスの構築

本章では，本論文における対話コーパスの構築方法について述べる．本研究では先行研究 [5] で構築した対面，遠隔の対話コーパスを利用した．この対話コーパスは，2 者対話における言語・非言語行動と褒めているシーンかそうでないかの判定値が含まれている．

4.1　対面対話コーパスの構築

4.1.1　2 者対話の収録

2 者対話の参加者は，20 代の大学生 34 名（男性 28 名，女性 6 名）であり，2 名 1 組のペアを 17 組構成した．17 組のうち，初対面が 14 組，顔見知りが 2 組，友人同士が 1 組であった．参加者に対話材料を準備させることを意図して，いままで頑張ってきたことに関するエピソードを 2 つ以上用意してもらった．対話収録時は，図 4.1 のように参加者が互いに向き合って着座した．このとき，参加者間の距離は 180cm とした．対話の収録は，各参加者の様子と 2 者対話全体の様子を撮影するためのビデオカメラ，各参加者の声を録音するためのマイクを用いて行った．

図 4.1: 対面対話における 2 者対話の様子

各組の参加者（参加者 A，参加者 B）は，撮影者の合図に従い，下記の対話 1〜対話 3 を行った．

対話 1　自己紹介 (5 分間)

対話 2　参加者 A が褒める人となり，参加者 B が褒められる人となる対話（5 分間）

対話 3　参加者 B が褒める人となり，参加者 A が褒められる人となる対話（5 分間）

対話 1〜対話 3 を 17 組分，計 255 分間収録した．なお，対話 1（自己紹介）は，各組の多くが初対面であることから，参加者の緊張をほぐす目的で行っているため，分析の対象外とする．対話 2 と対話 3 において，褒める人には，対話相手を積極的に褒めるように指

示した．しかし，一方的に褒めているだけのような不自然な対話にならないようにするために，自由に質問したり，リアクションをしたりすることを許可した．褒められる人には，事前に用意した自分がいままで頑張ってきたことに関連するエピソードを話すように指示した．加えて，対話の自然さや話題の多様性を担保するために，事前に用意していないエピソードについて話すことを許可した．

4.1.2　アノテーション

4.1.1 項で記録した映像データや音声データに対して注釈付けを行うツールである ELAN[22] を使用して，人手で発話シーンを付与した．発話シーンは，沈黙時間が 400 ミリ秒未満の連続した音声区間とした．著者らで対話データをもとに議論を行い，発話の区切りが自然になるように沈黙時間が 400 ミリ秒未満の場合は 1 つの発話とした．次に，ELAN を使用して話者の音声を参照しながら人手で発話内容の書き起こしを行った．

4.1.3　褒める行為の判定

2 者対話の収録に参加していない第三者のアノテータ 5 名が，褒めているシーンかそうでないかの判定を行った．アノテータは文系・理系の教育を受けている 20 代の大学生であり，日常生活に支障のないコミュニケーション能力を有している．アノテータは各発話シーンの映像と音声データを参照し，褒める人の発話シーンごとに対話の相手を褒めているシーンであるか，そうでないかの判定を行った．本研究では，各発話シーンにおいて，褒めていると判定したアノテータが 3 名以上のシーンを Praise シーン（褒める行為が行われているシーン）とした．また，褒めていると判定したアノテータが 3 名未満のシーンを非 Praise シーン（褒める行為が行われていないシーン）とした．Praise シーンの件数は 228 件，非 Praise シーンの件数は 2473 件であった．

4.2　遠隔対話コーパスの構築

4.2.1　2 者対話の収録

オンラインコミュニケーションツールである Zoom[23] を用いて，2 者対話を収録した．参加者は 20 代〜50 代の 40 名（男性 20 名，女性 20 名）であり，一人当たり 3 回異なる相手と対話を行った．合計 60 組の対話が記録された．各ペアの参加者の年齢層・性別は同じであり，初対面である．参加者は，図 4.2 に示すように，PC の前に座り会話をした．対話の収録は，参加者の前に設置された PC のカメラで各参加者の映像を，PC のマイクで各参加者の音声を記録した．

参加者にはいままで頑張ってきたことに関するエピソードを用意したもらった．対話収録時は，各組の参加者（参加者 A，参加者 B）は，撮影者の合図に従い，下記の対話 1〜対話 3 を行った．

図 4.2: 遠隔対話における 2 者対話の様子

対話 1　自己紹介 (5 分間)

対話 2　参加者 A が褒める人となり，参加者 B が褒められる人となる対話（10 分間）

対話 3　参加者 B が褒める人となり，参加者 A が褒められる人となる対話（10 分間）

対話 1〜対話 3 の対話を 60 組分，計 1500 分間収録した．なお，対話 1 の自己紹介は，各組の多くが初対面であることから，参加者の緊張をほぐす目的で行っているため，分析の対象外とする．対話 2 と対話 3 において，褒める人には，対話相手を積極的に褒めるように指示した．しかし，一方的に褒めているだけのような不自然な対話にならないようにするために，自由に質問したり，リアクションをしたりすることを許可した．褒められる人には，事前に用意した自分がいままで頑張ってきたことに関連するエピソードを話すように指示した．加えて，対話の自然さや話題の多様性を担保するために，事前に用意していないエピソードについて話すことを許可した．

4.2.2　アノテーション

4.2.1 項で記録した映像データと音声データに対して，自動音声認識システム*を使用して，自動で発話シーンの付与と発話内容の書き起こしを行った．発話シーンについて，4.2.2 節と同様に沈黙時間が 400 ミリ秒未満の場合を 1 つの発話とした．

*本研究では NTT が開発した自動音声認識システムを利用した．

4.2.3　褒める行為の判定

4.1.2 節と同様の手法で褒めているシーンか，そうでないかの判定を行った．本研究では，各発話シーンにおいて，アノテータ 5 名のうち 3 名以上が褒めていると判定したシーンを Praise シーンとする．また，褒めているシーンを Praise シーンとし，それ以外の発話シーンを非 Praise シーンとした．Praise シーンの件数は 236 件，非 Praise シーンは 16115 件であった．

第5章　推定モデルの構築

本章では，推定モデルの構築について述べる．推定モデルの推論の一連流れを図 5.1 に示す．

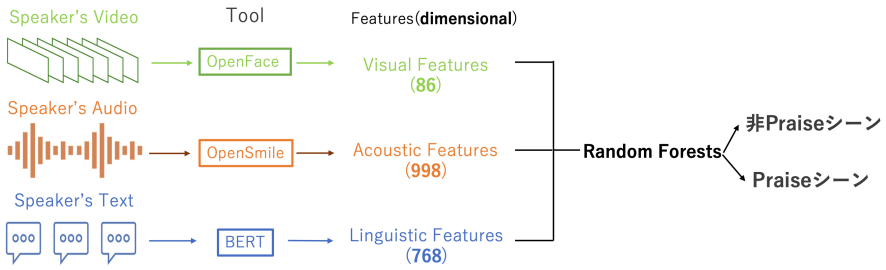

図 5.1: 推論の一連の流れ

はじめに，視覚的，韻律的，言語的特徴量を抽出する．次に，褒める行為を検出する機械学習モデルの構築を行う．

5.1 特徴量抽出

本研究では視覚的，韻律的，言語的特徴量を抽出する．これらの特徴量は，マルチモーダルインタラクション分析の分野でよく用いられている．

5.1.1 視覚的特徴量

各参加者の頭部・顔部の振舞いに関連する特徴量を抽出した．顔画像処理ツールである OpenFace[24] を用いて．褒める人の正面に設置したビデオカメラで撮影したの映像データから，頭部の向きや視線の角度と筋肉群の基本的な行動の単位を表す Action Units[25] に関する特徴量の抽出を行った．頭部の向きに関しては，ビデオカメラ側から顔を見て左から右方向を x 軸，下から上方向を y 軸，手前から奥方向を z 軸とした．Praise シーンの前後 1 秒ずつを含む範囲における頭部の x 軸，y 軸，z 軸回りの回転角度（pose_Rx，pose_Ry，pose_Rz）の分散（_var），中央値（_med），10 パーセンタイル値（_p10），90 パーセンタイル値（_p90）を用いた．視線の角度に関しては，ビデオカメラ側から顔を見て左から右方向を x 軸，下から上方向を y 軸とした．Praise シーンの前後 1 秒ずつを含む範囲における視線の向きの x 軸，y 軸回りの回転角度（gaze_Ax，gaze_Ay）の分散，中央値，10 パーセンタイル値，90 パーセンタイル値を用いた．Action Units に関しては，Praise シーンの前後 1 秒ずつを含む範囲における各 Action Units（表 5.1）の強度の分散，中央値，10 パーセンタイル値，90 パーセンタイル値を用いた．

表 5.1: Action Units の内容

項目	内容	項目	内容
AU01	眉の内側を上げる	AU14	笑窪を作る
AU02	眉の外側を上げる	AU15	唇の両端を下げる
AU04	眉を下げる	AU17	顎を上げる
AU05	上瞼を上げる	AU20	唇の両端を横に引く
AU06	頬を持ち上げる	AU23	唇を硬く閉じる
AU07	瞼を緊張させる	AU25	顎を下げずに唇を開く
AU09	鼻に皺を寄せる	AU26	顎を下げて唇を開く
AU010	上唇を上げる	AU45	瞬きをする
AU012	唇の両端を引き上げる		

5.1.2 韻律的特徴量

音声情報処理ツールである openSMILE[26] を用いて，褒める人が装着したマイクで録音された音声データから，音声の韻律の代表的な特徴量，発話に関する特徴量の抽出を行った．表 5.2 に示した音声の韻律の代表的な特徴量につき，表 5.3 に示す統計量を算出した特徴量と，これらの各特徴量を一次微分したもの（_de）の計 988 の特徴量を抽出した [27][28]．なお，メル周波数ケプストラム係数とは，対数ケプストラム（20 次）の低次成分（1〜12 次）に対して，人間の周波数知覚特性を考慮して重み付けをしたものであり，声道特性を表す特徴量である．また，ゼロ交差率とは，音の波形が音圧が 0 である軸を交差する割合を表す特徴量である．

表 5.2: 音声特徴量の内容の内容

特徴量	内容
pcm_intensity_sma	正規化された強度の値
pcm_loudness_sma	正規化された強度に 0.3 乗した値
mfcc_sma[1]–[12]	1〜12 次のメル周波数ケプストラム係数
lspFreq_sma[0]–[7]	8 つの LPC 係数から計算される周波数
pcm_zcr_sma	ゼロ交差率
voiceProb_sma	声である確率
F0_sma	基本周波数
F0env_sma	基本周波数のエンペローブ

表 5.3: 抽出可能な統計量

統計量	内容	統計量	内容
_max	最大値	_linregc1	線形近似の勾配
_min	最小値	_linregc2	線形近似のオフセット
_range	最大値と最小値の差	_linregerrQ	線形近似の二乗誤差
_maxPos	最大値の絶対位置	_linregerrA	線形近似と実際の値の誤差の差
_minPos	最小値の絶対値位置	_skewness	歪度
_amean	平均値	_kurtosis	尖度
_stddev	標準偏差	_iqr1-2	四分位範囲:quartile2-quartile1
_quartile1	25 パーセンタイル	_iqr2-3	四分位範囲:quartile3-quartile2
_quartile2	50 パーセンタイル	_iqr1-3	四分位範囲:quartile3-quartile1
_quartile3	70 パーセンタイル		

5.1.3 言語的特徴量

言語特徴量の抽出方法について，発話内容を単語ごとに分析する方法と発話内容をベクトル化する方法を用いた．発話内容をベクトル化する方法として，日本語事前学習済みの BERT モデル [29] を利用し，褒める人の発話を 768 次元のベクトルに変換し，特徴量として抽出した．

5.2 モデル構築

褒める行為を検出するために，対面および遠隔対話のそれぞれの環境で分類モデルの構築を行った．具体的には，説明変数を 5.1 節で抽出した特徴量，目的変数を 4.1.3 項や 4.2.3 項で判定した Praise シーン/非 Praise シーンとし，対面および遠隔対話ごとに褒める行為が行われているか否かを判定する機械学習モデルを構築した．分類モデルの構築には，少ないデータ数の場合でも性能が良くなることが見込まれる Random Forests[30] を用いて行い，ハイパーパラメータは Hyperopt[31] を用いてチューニングを行った．加えて，特徴量の取りうる値の範囲を揃える為に，各特徴量を平均値 0，分散 1 になるように正規化を行った．

本研究では，どのモダリティが褒める行為を検出するために寄与するのか明らかにするために，各モダリティを組み合わせた機械学習モデルの構築を行った．各モダリティを組み合わせた機械学習モデルについて，次の（1）〜（3）を 100 回ずつ行った．

(1) Praise シーンと非 Praise シーンの件数を揃えるために，シーンごとに無作為に 228 件[†]を選択し，データセットとする．

[†]対面対話における Praise シーンが 228 件のため，今回は Praise シーンと非 Praise シーンの件数を 228 件に揃えることとした．

(2) データセットを訓練データとテストデータを 9:1 に無作為に分ける.

(3) 訓練データで構築したモデルを用いてテストデータにおける目的変数を推定する.

.

第6章　構築した推定モデルの分析

本章では，構築した推定モデルの分析・考察を述べる

6.1 対面，遠隔対話における褒める行為の検出の推定性能について

6.1.1 対面対話における褒める行為の検出の推定結果

構築した機械学習モデルの推定結果を表 4 に示す．ベースラインは，データセットにおける各群 (褒めているシーン，褒めていないシーン) の割合に合わせて，褒めているシーンかそうでないかを出力するモデルを用いた．表 4 より，ベースラインの推定性能は F 値が 0.523 であった．本稿で提案する対面対話における機械学習モデルで最も推定性能の良かったモデルは Model G_F であり，推定性能の F 値が 0.816 であった．このモデルの構築に用いたモダリティは視覚，韻律，言語の 3 つであった．ベースラインの推定性能と比較した場合，Model G_F の推定性能の方が良いことがわかった．単体のモダリティ（A_F〜C_F）の推定性能と比較した場合でも Model G_F の推定性能の方が良いことがわかった．このとこから，単体のモダリティを用いるよりも複数のモダリティを組み合わせることが推定性能の向上につながることが明らかになった．

表 6.1: 対面対話において各モデルで利用した特徴量と推定性能の各指標の平均値 (N=100)

	視覚	韻律	言語	適合率	再現率	F 値
Baseline				0.543	0.537	0.523
Model A_F	✓			0.773	0.764	0.752
Model B_F		✓		0.657	0.641	0.638
Model C_F			✓	0.824	0.796	0.793
Model D_F	✓	✓		0.781	0.754	0.741
Model E_F	✓		✓	0.814	0.805	0.792
Model F_F		✓	✓	0.797	0.787	0.783
Model G_F	✓	✓	✓	**0.841**	**0.821**	**0.816**

6.1.2 遠隔対話における褒める行為の検出の推定結果

5.2 節で構築した機械学習モデルの推定結果を表 6.2 に示す．データセットにおける各群（褒めているシーン，褒めていないシーン）の割合に合わせて，褒めているシーンかそうでないかを出力するモデルを用いた．ベースラインの推定性能は F 値が 0.493 であった．本稿で提案する遠隔対話における機械学習モデルで最も推定性能の良かったモデルは Model G_R であり，推定性能の F 値が 0.867 であった．このモデルの構築に用いたモダリティは視覚，韻律，言語の 3 つであった．ベースラインの推定性能と比較した場合，Model

G_R の推定性能の方が良いことがわかった．対面対話でのモデルの推定性能と同様に，単体のモダリティ（A_R〜C_R）の推定性能と比較した場合でも Model G_R の推定性能の方が良いことがわかった．このことから，単体のモダリティを用いるよりも複数のモダリティを組み合わせることが推定性能向上につながることが明らかになった．

表 6.2: 遠隔対話における各モデルで利用した特徴量と推定性能の各指標の平均値 (N=100)

	視覚	韻律	言語	適合率	再現率	F 値
Baseline				0.501	0.493	0.493
Model A_R	✓			0.684	0.672	0.668
Model B_R		✓		0.808	0.795	0.793
Model C_R			✓	0.823	0.819	0.819
Model D_R	✓	✓		0.800	0.786	0.784
Model E_R	✓		✓	0.859	0.846	0.845
Model F_R		✓	✓	0.875	0.865	0.865
Model G_R	✓	✓	✓	**0.876**	**0.871**	**0.867**

6.2　対面，遠隔対話における褒める行為の推定結果の比較

表 6.1，6.2 で示した対面，遠隔対話の推定性能に対して，各モデル間で比較を行った．具体的には，各モデル間の F 値に対して 5%水準で対応のない t 検定を行い，その後ホルム補正を行った．これより得られた p 値を表 6.3，6.4 に示す．

表 6.3: 対面対話における検定結果の p 値

	Model A_F	Model B_F	Model C_F	Model D_F	Model E_F	Model F_F	Model G_F
Model A_F	-	**2.E-27***	**9.51E-45***	**7.5E-32***	**2.12E-47***	**1.62E-48***	**2.88E-48***
Model B_F	**2.E-27***	-	**1.E-8***	**0.011***	**1.07E-10***	**1.02E-10***	**0.0001***
Model C_F	**9.51E-4***	**1.E-8***	-	**5.38E-5***	0.55	0.137	**0.027***
Model D_F	**7.5E-32***	**0.011***	**5.38E-5***	-	**1.72E-6***	**2.31E-6***	**9.57E-8***
Model E_F	**2.12E-47***	**1.07E-10***	0.55	**1.72E-6***	-	**0.044***	**0.007***
Model F_F	**1.62E-48***	**1.02E-10***	0.137	**2.31E-6***	**0.044***	-	0.275
Model G_F	**2.88E-48***	**0.0001***	**0.027***	**9.57E-8***	**0.007***	0.275	-

6.3　考察

6.3.1　モデルの性能についての考察

対面，遠隔対話においてそれぞれの最も性能の良かったモデルの考察を述べる．対面，遠隔対話それぞれで，単体のモダリティを用いるより複数のモダリティを用いた方が検出

表 6.4: 遠隔対話における検定結果の p 値

	Model A_R	Model B_R	Model C_R	Model D_R	Model E_R	Model F_R	Model G_R
Model A_R	-	**7.E-21***	**6.24E-38***	**4.5E-20***	**5.94E-48***	**5.37E-31***	**1.23E-46***
Model B_R	**7.E-21***	-	**2.E-7***	**0.012***	**8.34E-9***	**1.63E-10***	**1.4E-4***
Model C_R	**6.24E-38***	**2.E-7***	-	**3.38E-4***	0.109	0.23	**0.022***
Model D_R	**4.5E-20***	**0.012***	**3.38E-4***	-	**1.05E-7***	**4.55E-7***	**2.87E-8***
Model E_R	**5.94E-48***	**8.34E-9***	0.109	**1.05E-7***	-	0.798	0.151
Model F_R	**5.37E-31***	**1.63E-10***	0.23	**4.55E-7***	0.798	-	0.061
Model G_R	**1.23E-46***	**1.4E-4***	**0.022***	**2.87E-8***	0.151	0.061	-

性能が向上することがわかった．このことから．遠隔，対面対話ともに人が相手を褒める際には視覚，韻律，言語のモダリティに変化が現れると考えられる．

6.3.2　対面，遠隔対話における振舞いの傾向についての考察

対面，遠隔対話のそれぞれのモデルについて比較を行った結果についての考察を述べる．対面対話では表 6.3 のモデル E_F・F_F・G_F の各性能間に有意差が認められたが，遠隔対話では表 6.4 のモデル E_R・F_R・G_R の各性能間に有意差が認められなかった．このことから，対面，遠隔対話での視覚・韻律モダリティの特徴量の傾向に差があると考えられる．傾向に差が生じた要因は対面，遠隔対話において各参加者の視覚．韻律的振舞いに差があった可能性が考えられる．対面対話では対話相手が目の前にいるため，視覚，韻律的な振舞いに配慮して対話した参加者が多かったと考えられる．遠隔対話では対話相手が PC モニタ内にいることから，視覚，韻律的な振舞いに配慮する参加者が少なかったと考えられる．褒める行為を行う際には対話をする環境によって話者の振舞いが異なると考えられる．

6.3.3　重要度の高い特徴量についの考察

対面，遠隔対話における重要度の高かった特徴量についての考察を述べる．対面，遠隔対話においてそれぞれ最も性能の良かった Model G_F，G_R における，重要度の高い特徴量上位 30 件を対面，遠隔対話ともに，図 6.3，6.4 に示す．図 6.3，6.4 において．橙色は韻律に関する特徴量．青色は言語に関する特徴量を表している．韻律，言語の特徴量が上位 30 件を占めており，視覚的特徴量が上位に現れなかった．この要因として，各特徴量の次元数に差があったためと考えられる．また，対面対話では上位に韻律的特徴量が言語的特徴量より多く現れているが，遠隔対話では言語的特徴量が韻律的特徴量より多く現れている．現状ではなぜ対面と遠隔対話で重要度の高い特徴量の傾向に差があるのか明らかになっていないため．今後の課題として調査が必要であると考えられる．

第7章　結論

人間の振舞いと褒める行為の判定値が含まれている対話コーパスを用いて，対面，遠隔対話において言語，非言語行動から褒める行為を検出できるか明らかにする取り組みを行った．加えて，対面，遠隔対話での褒める行為を検出するモデルの比較を行った．遠隔対話において，話者の視覚，韻律，言語に関する特徴量を抽出し，褒める行為を検出する機械学習モデルを構築した．その結果，視覚的，韻律的，言語的特徴量を用いることで褒める行為を検出できることが明らかになったまた，対面，遠隔対話におけて褒める行為特有の振舞いの共通点，相違点を明らかにするために対面，遠隔対話での推定性能の比較を行った．その結果，対面対話では，モデル E_F（視覚，言語）・F_F（視覚，韻律）・G_F（視覚，韻律，言語）の各性能間に有意差が認められたが，遠隔対話では，モデル E_R（視覚，言語）・F_R（視覚，韻律）・G_R（視覚，韻律，言語）の各性能間に有意差が認められなかった．このことから，対面，遠隔対話での視覚的・韻律的特徴量に関する振舞いの傾向に差があると示唆された．最後に，褒める行為の検出を行うためにどの特徴量が重要となるか分析した．その結果，対面，遠隔対話でモダリティの重要度の傾向に差がある可能性が示唆された．しかし，なぜ差があるのかは現状では明らかにできていないため，今後の課題として調査を続けていきたい．本稿では，対面，遠隔対話において褒めている発話に着目した．しかし，褒める行為では，褒められる人の発話も重要であると考えられている [32]．今後の研究方針として，褒められる人の情報に着目し，推定性能を向上させることが挙げられる．また，今回の研究で用いたデータセットに関して，データ数が少ないという問題がある．この問題を解決するために新たなデータセットを作成し，データ量を増やすことで機械学習モデルの推定性能の向上を図りたい．

謝辞

本研究は，日本電信電話株式会社 NTT 人間情報研究所との共同研究の成果である．

参考文献

[1] Jenkins, L.N., Floress, M.T. and Reinke, W.: Rates and Types of Teacher Praise: A Review and Future Directions, Psychology in the Schools, Vol.52, No.5, pp.463–476 (2015) .

[2] 大西俊輝，山内愛里沙，大串旭，石井亮，青野裕司，宮田章裕：褒める行為における頭部・顔部の振る舞いの分析，情報処理学会論文誌，Vol.62，No.9，pp.1620–1628（2021）.

[3] Onishi, T., Yamauchi, A., Ishii, R., Aono, Y. and Miyata, A.: Analyzing Nonverbal Behaviors along with Praising, Proc. 22nd ACM International Conference on Multimodal Interaction (ICMI '20), pp.609–613 (2020).

[4] Onishi, T., Yamauchi, A., Ogushi, A., Ishii, R., Fukayama, A., Nakamura, T., Miyata, A.: Modeling Japanese Praising Behavior by Analyzing Audio and Visual Behaviors. Frontiers in Computer Science, Vol.4 (2022).

[5] Onishi, T., Yamauchi, A., Ogushi, A., Ishii, R., Fukayama, A., Nakamura, T. and Miyata, A.: A Comparison of Praising Skills in Face-to-Face and Remote Dialogues, Proc. 13th Conference on Language Resources and Evaluation (LREC '22), pp.5805–5812

[6] Ogushi, A., Onishi, T., Tahara, Y., Ishii, R., Fukayama, A., Nakamura, T., Miyata, A.: Analysis of praising skills focusing on utterance contents, Proc. 23rd Annual Conference of the International Speech Communication Association (INTERSPEECH '22), pp.2643–2747 (2022).

[7] 大串旭，大西俊輝，田原陽平，石井亮，深山篤，中村高雄，宮田章裕：言語特徴に着目した褒め方の上手さの推定モデルの検討．グループウェアとネットワークサービスワークショップ 2021，Vol.2021，pp.1–8（2021）.

[8] Doherty, S.G., Anderson, A., O'malley, C., Langton, S., Garrod, S., Bruce, V.: Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance, Journal of experimental psychology, VoL.3, No.2, pp.105–125 (1997).

[9] Okada, S., Ohtake, Y., Nakano, Y.I., Hayashi, Y., Huang, H.H., Takase, Y. and Nitta, K.: Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets, Proc. 18th ACM International Conference on Multimodal Interaction (ICMI '16) , pp.169–176 (2016) .

[10] Wörtwein, T., Chollet, M., Schauerte, B., Morency, L.P., Stiefelhagen, R. and Scherer, S.: Multimodal Public Speaking Performance Assessment, Proc. 17th ACM on International Conference on Multimodal Interaction (ICMI '15), pp.43–50 (2015).

[11] Oya, A. and Daniel, G.P.: One of a kind: inferring personality impressions in meetings, Proc. 15th ACM on International conference on multimodal interaction (ICMI '13) , pp.11–18 (2013).

[12] Yagi, Y., Okada, S., Shiobara, S. and Sugimura, S.: Predicting multimodal presentation skills based on instance weighting domain adaptation, Journal on Multimodal User Interfaces, Vol.16, No.1, pp.1–16 (2021).

[13] Ishii, R., Otsuka, K., Kumano, S., Higashinaka, R. and Tomita, J.: Analyzing Gaze Behavior and Dialogue Act during Turn-taking for Estimating Empathy Skill Level, Proc. 20th ACM International Conference on Multimodal Interaction (ICMI '18) , pp.31–39 (2018).

[14] Soleymani, M., Stefanov, K., Kang, H.S., Ondras, J. and Gratch, J.: Multimodal Analysis and Estimation of Intimate Self-Disclosure, Proc. 21st ACM International Conference on Multimodal Interaction (ICMI '19) , pp.59–68 (2019).

[15] Chen, L., Feng, G., Joe, J., Leong, C.W., Kitchen, C. and Lee, C.M.: Towards Automated Assessment of Public Speaking Skills Using Multimodal Cues. Proc. 16th International Conference on Multimodal Interaction (ICMI '14), pp.200–203 (2014).

[16] Judee, K., Burgoon, T.B. and Michale, P.: Nonverbal Behaviors, Persuasion, and Credibility, Human Commnication Reserch, Vol.17, No.1, pp.140–169 (1990).

[17] Park, S., Shim, H.S., Chatterjee, M., Sagae, K. and Morency, L.P.: Computational Analysis of Persuasiveness in Social Multimedia: A Novel Dataset and Multimodal Prediction Approach, Proc. 16th International Conference on Multimodal Interaction (ICMI '14) , pp.50–57 (2014).

[18] Leena, M. and Maja, J.M.: Introducing Representations of Facial Affect in Automated Multimodal Deception Detection, Proc. 2020 International Conference on Multimodal Interaction(ICMI '20), pp.305–314 (2020).

[19] Kalis, T.M., Vannest, K.J. and Parker, R.: Praise Counts: Using Self-Monitoring to Increase Effective Teaching Practices, Preventing School Failure: Alternative Education for Children and Youth, Vol.51, No.3, pp.20–27 (2007).

[20] Charles, H.M., Wesley, C.B. and Don, R.T.: Rules, Praise, and Ignoring: Elements of Elementary Classroom Control, Applied Behavior Analysis, Vol.1, no.2, pp.139–150 (1968).

[21] Brophy, J.: Teacher praise: A functional analysis, Review of Educational Research, Vol.51, No.1, pp.5–32 (1981).

[22] Brugman, H. and Russel, A. : Annotating Multimedia / Multi-modal resources with ELAN, Proc. 4th International Conference on Language Resources and Language Evaluation (LREC '04), pp.2065–2068 (2004).

[23] Zoom Cloud Meetings, Zoom Video Communications Inc., available from 〈https://zoom.us/〉 (accessed 2022-11-01).

[24] Baltrusaitis, T., Robinson, P. and Morency, L.P.: OpenFace: An open source facial behavior analysis toolkit, IEEE Winter Conference on Applications of Computer Vision (WACV '16), pp.1–10 (2016).

[25] Ekman, P. and Friesen, W.: Manual for the facial action coding system, Palo Alto: Consulting Psychologists Press (1977).

[26] Eyben, F., Wöllmer, M., and Schuller,B.: openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor, Proc. ACM Multimedia, pp.1459–1462 (2010).

[27] Schuller, B., Steidl, S. and Batliner, A.: The interspeech 2009 emotion challenge (2009).

[28] Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Müller, C. and Narayanan, S.,: The INTERSPEECH 2010 paralinguistic challenge, Proc. INTERSPEECH '10, pp.2794–2797 (2010).

[29] Devlin, J., Chang, M., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT '19) , pp.4171–4186 (2019).

[30] Breiman, L.: Random Forests, Machine Learning, Vol.45, No.1, pp.5–32 (2001).

[31] Bergstra, J.and Yamins, D. and Cox, D.: Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms, Proc. 12th Python in Science Conferences (SciPy '13), pp.13–20 (2013).

[32] Henderlong, J. and Lepper, M.: The effects of praise on children's intrinsic motivation: A review and synthesis, Psychological Bulletin, Vol.128, No.5, pp.774–795 (2002).

研究業績

査読付き国際会議

(1) Asahi Ogushi, Toshiki Onishi, Yohei Tahara, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata: Analysis of Praising Skills Focusing on Utterance Contents. : Proc. 23rd Annual Conference of the International Speech Communication Association (Interspeech '22), pp.2743–2747 (2022 年 9 月).

(2) Toshiki Onishi, Asahi Ogushi, Yohei Tahara, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata: A Comparison of Praising Skills in Face-to-Face and Remote Dialogues. Proc. 13th Language Resources and Evaluation Conference (LREC '22), pp.5805–5812 (2022 年 6 月).

査読付き国内会議

(1) 田原陽平, 大西俊輝, 大串旭, 石井亮, 深山篤, 中村高雄, 宮田章裕: 対面・遠隔対話からの称賛行為検出の基礎検討. 情報処理学会グループウェアとネットワークサービスワークショップ 2022 論文集 (2022 年 11 月掲載予定)

(2) 大串旭, 大西俊輝, 田原陽平, 石井亮, 深山篤, 中村高雄, 宮田章裕: 言語特徴に着目した褒め方の上手さの推定モデルの検討. 情報処理学会グループウェアとネットワークサービスワークショップ 2021 論文集, Vol.2021, pp.1–8 (2021 年 11 月).

研究会・シンポジウム

(1) 田原陽平, 大西俊輝, 大串旭, 石井亮, 深山篤, 中村高雄, 宮田章裕: マルチモーダル情報に基づく褒める行為の判定の基礎検討. 情報処理学会シンポジウム論文集, マルチメディア, 分散, 協調とモバイル (DICOMO '22), Vol.2022, pp.576–581 (2022 年 7 月).

受賞

(1) 情報処理学会グループウェアとネットワークサービスワークショップ 2022 ベストペーパー賞: 対面・遠隔対話からの称賛行為検出の基礎検討, 受賞者: 田原陽平, 大西俊輝, 大串旭, 石井亮, 深山篤, 中村高雄, 宮田章裕 (2022 年 11 月).

(2) 情報処理学会マルチメディア、分散、協調とモバイルシンポジウム (DICOMO2022) 優秀論文賞, マルチモーダル情報に基づく褒める行為の検出の基礎検討, 受賞者: 田原陽平 (2022 年 7 月).

(3) 情報処理学会グループウェアとネットワークサービスワークショップ 2021 ベストペーパー賞: 言語特徴に着目した褒め方の上手さの推定モデルの検討, 受賞者: 大串旭, 大西俊輝, 田原陽平, 石井亮, 深山篤, 中村高雄, 宮田章裕 (2021 年 11 月).