対面・遠隔対話における
上手い褒め方の振舞いの分析

令和5年度　修士論文

日本大学大学院　総合基礎科学研究科
地球情報数理科学専攻　宮田研究室

6122M09　大串　旭

概要

多くのシーンで利用されるコミュニケーション手段として褒める行為がある．褒める行為は，言語的・非言語的な行動を用いて褒める相手の行動や所持品などに向けられた称賛，承認を表現するポジティブなフィードバックとされている．褒める行為の効果として，褒める相手の行動やモチベーションを強化することや，褒める相手との関係性を良好にすることが明らかにされている．このような効果が見込める褒める行為は事前に準備を必要とせず，どのような場所でも行うことが可能であり，多くの人が日常的に褒める行為を実行している．だが，褒める際にどのように振る舞えばよいかわからない人も多く存在している．加えて，心理学や教育学の分野では，褒める行為の機能的な側面や褒める行為の効果が明らかにされているが，どのように振舞えば上手く褒められていると褒められる人が感じられるのか明らかにされていない．さらに，近年普及している遠隔対話における褒める行為と対面対話における褒める行為に相違点があるか明らかにされていない．そこで，本研究では実際に対話相手を褒めている際のデータを用いて，対話相手を上手く褒めるために重要な言語的・非言語的行動を明らかにする取り組みを行う．本論文では，対面・遠隔対話中の褒める人と褒められる人の視覚的（頭部や顔部，視線の動き），韻律的（話し方や声のピッチなど），言語的（発話内容）な行動に着目して対話相手を上手く褒めるために重要な言語的・非言語的行動を明らかにする取り組みを行う．加えて，本論文では，対面対話と遠隔対話との間で上手く褒めるために重要な振舞いに違いがあるのか明らかにする取り組みを行う．具体的には，まず，対面・遠隔対話における言語的・非言語的行動と褒め方の上手さの評価値を含む対話コーパスを用意した．次に，言語・非言語行動に関連する特徴量を用いて褒め方の上手さの評価値を推定する機械学習モデルを構築し，どのような振る舞いが重要であるかを検証した．その結果，上手く褒めるために特に重要であった行動は対面対話においては言語的行動であり，遠隔対話においては韻律的行動であった．

本論文の貢献は，下記の通りである．

- 対面・遠隔対話において言語的・非言語的行動に着目して，対話相手を上手く褒めるために重要な振舞いを明らかにした．
- 対面・遠隔対話において対話相手を上手く褒めるために重要な振舞いに違いがあることを明らかにした．

目 次

第 1 章　序論 **1**
1.1　研究の背景 2
1.2　研究の目的 2
1.3　本論文の構成 3

第 2 章　関連研究 **4**
2.1　褒める行為に関する研究事例 5
2.2　人の個人特性やコミュニケーションスキルを推定する研究 7
2.2.1　人の個人特性を推定する研究 8
2.2.2　人のコミュニケーションスキルを推定する研究 9

第 3 章　研究課題 **13**
3.1　問題の定義 14
3.2　研究課題の設定 15

第 4 章　対話コーパス **16**
4.1　対面対話コーパス 17
4.1.1　対面対話の収録方法 17
4.1.2　アノテーション 18
4.2　遠隔対話コーパス 20
4.2.1　遠隔対話の収録 20
4.2.2　アノテーション 22

第 5 章　対面対話コーパスにおける上手く褒めるための振る舞いの分析 **24**
5.1　特徴量の抽出 25
5.1.1　視覚的特徴量 25
5.1.2　韻律的特徴量 26
5.1.3　言語的特徴量 27
5.2　褒め方の上手さ予測モデル構築 28
5.3　対面対話の分析 30

第 6 章　遠隔対話コーパスにおける上手く褒めるための振る舞いの分析 **35**
6.1　特徴量の抽出 36

6.1.1 視覚的特徴量 36
6.1.2 韻律的特徴量 37
6.1.3 言語的特徴量 37
6.2 褒め方の上手さ予測モデル構築 37
6.3 遠隔対話の分析 38

第 7 章 対面対話と遠隔対話の振舞いの比較 42
7.1 機械学習モデルの性能 43
7.2 振舞いの比較 44

第 8 章 結論 47

参考文献 50

研究業績 56

図 目 次

4.1 2 者による対面対話の様子 18
4.2 2 者による遠隔対話の様子 21

5.1 モデルの概要 25
5.2 視覚的特徴量の抽出範囲 27
5.3 韻律的特徴量の抽出範囲 28
5.4 言語的特徴量の抽出範囲 29
5.5 BERT を用いた文章ベクトルの抽出 29
5.6 M29 における重要度の高い特徴量上位 20 件 34

6.1 モデルの概要 36
6.2 M15 における重要度の高い特徴量上位 20 件 41

7.1 対面・遠隔対話コーパスの褒める人の発話内容から抽出した言語的特徴量を UMAP を用いて 2 次元に削減した結果 45
7.2 対面・遠隔対話における Praise スコア高群と低群の発話内容の単語数 . . . 46

表 目 次

2.1 本研究が属する研究領域 5

3.1 褒める行為を工学的に分析している先行研究と本論文との差分 15

4.1 対面対話における発話シーンのシーン数，平均値，最大値，最小値 19
4.2 収録機材のスペック 22
4.3 発話シーンと Praise シーンの情報 23

5.1 Action Units の内容 26
5.2 韻律的特徴量の内容 27
5.3 抽出可能な統計量 28
5.4 対面対話コーパスを用いた機械学習モデルの性能 33

6.1 遠隔対話コーパスを用いた機械学習モデルの性能 40

7.1 対面対話コーパスにおいて最も性能の良かったモデル (M29) の confusion matrix 43
7.2 遠隔対話コーパスにおいて最も性能の良かったモデル (M15) の confusion matrix 43
7.3 対面・遠隔対話コーパスにおける各言語的特徴量と Praise スコアとの間に相関関係が確認できた個数 45

第1章　序論

1.1 研究の背景

日常生活や社会活動において褒める行為は多くの人が行う大切なコミュニケーションの 1 つである．褒める行為とは対象の行動や性格に向けられたものであること [1]，送り手から受け手への一方的な意思の伝達ではなく複雑なコミュニケーションであると考えられている [2]．褒めることにより，相手との関係性を円滑にすることや，相手の能力を最大限発揮してもらうなどの効果が期待されている [3][4]．そのため，多くの人が日常的に褒める行為を行っていると考えられる．その一方で褒め方に自信が無い人や，より上手く褒められるようになりたいと考えている人が存在している．そういった人が褒め方の上手さを向上させるためには，自分自身の褒め方の上手さを把握することが必要であると考えられるが，自分自身の褒め方の上手さを把握することは困難である．実際に自分自身の褒め方の上手さを把握するためには，家族や友人といった他者からの評価を聞くことや専門家の指導を仰がなくてはならない．しかし，他者から評価を聞くためには時間や場所，多額のコストといった問題やその評価が定性的になってしまうという問題が存在している．手軽に気兼ねなく自分自身の褒め方の上手さを把握することや誰にとっても同一の基準で褒め方の上手さを判断することは困難である．そのため，上手い褒め方のモデリングが必要であり，褒める行為における言語・非言語行動をどのように用いれば良いのかわからないという問題がある．また，COVID-19 の影響により日常的なコミュニケーション手段として遠隔対話を行うことが増加した [5]．これにより，褒める行為を遠隔対話中に行うことが増加した．しかし，対面対話と遠隔対話では言語・非言語行動の伝達のされ方に差異があるとされている [6][7]，対面対話における上手く褒めるための振舞いは，遠隔対話における振舞いと相違点がある可能性が考えられる．そのため，対面対話と遠隔対話における上手く褒めるための振舞いに相違点があるのかわからないという問題がある．

1.2 研究の目的

1.1 節で述べたように，褒める行為は日常的に行われているコミュニケーションの 1 つである．しかし，褒める行為を行う際に言語・非言語行動をどのように用いることが上手い褒め方となるのか定量的に明らかにされていない．このため，どのような人でも同一の基準で褒め方の上手さを判断することや改善することが困難であると考えられる．褒め方の上手さを定量的に分析することで，いかなる条件でも褒め方の上手さを判断できる同一の基準を発見できると考えた．そのため，多くの人にとって上手く褒められていると感じるための振舞いを定量的に明らかにすることが重要であると考えた．心理学や教育学，社会学の分野では，褒める行為に関する研究が多く行われている [4][8][9][10]．だが，これらの研究では，褒める行為の機能や効果に着目している．さらに，分析手法はアンケート結果の集計や主観による評価である．これらから明らかになった知見では，人により解釈が変化してしまう可能性がある．そこで，上手く褒めるためにどのように振舞えばよいのか客観的な指標が必要であると考えた．しかし，これまで上手く褒めるためにどのうように振舞えば良いのか客観的な指標を用いた分析は行われていない．さらに，近年増加した遠

隔対話と対面対話における上手く褒めるの振舞いの相違点を明らかにする必要があると考えた．本研究では，対面・遠隔対話における褒める行為の言語・非言語行動を工学的に分析することで，上手く褒めるためには言語・非言語行動をどのように用いれば良いのか明らかにすることを目指す．本論文では，上手く褒めるために言語的・非言語的行動をどのように用いればよいのか明らかにすることと対面・遠隔対話における上手く褒めるための振舞いの相違点を明らかにすることを目的とする．上記のことを明らかにするために，話者（褒める人・褒められる人）の言語・非言語行動に着目し，相手を上手く褒めるために重要な行動を調査・分析する取り組みを行う．まず，対面・遠隔対話における話者の言語・非言語行動と褒め方の上手さの評価値を含む対話コーパスを用意・作成する．次に，言語・非言語行動に関する特徴量を用いて褒め方の上手さの評価値を予測する機械学習モデルを構築し，上手く褒めるために重要な言語・非言語行動の振舞いを検証する．最後に，対面・遠隔対話における上手く褒めるために重要な振舞いの相違点を検証する．

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，対話中の振る舞いを用いて人間の行動・能力・感情を分析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設定と研究課題について述べる．

4 章では，2 者対話における褒め方の上手さの評価と各モダリティの記録を含む対話コーパスの構築について述べる．

5 章では，対面対話コーパスを用いて上手く褒めるための振る舞いの分析について述べる．

6 章では，遠隔対話コーパスを用いて上手く褒めるための振る舞いの分析について述べる．

最後に 8 章にて，本論文の結論を述べる．

第2章　関連研究

本研究は，対話中における話者の言語・非言語行動をもとに，褒め方の上手さを分析する取り組みである．本章では，褒める行為に関する研究事例と，人の言語・非言語行動をもとに，話者の感情や個人の能力を推定する研究事例について述べる．本研究が属する研究領域を表 2.1 に示す．2.1 節では，褒める行為に関する研究事例について紹介する．これらは，褒める行為について研究をしているという点で本研究と関係している．2.2 節では，人の言語・非言語行動を用いて，人の個人特性やコミュニケーションスキルを推定する研究事例について紹介する．これらは，対話中の話者の言語・非言語行動を用いてコミュニケーションスキルを推定するという点で本研究と関係している．

表 2.1: 本研究が属する研究領域

研究分野	研究事例
褒める行為に関する研究	・褒める行為の機能に関する研究 ・褒める行為の効果に関する研究
人の個人特性や コミュニケーションスキルを推定する研究	・人の言語・非言語行動を用いて 個人特性を推定する研究 ・人の言語・非言語行動を用いて コミュニケーションスキルを推定する研究

2.1　褒める行為に関する研究事例

本節では褒める行為に関する研究事例について説明する．褒める行為は，主に教育学や心理学の分野で研究されている．褒める行為の機能や褒める行為による効果を明らかにする研究事例が多く存在している．特に，未就学児や学生を対象とした研究事例が多く存在している．

まず，褒める行為の機能に関する研究事例を紹介する．Brophy[4] は．教育現場における教師の褒める行為は，学生の行動を強化し，自尊心を高め，親密な関係を築くのに推奨される行為であると述べている．教示が褒める行為の機能として，(1) 自発的な驚きや感嘆の表現，(2) 過去の批判や期待に対する結果，(3) 学生の行動を変化させたりコントロールする，(4) 否定語を避けるための肯定表現，(5) コミュニケーションを確立するためのきっかけ，(6) 学生からの誘発，(7) 移行儀式としての褒め言葉，(8) 慰めや励ましの 8 つの機能があるとした．このような褒める行為は，特定の行動を対象としたポジティブなフィードバックであると考えられており，一般的な褒める行為に比べてより有用であることを示した．

Floress ら [11] は，教師が学生の行動を促進させるために褒める行為が効果的であると述べている．学生の行動を促進させるための褒める行為の特徴として．下記の特性を挙げている．

- 行動特定褒め（BSP）：BSP は，学生の特定の行動に対して具体的に褒めることで

ある．BSP は，学生にとってより意味のあるフィードバックを提供し、学習意欲を高めることできると述べられている．

- 条件付き褒め：条件付き褒めは，学生が望ましい行動を示した場合にのみ褒めることである．条件付き褒めは，学生に望ましい行動を促すために使用される．
- 真摯な褒め：真摯な賞賛は，教師が本当に感謝していることを示す際の褒めることである．真摯な褒めは，学生に対する教師の関心や尊重を示し，学生の自尊心を高めることが可能であると述べられている．
- 学術的なスキルやタスクに向けられた褒め：学術的なスキルやタスクに向けられた褒めは、学生が学術的なスキルやタスクを達成した場合に与えられる褒めである．この種の褒めは，学生の学習意欲を高め，学術的な成果を促進することが可能である．
- 即時性：褒めは，学生が望ましい行動を示した直後に与えられることが望ましい．即時性のある褒めは，学生にとってより意味のあるフィードバックを提供し，学習意欲を高めることができる．
- 頻度：褒めは，学生が望ましい行動を示した場合にのみ与えられるべきである．過剰な褒めは，学生にとって意味のないものになる可能性がある．
- 真実性：褒めは，教師が本当に感謝していることを示すことが重要である．真実性のある褒めは，学生に対する教師の関心や尊重を示し，学生の自尊心を高めることができる．

Floress と Beschta[12] は．一般的な褒める行為には 8 つのカテゴリーに定義しており，教師が褒める行為の各カテゴリーの割合の分析を行っている．具体的には，(1) 行っているまたは完了した働きに対して褒める，(2）単一の肯定的な形容詞を用いて褒める，(3）相手が努力していることを強調するために褒める，(4）相手が行ったことへの感謝を伝える，(5）言語・非言語行動ジェスチャーを行う，(6）具体的な行動を示す，(7）ハイタッチなどの物理的な非言語的行動を行う， (8) その他の 8 つのカテゴリーがあるとした．その結果，教師が特に使用する褒める行為のカテゴリーは，(2）,（4）であった．

Morton ら [8] は，褒める行為のプロセスを定義し，各プロセスについて分析を行っている．前提として，褒める行為とは褒め手と受け手の複雑なインタラクションであり，褒め手が一方的に行う行動ではないとしている．Morton らが定義した褒める行為のプロセスは，(1）メリットの評価，(2）褒める意図，(3）褒める行為，(4）評価，(5）褒める行為へのリアクション，(6）褒める行為の調整の主に 6 つがあるとしている．人が褒める行為を行う前には褒めることによるメリットを考え，どのように褒めていくかを考えるとしている．この時，褒める人の個人特性や文化的背景が褒める意図に影響を及ぼしているとしている．その後褒める行為が実行され，受け手は褒め手の褒める行為の評価が行われる．評価は（1）褒める発言が自身の意見と一致しているか，(2）褒め手が褒めた内容に関して専門的知識を有するか，(3）褒める行為に誠実性があり尚且つ自発的か，(4）褒める発

言が文脈上関連していたかの 4 段階であるとした．特に（1）が中心的な評価基準であるとした．受け手が褒める行為を評価した後に，それにもとづいたリアクションが表出される．リアクションとして，感情的，行動的（モチベーションなど），関係性（褒め手との関係が良好になるなど）がある．受け手の性格特性によっては，褒める行為の評価が変化しリアクションも異なってくることが考えられる．褒める人は受け手のリアクションを観察し，自身の想定していたリアクションとの差異を考慮した褒める行為の調整が行われる．望ましい効果が発揮された場合は，再度同様の褒める行為をし，望ましくない効果が発揮された場合は，褒める人による褒める行為の修正が行われる．

Herbert[13] は．性差による褒め言葉の表現の違いの分析を行っている．褒め言葉の対象として 4 つのカテゴリーがあることを特定している．(1）外見，(2）所有物，(3）業績，(4）褒められた人自身が提供した努力の 4 つに分類できるとした．また，褒める行為の効果を調整するために，(1）人称代名詞の選択，(2）褒める内容の長さ，(3）修辞技法（誇張，比喩など）が含まれるとした．

次に，褒める行為の効果に関する研究事例を紹介する．Reinke ら [9] は，教室内での教師の褒める回数の変化と生徒の行動の変化を調査している．実際の小学校教師と小学校を対象として，教師の褒め，叱責回数，生徒が授業中に妨害行為を行ったかを観察する実験を行っている．教師には，事前に日常的に生徒に対して，褒める回数を増加させる行動計画を建てさせた．実験では，教師に対して上記の結果を毎日フィードバックした．その結果，著者らは教師が生徒を褒める回数が増加すると，生徒が授業を妨害する行為を減少させることを明らかにした．また，褒める回数の増加は叱責の回数を減少させることにも貢献した．

Luiselli ら [10] は，教師が生徒を褒めることによる生徒の学業の生成への影響を調査している．実験は，数学の問題を解答した際に教師が生徒を褒める行為を行う期間と行わない期間を交互に 2 回繰り返すことで生徒の成績に変化が生じるのか比較を行っている．その結果，褒める期間において正解した問題数が増加することを明らかにしている．また，褒めない期間に戻した際に生徒の正解した問題数が減少したことも報告している．

2.2 人の個人特性やコミュニケーションスキルを推定する研究

本節では，人の個人特性やコミュニケーションスキルを推定する研究事例について説明する．2.2.1 項では，人の言語的・非言語的行動を用いて人の個人特性（感情，共感，性格など）を推定する研究について説明する．2.2.2 項では，人の言語的・非言語的行動を用いて人のコミュニケーションスキル（プレゼンテーションスキル，説得力リーダーシップなど）を推定する研究について説明する．

2.2.1 人の個人特性を推定する研究

本項では，人間の言語・非言語行動を利用して，個人の性格特性を推定する研究事例について述べる [14][15][16][17]．これらの研究事例では，個人の性格特性を推定するために，人間の言語・非言語行動を記録したコーパスを利用もしくは作成し，コーパスから抽出した特徴量を用いて機械学習モデルを構築している．

Tan ら [14] は，対話における言語・非言語行動を利用して，聞き手が相手の話に共感しているかを予測するモデルの構築・評価を行っている．分析する対象として 2 者対話の収録を行った．2 者対話では，話し手と聞き手に分かれ，話し手は自分自身の過去の出来事について話すようにした．この対話は話し手と聞き手を入れ替えながらを合計 80 回行った．収録後，聞き手には対話中の自身の映像を参照して話し手が話している際の自分自身の共感度を-1〜1 の間で評価した．収録した映像から抽出した音声，言語，顔部の特徴を用いて聞き手の共感度を推定するマルチモーダル LSTM (Long Short Term Memory) モデルを構築している．その結果，聞き手の共感度を予測するには音声と発話内容，特に話者の発話内容が重要であることを明らかにした．

PerezRosas ら [15] は，オンラインビデオ中の話者の感情を言語，非言語情報を用いて分析を行っている．YouTube から収集した動画から，音声，視覚，文字起こしされた発話内容といった 3 つのモダリティを抽出した．抽出した特徴量を用いてサポートベクターマシン（SVM）で話者の感情を推定する機械学習モデルの構築を行った．その結果，感情を推定するためには音声，視覚，発話内容のモダリティを利用した場合が最も精度が良いことが明らかになった．

Lin ら [16] は Group Formation Task の自発的表情データベースから得た，3 人の初対面の人が対話を通じてお互いを知るという実験の映像を利用して話者の感情表現についての分析を行った．映像から，頭部，顔部の動きと書き起こされた発話内容から特徴量を抽出し複数のアルゴリズムを用いて感情特性を推定する回帰モデルを構築した．その結果，Elasti Net を用いたモデルが最も性能が高く，全てのモダリティを用いることが感情特性を予測することに適していることを明らかにした．

Mawalim ら [17] は，グループディスカッションにおける話者の性格特性を予測する，分類モデルの構築と各モダリティの評価を行っている．性格特性の指標は BIG5 を用いている．BIG5 とは Openness（開放性）, Conscientiousness（誠実性）, Extraversion（外向性）, Agreeableness（協調性）, Neuroticism（神経症的傾向）の 5 つの因子から構成されている．コーパスとして，グループディスカッションデータセットである MATRICS コーパスと ELEA-AV コーパスを用いている．特徴量として，音声に関する特徴量（話者識別，MFCC，基本周波数，線形予測係数，線スペクトルペア，音響パワー，話者交代）と言語に関する特徴量（品詞，12 個の対話行為タグ，3 個の発話行為タグ，2 つの意味タグ），視覚に関する特徴量（Action Units，視線，頭の動き），を抽出した．人手で付与された，コミュニケーションスキルスコアとリーダーシップスキルも特徴量として用いた．これらの特徴量を用いて．BIG5 を予測する分類モデルを構築した．アルゴリズムとして，SVM，ランダムフォレスト，ナイーブベイズ，決定木アルゴリズムを用いた．その結果，音声に

関する特徴量を使用する，もしくは音声と他のモダリティを組み合わせることが，重要であることを明らかにした．

2.2.2 人のコミュニケーションスキルを推定する研究

本項では，人間の言語・非言語行動を利用して，個人のコミュニケーションに関するスキルを推定する研究事例について述べる．説得力 [18]，自己開示 [19]，面接スキル [20]，コミュニケーションスキル [21][22]，プレゼンテーションスキル [23][24][25][26] に関する研究事例について述べる．これらの研究事例では，個人の性格特性を推定するために，人間の言語・非言語行動を記録したコーパスを利用もしくは作成し，コーパスから抽出した特徴量を用いて機械学習モデルを構築している．

Park ら [18] は，オンライン・ソーシャル・マルチメディア・コンテンツにおける話者の説得力を分析している．ソーシャル・マルチメディア・ウェブサイトから 1000 本の映画レビュー動画と映画に対する話者自身の 5 段階評価を取得しコーパスを構築した．コーパスから取得した言語・非言語行動に関する特徴量を用いて SVM にて説得力のある話者か説得力の無い話者かを分類する 2 値分類モデルを構築した．その結果，全てのモダリティを用いたモデルの推定精度が 70.84%となり，話者の説得力を予測するためには言語・非言語行動の全てを考慮することが重要であることを明らかにしている．

Soleymani ら [19] は，対話中の言語・非言語行動を用いて自己開示について分析している．分析を行うために，102 名のインタビュー動画を収録した．第三者のアノテータがインタビュー対象者の回答をもとに自己開示の親密さのレベルを 4 段階で評価した．そして，収録したデータから言語・非言語行動の特徴量を自動で抽出した．自己開示のレベルを推定するために，ユニモーダルモデルとマルチモーダルディープニューラルネットワークを構築している．その結果，言語行動のみを用いたモデルが r=0.66 となり，自己開示レベルを推定するためには言語行動のみを用いたユニモーダルモデルが最も精度が高いことが明らかになった．

Naim ら [20] は，言語・非言語行動を利用して，就職面接での社会的スキルを分析している．分析のために，キャリアカウンセラーと大学生が面接する動画を 138 本収録した．動画から，インタビュー対象者の表情，音声，言語の特徴量を自動で取得し，それらを用いて社会的スキルを推定するための回帰モデルを構築した．その結果就職面接で特に重要とされる総合評価と採用の推薦度の予測精度が r=0.70 となった．

Okada ら [21] は，対話中の話者の言語・非言語情報を用いて，コミュニケーションスキルの推定モデルを構築している．1 グループ 4 名の組み合わせを 10 組作り課題解決を目的とした会議を 3 種類行い，その際の参加者の映像と音声を記録した．人事管理経験者が記録した映像，音声を参照し，参加者のコミュニケーションスキルを判断した．さらに，記録した映像，音声から話者交替，韻律情報，頭部の動き，品詞，対話行為を特徴量として抽出した．人手で判断したコミュニケーションスキルのスコアと，抽出した特徴量を用いてコミュニケーションスキルのスコアを予測する回帰モデルを構築した．

Rasipuram ら [22] は，言語・非言語情報を抽出して機械学習をすることによって，非同期インタビューと対面インタビューでの人のコミュニケーションスキルを予測するシステムを構築している．非同期インタビューと対面インタビューの映像，音声を 69 件収録した．インタビュー映像から参加者の音声，視覚，言語に関する特徴量を自動で抽出した．コミュニケーションスキルとして 13 個の評価基準を設け，3 名のアノテータによって評価された．アノテータの評価と抽出した特徴量を用いて線形回帰，SVM，ロジスティック回帰などの機械学習アルゴリズムを用いてコミュニケーションスキルとして 13 個の評価基準を予測するシステムを構築した．その結果，非同期インタビュー動画から自動で特徴量を抽出した場合でも 80 %の精度となりコミュニケーションスキルを予測するために十分であることが明らかになった．

Ramanarayanan ら [23] は，時系列に基づいて計算された 3 つの異なる特徴量を用いて，人間が評価したプレゼンテーション能力のスコアを予測する際の比較を行っている．データセットとして，プレゼンテーション中の 3D 身体動作，音声，映像を記録した 56 個のプレゼンテーションから構成されている．上記のデータセットから，頭部方向，視線方向，表情，運動量や手の動きの程度を特徴量として抽出している．これらの特徴量を用いて，プレゼンテーション能力のスコアを予測するタスクを行ったところ，すべての特徴量を組み合わせた機械学習モデルにおいて最も良いスコアが予測されることが確認された．

Yagi ら [24] は，プレゼンテーションシーンから得られる言語・非言語情報からプレゼンテーションでメッセージ作成スキルを推定するモデルの構築・評価を行っている．データセットとして，プレゼンテーション中の 3D 身体動作，音声，映像を記録した．プレゼンテーションのテーマは自身の仕事内容や仕事の魅力の紹介であった．収録したデータセットに対して，専門家によるプレゼンテーションの評価を行った．評価項目は，プレゼンテーションの目標設定に関するもの，プレゼンテーションの構成に関するもの，スピーチ中の振舞いに関するものであった．上記のデータセットから，韻律情報，言語情報（品詞の種類と使用回数，単語埋め込み表現），視覚情報（上半身 19 部位の動き，表情，視線方向）を特徴量として抽出している．これらを用いて，プレゼンテーションにおけるメッセージ作成スキルを予測する SVR モデルを構築した．その結果，評価項目によって重要な特徴量の組み合わせに違いがあることが確認された．目標設定に関する項目は，韻律に関する特徴量と言語に関する特徴量を組み合わせた機械学習モデルが最も性能が高い結果となった．構成に関する項目では，韻律に関する特徴量のみを使用した機械学習モデルが最も性能が高い結果となった．振舞いに関する項目では，全ての特徴量を組み合わせた機械学習モデルが最も性能が高い結果となった．

Schneider ら [25] は，非言語コミュニケーションに関するリアルタイムフィードバックシステムの構築を行っている．プレゼンテーション中の 3D 身体動作，韻律情報を入力することで，非言語コミュニケーションに関するフィードバックを生成している．提案しているシステムのフィードバック内容は身体の姿勢、ジェスチャーの使用、声の大きさ、間の使用、発話の一時停止の使用、片足からもう片方の足に体重を移動させずに地に足をつけている能力など非言語的行動である．身体の姿勢，ジェスチャー，足の体重移動に関しては 3D 身体動作の値を使用して事前に設定した閾値との比較をすることでフィードバッ

クを生成している．声の大きさ，間の使用に関しては韻律情報の値を使用して，事前に設定した閾値との比較をすることでフィードバックを生成している．発話の一時停止の使用に関しては，発話中に “えー” や “あー” といったフィラーが発生したかを使用してフィードバックを生成している．フィードバックは自身の動きがリアルタイムに表示された画面上に重畳するように提示された．画面にはどの種類の行動を修正して欲しいのかと，どのように修正すればよいのかを説明する短い文章が表示された．効果検証として，提案システムを使用したグループと単に自身の動きを画面に表示しているグループとの比較を行った．その結果提案システムを使用したグループが有意に修正される行動の回数や時間が減少した．

Chen ら [26] は，スピーチパフォーマンスの自動スコアリングモデルを構築している．データセットとして，4 つの課題（ビジネスプレゼンテーション，中学生を対象とした模擬授業，説得プレゼンテーション，即興プレゼンテーション）中の 3D 身体動作，映像，音声を記録した．スピキーングスキルの評価方法として Public Speaking Competence Rubric を使用した．特徴量として，言語に関する特徴量（動詞句や T-units，節などの 9 種類の構文構造の出現頻度と平均節長や T-unit あたりの動詞句などの 14 種類の構文の複雑さ）と音声に関する特徴量（発話速度，発音，韻律），視覚に関する特徴量（腰の動き，手の動きの平均値と標準偏差及びの向き）を抽出した．これらの特徴量を用いて，スピーキングスキルを予測する回帰モデルを構築した．機械学習モデルのアルゴリズムとして，ランダムフォレストと SVM を用いた．その結果，全てのモダリティを使用することが重要であることを明らかにした．

Saga ら [27] は，ソーシャルスキルの自動評価及びフィードバックシステムの構築を行っている．データセットとして，human-human SST[28] を使用している．データセットは，49 名の参加者が (1) 他者に注意を払う，(2) ポジティブな感情を伝える，(3) 他者に要求する，(4) 要求を拒否するという 4 つのタスクを行ったセッションを 2，3 回繰り返した際の 3D 身体動作，映像，音声を記録した．ソーシャルスキルは専門家によって評価されている．評価項目は，アイコンタクト，体の方向/距離，顔の表情，声の変化，明瞭さ，流暢さ，各タスクの社会的適切性の 7 つである．データセットから特徴量として，言語的特徴量（BERT を用いた文章の類似性，発話内単語数，感謝・謝罪・拒絶フラグ，バックチャネル，1 分あたりの単語数）と韻律的特徴量（声の強さ，F0），視覚的特徴量（頭部の動き，表情，全身の運動量，上半身の運動量，腕の運動量，お互いの笑顔の頻度）を抽出した．これらの特徴量を用いて，ソーシャルスキルを予測する回帰モデルを構築した．機械学習モデルを構築するために,，部分最小二乗回帰，SVR，ランダムフォレスト，ニューラルネットワーク，回帰モデルの変分オートエンコーダ（VAE）を使用した．その結果，ランダムフォレストが最も高い性能となった．さらに，ランダムフォレストにおいて最も性能の高い性能となった特徴量の組み合わせはタスクごとに異なっていた．具体的には，(1) は視覚的特徴量と韻律的特徴量，(2) と (3) は視覚的特徴量と言語的特徴量，(4) は韻律的特徴量と言語的特徴量であった．フィードバックは，推定された結果から事前に設定された閾値を使用することで表示する内容を変化させた．予測結果が閾値より高ければポジティブな内容となり，閾値より低ければ改善を促すような内容とした．フィードバック

する内容は，項目ごとに事前に準備した.

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

日常生活や社会活動において褒める行為は多くの人が行う大切なコミュニケーションの 1 つである．心理学や社会学の分野で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表現する言語・非言語行動であり [1][29]，送り手から受け手への一方的な意思の伝達ではなく複雑なコミュニケーションであると考えられている [2]．褒めることにより，相手との関係性を円滑にすることや，相手の能力を最大限発揮してもらうなどの効果が期待されている [3][4]．2.1 節で述べたように心理学や教育学における褒める行為の分析は，アンケートによる回答結果や，特定の行動の変容の観測，実際の対話を参照し傾向を観察するといった手法を用いて分析を行い，褒める行為の機能や効果について明らかにしている．褒める行為の表現の種類 [4][11][12] や褒める行為の流れ [8]，性差による褒める表現の違い [13] について調査，分析している．Reinke ら [9] や Luiselli ら [10] は褒めることによる効果について調査，分析している．しかし，これらの研究で明らかにされた効果を最大限に発揮するために褒める人はどのように言語・非言語行動を行えば良いのか定量的に明らかになっていない．さらに，褒める行為中に褒められる人がどのような振舞いを行っているのか明らかでない．さらに，対話中の人間の能力を分析している研究は 2.2 節で述べたように多く存在している．だが，相手を上手く褒めるために褒める人が言語・非言語行動をどのように用いれば良いのかを明らかにする取り組みはまだ行われていない．そのため，相手を上手く褒めるために，頭部の動きや表情，話す速度，声の抑揚，発話内容などの言語・非言語行動をどのように用いれば良いのか定量的には明らかでないという問題がある．この問題が明らかになっていないため，褒め方の上手さを把握，改善することが困難であると考えられる．先行研究 [30][31] では，頭部，顔部の動きや表情，音声情報といった非言語的行動を用いて褒め方の上手さの分析を行っている．しかし，褒める行為は非言語的行動だけではなく，言語行動も伴っているが，先行研究では言語的行動に関する分析は行われていない．そのため，話者が言語的行動をどのように用いれば良いのか明らかでない．加えて，先行研究では言語的・非言語的行動を組み合わせた分析を行っていない．そのため，褒める行為における言語的・非言語的行動の相互効果が褒め方の上手さと関係しているか明らかでない．そこで，本論文では先行研究で明らかにできていない下記内容を問題として設定する．

- 上手く褒めるために話者の言語的行動をどのように用いれば良いのか定量的に明らかでない
- 上手く褒めるために話者の言語的・非言語的行動をどのように組み合わせれば良いのか定量的に明らかでない
- 対面・遠隔対話における上手く褒めるための振舞いに相違点があるか定量的に明らかでない

3.2　研究課題の設定

3.1 節で述べた問題を解決するためには，相手を上手く褒めるために重要な言語・非言語行動を明らかにする必要がある．先行研究 [30][31][32][33][34] では，頭部，顔部の動きや表情，音声情報といった非言語行動を用いて褒め方の上手さの分析を行っている．だが，著者が調査したところ，相手を上手く褒めるためにはどのように言語行動を用いれば良いのかを工学的に分析し，定量的に明らかにする取り組みはまだ行われていない．褒める行為は非言語行動だけでなく言語行動も伴っていると考えられている [1][29]．また，2.2 節で述べたように，人間の行動や能力を分析するためには言語行動も重要なモダリティであることを示している．加えて，褒める行為は褒める人から褒められる人への一方的な意思の伝達ではなく，複雑な社会的コミュニケーションであると考えられている [2]．これらのことから，褒める人と褒められる人の発話内容に着目して言語行動を用いれば上手く褒めることができるのかを明らかにする必要がある．さらに，2.2 節で紹介した多くの研究では，言語・非言語行動を組み合わせて人間の行動や能力の分析を行っている．しかし，先行研究では非言語的行動に着目した分析を行っており，言語的行動や言語・非言語的行動を組み合わせた分析が行われていない．そのため，相手を上手く褒めるために言語・非言語的行動をどのように用いれば良いのか明らかでない．そして，先行研究の多くは対面対話における褒め方の上手さを対象としていた．そのため，COVID-19 の影響により普及してきた遠隔対話 [5] における褒め方の上手さの分析は行われておらず，遠隔対話において上手く褒めるために言語・非言語行動をどのように用いれば良いのか明らかでない．対面対話と遠隔対話では，言語・非言語情報の伝達のされ方に差異があるとされている [6][7]．先行研究 [32] では，遠隔対話コーパスの作成と視覚的特徴量に関する振舞いの分析にとどまっており，相手を上手く褒めるためにはどのような振舞いが重要であるかなどの分析は行われていない．先行研究と本論文との差分を表 3.1 示す．

表 3.1: 褒める行為を工学的に分析している先行研究と本論文との差分

	対面対話						遠隔対話					
	褒める人			褒められる人			褒める人			褒められる人		
	視覚	韻律	言語	視覚	韻律	言語	視覚	韻律	言語	視覚	韻律	言語
先行研究 [30][31][32][33][34] で着目された行動	✓	✓		✓	✓		✓					
本論文でのみ着目している行動			✓			✓		✓	✓	✓	✓	✓
本論文で着目した行動	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

上記を踏まえ本研究では，**対面・遠隔対話における褒める人と褒められる人の視覚的，韻律的，言語的行動に着目して，相手を上手く褒めるために重要な振舞いを明らかにすること**を研究課題として設定する．具体的には，対面・遠隔対話中に褒めているデータセットから言語的・非言語的行動に関する特徴量を抽出し褒め方の上手さを推定する機械学習モデルを構築して褒め方の上手さに重要な特徴量を明らかにすることを目指す．加えて，対面・遠隔対話において上手く褒めるために重要な振舞いに相違点があるか明らかにすることを目指す．

第4章　対話コーパス

本章では，本研究で用いる対話コーパスについて述べる．本研究では，対面・遠隔対話における褒め方の上手さの評価と言語・非言語行動の記録を含む対話コーパスを用いた．対面対話コーパスは，既存で構築されたコーパスを用いた [32]．4.1 節にて対面対話コーパスの中身について述べる．遠隔対話コーパスは，新たに構築を行った．具体的には，言語・非言語行動を記録するために 2 者による遠隔対話の収録を行い，収録した対話データを参照しアノテーションと褒め方の上手さに関する評価を行った．4.2 節にて遠隔対話コーパスの構築について述べる．

4.1　対面対話コーパス

本節では，本論文で用いた対面対話コーパスの具体的な内容について述べる．本論文で用いた対面対話コーパスは，Onishi ら [32] が構築したコーパスである．コーパスには，対面対話における，話者の言語・非言語行動と，褒め方の上手さに関する評価値が含まれている．具体的な構築方法については次節以降で述べる．4.1.1 項では，対面対話の収録方法について述べる．4.1.2 項では，収録された対話データに対してアノテーションした方法について述べる．

4.1.1　対面対話の収録方法

本項では，対面対話の収録方法について述べる．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．対話の収録を始めるにあたり，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てもらった．このときの参加者間の距離は 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．各組の話者（話者 A，話者 B）は，撮影者の合図に従って下記の対話 1〜対話 3 を行った．

対話 1　自己紹介（5 分間）

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

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

対話 1〜対話 3 の対話を 17 組分，計 255 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り，参加者の緊張をほぐす目的で行っているため，分析の対象外とした．対話 2・3 では，褒められる人が事前に用意した「今まで頑張ってきたこと」に関するエピソードを話し，褒める人はその話を聞いて随時発言を行うように教示した．このとき，褒められる人に対しては，スピーチのように一方的に話さず，相手との対話であるこ

図 4.1: 2 者による対面対話の様子

とを意識して話してもらうよう指示し，会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．褒める人に対しては，対話相手の話を聞いている中で褒めることがあれば積極的に褒めるように指示した．ただし，褒める発言しかしないといった不自然な会話にならないよう，褒める以外の発言や質問などを自由に行うことを許可した．

4.1.2 アノテーション

本項では対面対話データに対するアノテーション及び褒め方の上手さの評価方法について述べる．4.1.1 項で収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[35] を利用して人手で発話シーンの注釈付けを行った（表 4.1）．本研究の分析単位は話者の各発話とするため，対話中の発話の切れ目を定義することで対話中の発話シーンを抽出した．発話シーンとは，発話の開始から終了までの場面のことである．発話シーンは，沈黙時間が 400ms 未満の連続した音声区間とした．このとき，話者の言語的行動を分析するため，どのような発話をしているのかが重要な要素である．そのため発話シーンは，文章として意味が理解できる発話の区切れとする必要がある．そこで，最も自然に発話内容を区切るための閾値を実際の発話を用いて検討した．日本語の対話コーパスを構築する際、無声区間 400ms 以上を発話区切りの基準としている研究がある [36]．この研究を参考に，発話シーンを区切る閾値を 400ms 程度に絞り込んだ．適切な閾値を確認するために，200ms，300ms，400ms，500ms を閾値として設定した．実際のデータを分析した結果，400ms が妥当であると判断した．次に，ELAN を利用して人手で発話内容

の書き起こしを行った．書き起こしを行う際，人手による書き起こし内容の感動詞とフィラーの表記ゆれが分析結果に影響を及ぼすことが考えられたため，感動詞とフィラーの表記ゆれを減らすために辞書を参照しながら作業を行った．本研究では，UniDic 話し言葉辞書 [37] に載っている感動詞とフィラーの単語を抜き出したものを辞書として扱った．このときアノテータには，発話内容の文頭が辞書内の単語の読みに近い場合は，辞書内の単語で書き起こすように指示した．たとえば，“あー”，“あ〜” などは “あー” と書き起こす．なお，辞書に載っていない単語については，聞き取った単語をそのまま書き起こすように指示した．

表 4.1: 対面対話における発話シーンのシーン数，平均値，最大値，最小値

	Scenes	Mean(sec)	Max(sec)	Min(sec)
発話シーン（褒める人）	2701	1.324	23.117	0.062
発話シーン（褒められる人）	3413	2.040	26.234	0.018
Praise シーン	228	2.018	9.172	0.368

褒め方の上手さの評価は，2 者対話の収録に参加していない 20 代の大学生 5 名（男性 5 名）が行った．第三者による客観的な観察に基づく評価と参加者本人による主観的な評価には，一定の相関関係があることも報告されている [38]．また，実際に褒められた人の評価と第三者との評価の相関係数を算出したところ，$r = 0.773$ となり，正の相関があることが確認された．また，個人の評価だけでは，その人の趣味趣向が反映された偏った評価となることが考えられる．本論文では，多くの人が良いと思う褒め方を分析することを目的としている．そのため，褒め方の上手さの評価は第三者が行った．具体的には，褒める人の正面に設置したビデオカメラから撮影した映像データと褒める人に取り付けたマイクから録音した音声データを参照し，褒める人の発話シーンごとに下記の判定・評価を行った．

- 対話相手を褒めているシーンであるか，そうでないかの判定
- 褒めているシーンであると判定した場合，1（上手く褒められていない）〜7（上手く褒められている）の 7 段階での褒め方の上手さの評価

本研究では多くの人が共通して考える上手い褒め方の振舞いを分析したかったので，アノテータには上手い褒め方について指示をしていない．評価基準を提示することで，アノテータが評価する項目が偏る可能性が考えられたためである．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 名以上であるシーンを Praise シーン（表 4.1）とする．Praise シーンにおいて褒めているシーンであると判定したアノテータの評価の平均値を Praise スコアとする．アノテータ間の Praise スコアの一致率を評価するために，級内相関係数（ICC）[39] を利用した．まず，3〜5 人のアノテータのそれぞれの組み合わせごとに級内相関係数を算出した．次に，サンプル数を考慮して組み合わせごとの級内

相関係数の重み付き平均を算出した．その結果，Praise スコアの級内相関係数の平均は $ICC(2,k) = 0.571$ であった．この結果から，Praise スコアはアノテータ間で中程度の一致率（[40]）があり，信頼性のあるデータであることが示唆された．

4.2 遠隔対話コーパス

本節では，遠隔対話コーパスについて述べる．本研究は対面対話における上手く褒めるための振舞いと遠隔対話における上手く褒めるための振舞いの比較を行う．そのためには，同様の手順で収録された対面対話コーパスと遠隔対話コーパスが必要である．そこで，本研究では 4.1 節で紹介した対話コーパスを参考に遠隔対話コーパスの構築した．具体的な構築方法については，次項以降で説明する．4.2.1 項では，遠隔対話における 2 者対話の収録について述べる．4.2.2 項では，収録した対話データに対するアノテーションについて述べる．

4.2.1 遠隔対話の収録

本項では，遠隔対話の収録について述べる．オンラインコミュニケーションツールである Zoom[41] を用いて，2 者遠隔対話中の映像・音声を収録した．Zoom を用いた理由としては，普及率が高いシステムである点，通話品質が十分である点，話者ごとに音声を録音が可能である点が挙げられる．収録に際して，話者に対して下記の点を注意するように指示をした．

- 前髪はヘアピンなどで上げておく
- 髭は剃っておく
- メガネを使用している人はコンタクトレンズの使用や裸眼にする
- 背景に撮影参加者以外となるべく家具や物などが映らないようにする
- 肩から上が写っている
- ピントが話者の顔に合っており明瞭に写っている
- 話し手聞き手ともに，原則として Zoom 越しに相手の顔を見て話す（対話中の自然な動作として，目をそらす，頷く，驚いて仰け反る等は可）
- 対話とは関係ない行動（例：紙媒体のメモを読む，スマートフォンで検索する，Zoom の画面共有機能で自分の PC 画面を共有する）は禁止する

適切に視覚的な行動を分析するためには，頭髪や装飾品などにより顔全面が隠れていないこと，話者がなるべく正面を向いていること，話者以外の情報が画面に存在しないことが重要であると考えたので，上記の指示を行った．話者は 20 代〜50 代の 40 名（男性 20 名，女性 20 名）であり，それぞれ異なる相手と 3 回ずつ対話を行った．これより，合計 60 組の対話が記録された．各組の参加者の年齢層・性別は同じであり，初対面であった．話者は，図 4.2 に示すように，一般的な遠隔対話シーンを想定し机上に置かれた PC（詳細なスペックは表 4.2 に示す）の前に座り画面に映る対話相手を見ながら会話をした．話者の前に設置された PC の内蔵カメラで各話者の映像を，PC の内蔵マイクで各話者の音声を記録した．対話は雑音のない防音室内で行われた． 各組の話者（話者 A，話者 B）

図 4.2: 2 者による遠隔対話の様子

は，撮影者の合図に従って下記の対話 1〜対話 3 を行った．

対話 1 自己紹介（5 分間）

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

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

表 4.2: 収録機材のスペック

	性能	
PC	メーカー	DELL
	型番	Vostro 5502
	CPU	Core i7-1165G7
	RAM	16GB(DDR4,1x16GB)
	ROM	512GB(M.2 PCIe NVMe SSD)
	GPU	NVIDIA GeForce MX330(VRAM:2GB)
	液晶	15.6 インチ
映像	解像度	640x360
	ビットレート	256kbps
音声	サンプルレート	32kHz

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，各組が初対面であることから，話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・3 では，褒められる人が事前に用意した「今まで頑張ってきたこと」に関するエピソードを話し，褒める人はその話を聞いて随時発言を行うように教示した．このとき，褒められる人に対しては，スピーチのように一方的に話さず，相手との対話であることを意識して話してもらうよう指示し，会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．褒める人に対しては，対話相手の話を聞いている中で褒めることがあれば積極的に褒めるように指示した．ただし，褒める発言しかしないといった不自然な会話にならないよう，褒める以外の発言や質問などを自由に行うことを許可した．

4.2.2 アノテーション

本項では遠隔対話データに対するアノテーション及び褒め方の上手さの評価方法について述べる 4.2.1 項で記録した映像データと音声データに対して，自動で音声認識を行うシステムである Automatic speech recognition (ASR) *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．発話シーンは，対面対話コーパスと同様に沈黙時間が 400ms 未満の連続した音声区間とした（表 4.3）．

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 名による，対面対話コーパスと同様の褒め方の上手さの評価を行った．各アノテータは文系・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．アノテータは収録した映像データと音声データを参照し，褒める人の発話シーンごとに下記の判定・評価を行った．

*ASR は日本電信電話株式会社によって開発された音声認識システムである．

表 4.3: 発話シーンと Praise シーンの情報

	Scenes	Mean (sec)	Max (sec)	Min (sec)
発話シーン（褒める人）	16351	1.233	11.180	0.450
発話シーン（褒められる人）	27834	1.959	13.430	0.450
Praise シーン	236	1.998	10.610	0.470

- 対話相手を褒めているシーンであるか，そうでないかの判定
- 褒めているシーンであると判定した場合，1（上手く褒められていない）〜7（上手く褒められている）の 7 段階での褒め方の上手さの評価

本研究では多くの人が共通して考える上手い褒め方の振舞いを分析したかったので，アノテータには上手い褒め方について指示をしていない．評価基準を提示することで，アノテータが評価する項目が偏る可能性が考えられたためである．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 名以上であるシーンを Praise シーン（表 4.3）とする．さらに，Praise シーンにおいて褒めているシーンであると判定したアノテータの評価の平均値を Praise スコア（平均値と標準偏差：4.280 ± 0.931）とする．アノテータ間の Praise スコアの一致率が低いと後述の分析の信頼性が下がるため，彼らのスコアの一致率を ICC で評価した．具体的には，3〜5 人のアノテータのそれぞれの組み合わせごとに級内相関係数を算出した．次に，サンプル数を考慮して組み合わせごとの級内相関係数の重み付き平均を算出した．その結果，Praise スコアの級内相関係数の平均は $ICC(2, k) = 0.701$ であった．この結果から，Praise スコアはアノテータ間で高い一致率があり，信頼性のあるデータであることが示唆された．

第5章　対面対話コーパスにおける上手く褒めるための振る舞いの分析

本章では，対面対話コーパスを用いた上手く褒めるための振る舞いの分析について述べる．具体的には，4 章で構築した対話コーパスから，言語・非言語行動に関する特徴量を抽出し，抽出した特徴量を用いて褒め方の上手さを評価した値を推定する機械学習モデルの構築を行った．構築する褒め方の上手さ推定モデルの概要を図 5.1 に示す．

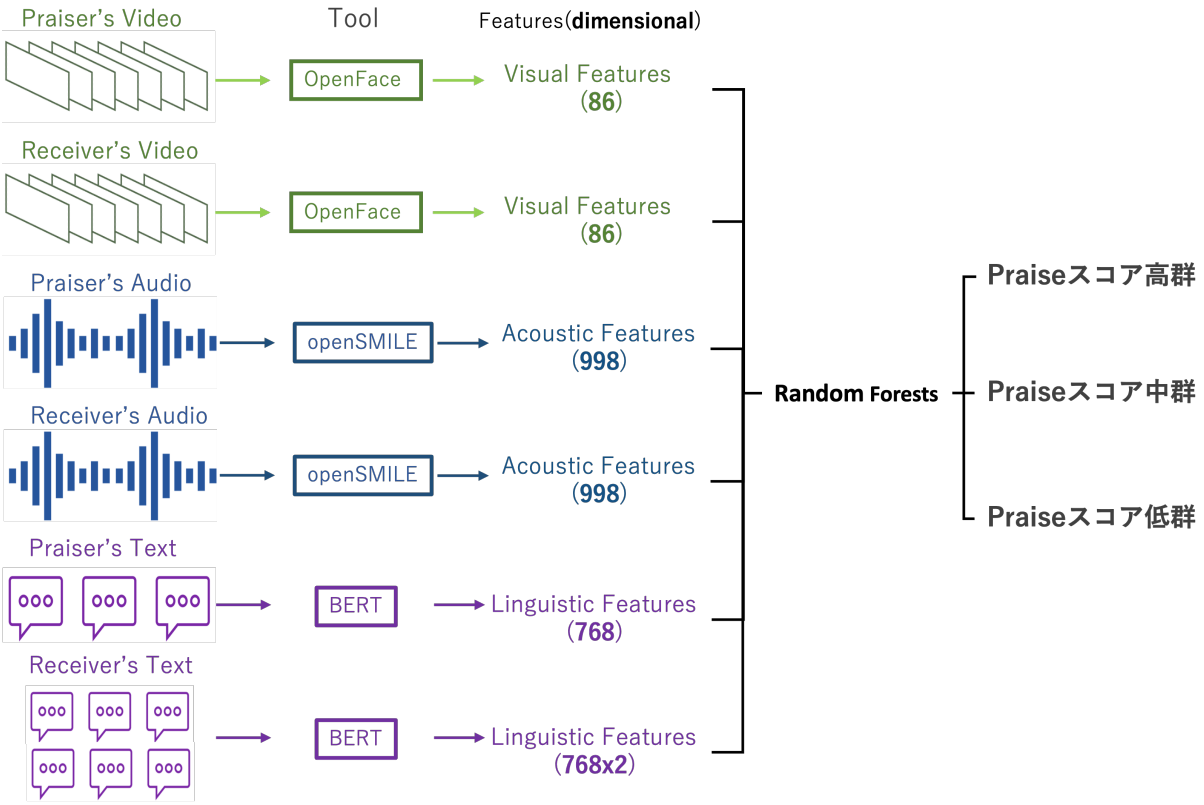

図 5.1: モデルの概要

5.1　特徴量の抽出

本節では，機械学習モデルを構築するために使用する特徴量の抽出について述べる．本研究では，2.2 節にて述べた研究を参考に，視覚的，韻律的，言語的行動に関する特徴量を用いる．以降，5.1.1 項，5.1.2 項，5.1.3 項にて，それぞれ視覚的行動，韻律的行動，言語的行動に関する特徴量の抽出方法について述べる．

5.1.1　視覚的特徴量

本項では，視覚的行動に関する特徴量の抽出について述べる．顔画像処理ツールである OpenFace[42] を用いて，話者の正面に設置したカメラで撮影した映像データから，各

話者の頭部・顔部の振舞いに関連する特徴量と Action Units（AU）*[43] に関する特徴量の抽出を行った．頭部の向きに関しては，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（図 5.2）における褒める人と褒められる人の頭部の 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，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．AU 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 AU（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．Praise シーンの前後 1 秒ずつを含む範囲から視覚的特徴量を抽出した理由は，褒める行為の直前と直後の表情や顔部の動きも重要であると考えたからである．表情や顔部の動きは発話中でなくても絶え間なく変化している．そのため，褒める発話をする直前から褒める行為特有の表情や顔部の動きを行っている可能性があると考えた．褒める発話をした直後にも同様のことが考えられる．そのため Praise シーンの前後 1 秒ずつを含む範囲から視覚的特徴量の抽出を行った．

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

本項では，韻律的行動に関する特徴量の抽出について述べる．音声情報処理ツールである openSMILE[44] を用いて，話者に装着され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．Praise シーンの範囲（図 5.3）における褒める人と褒められる人の表 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

*Action Units とは，顔の個々筋肉や筋肉群の基本的な行動の単位を表す．

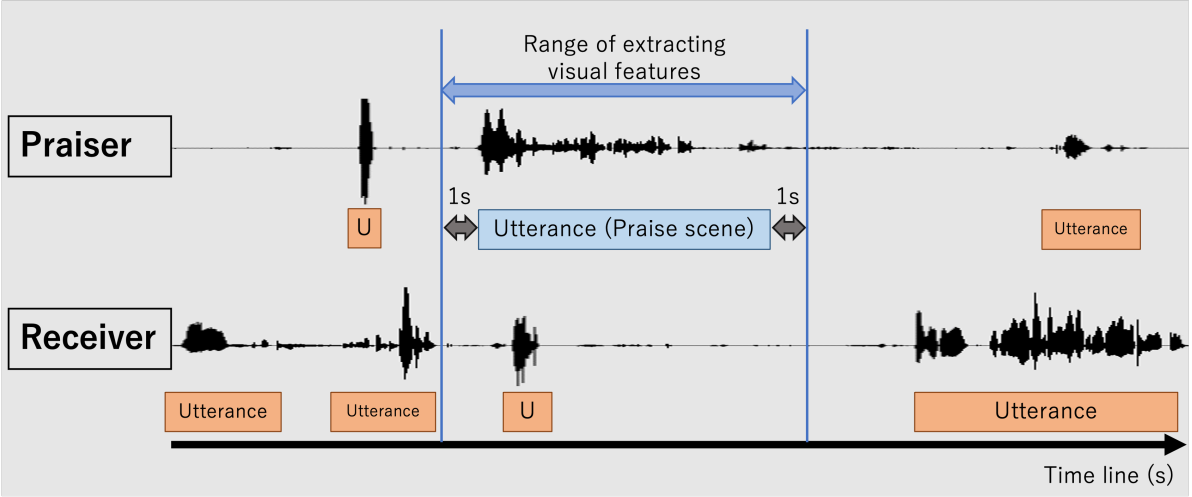

図 5.2: 視覚的特徴量の抽出範囲

の（_de）の計 988 の特徴量 [45][46] として用いた．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．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: 韻律的特徴量の内容

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

5.1.3　言語的特徴量

本項では，言語的行動に関する特徴量の抽出について述べる．言語的特徴量の抽出方法として，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，soleymani ら [19] の研究を参考に，日本語事前学習済みの BERT モデル [47] を利用して Praise シーンの褒める人の発話内容と Praise シーンの直前の褒められる人の発話内容（B），Praise シーンの直後の褒められる人の発話内容（A）（図 5.4）をそれぞれ 768 次元

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		

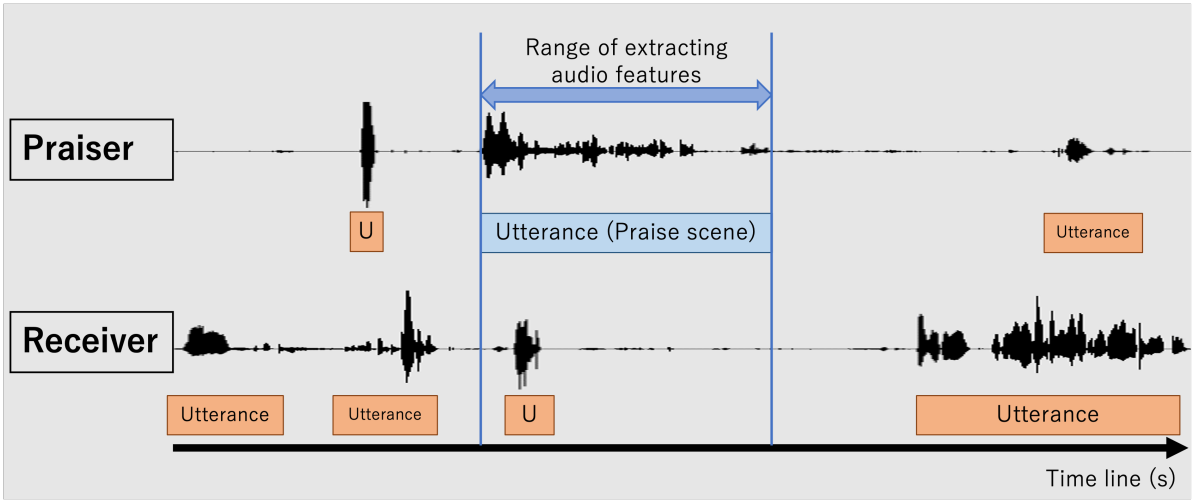

図 5.3: 韻律的特徴量の抽出範囲

のベクトル（文章埋め込み）に変換する方法を用いた．BERT とは，大規模コーパスを用いて事前学習したモデルを転移学習により様々なタスクに応用できる言語モデルである．具体的な抽出方法は，入力した文章を BERT が処理した最終層の先頭にある特殊トークンである [CLS] の埋め込み表現を取得し言語的特徴量とした（図 5.5）．

5.2　褒め方の上手さ予測モデル構築

本節では，機械学習モデルの構築について述べる．5.1 節にて抽出した特徴量を用いて 4.1.2 項にて評価した褒め方の上手さの評価値を予測する機械学習モデルを構築した．本研究では機械学習モデルを構築することで，褒め方の上手さの予測に寄与する特徴量の組み合わせから，褒め方の上手さに重要な行動を明らかにする．機械学習モデルを構築する

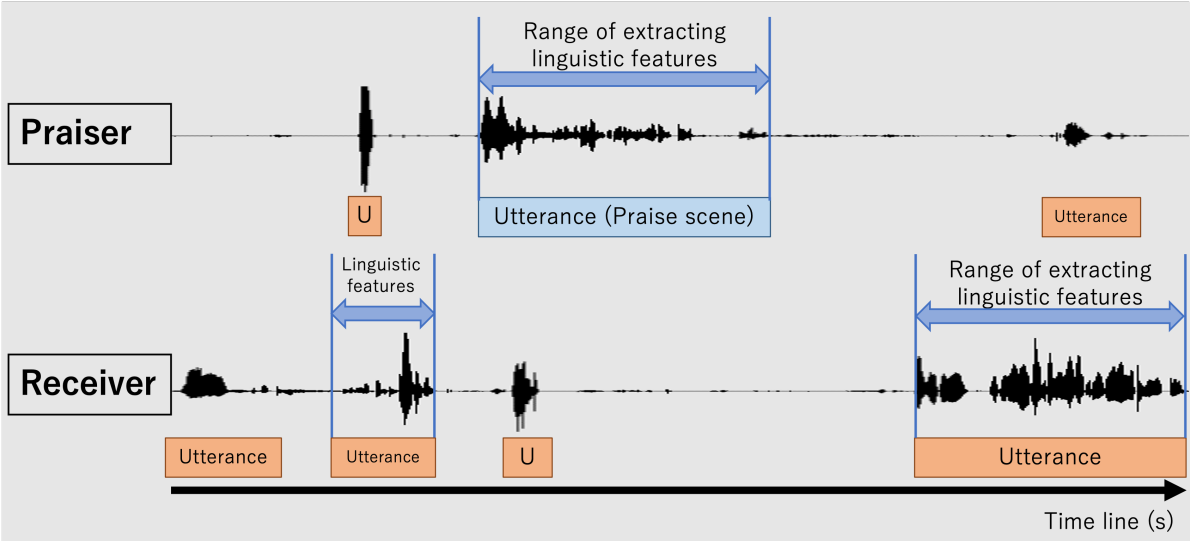

図 5.4: 言語的特徴量の抽出範囲

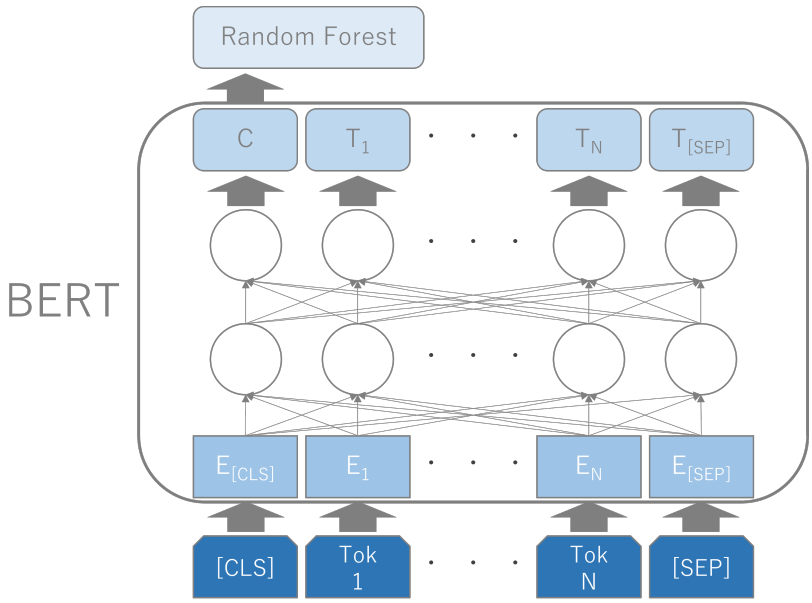

図 5.5: BERT を用いた文章ベクトルの抽出

にあたり，既存研究 [48][49][50] を参考に praise シーンを praise スコア低群，中群，高群の 3 クラスに分割した．対面対話コーパスにおける各群に属する Praise シーン数（全 228 シーン）ができるだけ均等になるように，下記のように praise スコア低〜高群を定義した† ．

praise スコア低群 Praise スコアが 3.8 点以下の praise シーン（計 82 シーン）

praise スコア中群 Praise スコアが 3.8 点より大きく 4.4 点未満の praise シーン（計 65 シーン）

praise スコア高群 Praise スコアが 4.4 点以上の praise シーン（計 81 シーン）

本論文ではモデルを構築するアルゴリズムとしてアンサンブル学習のバギング手法である Random Forest[51] を利用し，分類モデルを構築した．具体的には，目的変数を Praise スコア低群，中群，高群の 3 クラスとし，説明変数を 5.1 節で抽出した特徴量とする分類モデルを構築した．決定木の木の本数や木の深さといったハイパーパラメータは Hyperopt[52] を用いて最適化した．ハイパーパラメータのチューニングは 90%のトレーニングデータと 10%のテストデータに分割されたデータを使用して 1000 回行った．1000 回の試行結果に基づいて，最も性能の良いパラメータを用いて機械学習モデルを構築した．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，褒め方の上手さを推定するために寄与するモダリティを明らかにするために，各モダリティの組み合わせを変えて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）〜（2）を 100 回ずつ行った．

(1) 特徴量が含まれている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．

(2) 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．

このとき，ベースライン（チャンスレベル）は，データセットにおける各群の割合にあわせて，36%，28%，36%の確率で praise スコアの低，中，高群を出力するモデルを用いた．

5.3 対面対話の分析

本節では対面対話コーパスから 5.1 節にて抽出した特徴量を用いて 5.2 節にて構築した機械学習モデル結果について述べる．構築した分類モデルの推定結果を表 5.4 に示す．モデルの性能を評価するにあたり，F 値の重み付き平均を用いた．F 値とは，再現率 ($recall$) と適合率 ($precision$) の調和平均である．$recall$ とは，機械学習モデルが真であると予測した数を分母に，正しく予測できていた数を分子にした値である．$precision$ とは，正解

†各群に属するシーン数は等しいことが理想ではあるが，スコアが同じシーンが多数存在したため，各群のシーン数を等しくすることができなかった．

データの真である数を分母に，モデルが正しく予測できていた数を分子にした値である．具体的には下記の式で算出する．この時，正しく予測できた値を TP，真であるのに偽と予測した値を FN，偽であるのに真と予測した値を FP とする．

$$recall = \frac{TP}{TP+FN} \tag{5.1}$$

$$precision = \frac{TP}{TP+FP} \tag{5.2}$$

$$F\text{値} = \frac{2 \cdot recall \cdot precision}{recall+precision} \tag{5.3}$$

求まった F 値に対して重み付き平均を算出する．具体的には下記の式で算出する．この時，予測する各クラスのラベル 0,1,2 の F 値を F_0, F_1, P_2, ラベルの個数を n_0, n_1, n_2 とする．

$$F\text{値の重み付き平均} = \frac{F_0 n_0 + F_1 n_1 + F_2 n_2}{n_0 + n_1 + n_2} \tag{5.4}$$

以降 F 値は重み付き平均された F 値とする．加えて，各モデルの F 値の組に対して有意水準 5%のマン・ホイットニーの U 検定を行い，多重検定の問題を回避するために p 値をホルム法で補正した．

単一の特徴量のみを用いた機械学習モデル（M1-M7）の結果を確認する．M2，M6，M7 を除く全てのモデルとベースラインとの間に 5%水準で有意な差が確認できた．また，M2，M6，M7 を除く全てのモデルがベースラインよりも高い F 値となった．よって，褒め方の上手さを予測するためには言語・非言語行動が重要であることが示唆された．最も F 値の高いモデルは，褒める人の言語的特徴量を用いたモデル（M3）であった．M3 と M1，M2，M4，M6，M7 との間に 5%水準で有意な差が確認できた．このことから，褒める人の言語的行動が他の行動に比べて褒め方の上手さに関係していると考えられる．

複数の特徴量を用いた機械学習モデル（M8-M127）の結果を確認する．M8, M12, M13, M15, M17, M18, M24, M25, M28, M30, M32, M33, M39, M40, M43, M49, M50, M53, M62, M69, M70, M73, M82, M92, M107 を除く全てのモデルとベースラインとの間に 5%水準で有意な差が確認できた．また，M24，M28，M62 を除く全てのモデルがベースラインよりも高い F 値となった．よって，褒め方の上手さを予測するためには複数の言語・非言語行動を組み合わせることが重要であることが示唆された．最も F 値の高いモデルは，褒める人の視覚的，韻律的，言語的特徴量を用いたモデル（M29）であった．M29 の性能は F 値 $= 0.556$ であった．M29 との間に有意な差が確認されたモデルに多くに共通していることは，褒める人の言語的特徴量を用いていないという点である．褒める人の言語的行動は，F 値が高い機械学習モデル上位 5 位の全てにおいて用いられている．加えて，M29 における重要度の高い特徴量上位 20 件を図 5.6 に示す．図 5.6 より，重要度の高い特徴量上位 20 件は全て言語的特徴量であった．このことから褒め方の上手さを予測するためには，褒める人の言語的行動と他の行動を組み合わせることが重要であると考えられる．一方で M29 との間に有意な差が確認されたモデルのうち，褒める人の言語的特徴量を用いたモデル（M59，M88，M115，M117）が存在していた．これらのモデルで共通している点は褒められる人の Praise シーン直後の言語的特徴量を用いていることである．しかし，褒める人の言語的特徴量と褒められる人の Praise シーン直後の言語的特徴量だけを

組み合わせたモデル M24 と M29 との間には有意な差が確認されていない．そのため，この 2 つの特徴量に加えて他の特徴量を組み合わせることによる影響が考えられる．

先行研究 [31] では，褒める人の視覚的，韻律的特徴量と，褒められる人の視覚的特徴量を用いること（これらの特徴量を用いたモデルは，本論文では M30 である）が褒め方の上手さを予測するために重要であることを示していた．しかし，M29 には褒められる人の特徴量が用いられていないという結果となった．M30 に言語的特徴量を追加したモデルは M64 である．M30 と M64 との間に 5%水準で有意差が確認できた．このことから，M30 に言語的特徴量を組み合わせることが褒め方の上手さを予測するために重要であることが確認できた．一方，M29 と M30 との間にも 5%水準で有意差が確認できた．よって，褒める人の視覚的・韻律的特徴量は褒める人の言語的特徴量と組み合わせることが特に重要であると考えられる．このことは，上手く褒めるためには褒める人が言語・非言語行動を複雑に組み合わせて振舞う必要があることを示唆している．

表 5.4: 対面対話コーパスを用いた機械学習モデルの性能

	褒める人			褒められる人				F 値		褒める人			褒められる人				F 値		褒める人			褒められる人				F 値
	非言語		言語	非言語		言語				非言語		言語	非言語		言語				非言語		言語	非言語		言語		
	視覚	韻律	言語	視覚	韻律	言語 (B)	言語 (A)			視覚	韻律	言語	視覚	韻律	言語 (B)	言語 (A)			視覚	韻律	言語	視覚	韻律	言語 (B)	言語 (A)	
Baseline								0.346*	M43	✓					✓	✓	0.367*	M86		✓	✓	✓			✓	0.496
M1	✓							0.457*	M44		✓	✓	✓				0.551	M87		✓	✓		✓	✓		0.52
M2		✓						0.382*	M45		✓	✓		✓			0.514	M88		✓	✓		✓		✓	0.482*
M3			✓					0.518	M46		✓	✓			✓		0.521	M89		✓	✓			✓	✓	0.504
M4				✓				0.423*	M47		✓	✓				✓	0.512	M90		✓		✓	✓	✓		0.492
M5					✓			0.493	M48		✓		✓	✓			0.497	M91		✓		✓	✓		✓	0.486*
M6						✓		0.326*	M49		✓		✓		✓		0.375*	M92		✓		✓		✓	✓	0.366*
M7							✓	0.336*	M50		✓		✓			✓	0.364*	M93		✓			✓	✓	✓	0.46*
M8	✓	✓						0.412*	M51		✓			✓	✓		0.481*	M94			✓	✓	✓	✓		0.509
M9	✓		✓					0.547	M52		✓			✓		✓	0.471*	M95			✓	✓	✓		✓	0.503
M10	✓			✓				0.475*	M53		✓				✓	✓	0.357*	M96			✓	✓		✓	✓	0.522
M11	✓				✓			0.49	M54			✓	✓	✓			0.535	M97			✓		✓	✓	✓	0.499
M12	✓					✓		0.368*	M55			✓	✓		✓		0.508	M98				✓	✓	✓	✓	0.497
M13	✓						✓	0.375*	M56			✓	✓			✓	0.516	M99	✓	✓	✓	✓	✓			0.522
M14		✓	✓					0.538	M57			✓		✓	✓		0.517	M100	✓	✓	✓	✓		✓		0.521
M15		✓		✓				0.395*	M58			✓		✓		✓	0.5	M101	✓	✓	✓	✓			✓	0.523
M16		✓			✓			0.501	M59			✓			✓	✓	0.474*	M102	✓	✓	✓		✓	✓		0.502
M17		✓				✓		0.375*	M60				✓	✓	✓		0.465*	M103	✓	✓	✓		✓		✓	0.516
M18		✓					✓	0.361*	M61				✓	✓		✓	0.484*	M104	✓	✓	✓			✓	✓	0.5
M19			✓	✓				0.546	M62				✓		✓	✓	0.324*	M105	✓	✓		✓	✓	✓		0.502
M20			✓		✓			0.525	M63					✓	✓	✓	0.457*	M106	✓	✓		✓	✓		✓	0.499
M21			✓			✓		0.514	M64	✓	✓	✓	✓				0.539	M107	✓	✓		✓		✓	✓	0.379*
M22			✓				✓	0.503	M65	✓	✓	✓		✓			0.524	M108	✓	✓			✓	✓	✓	0.497
M23				✓	✓			0.502	M66	✓	✓	✓			✓		0.506	M109	✓		✓	✓	✓	✓		0.513
M24				✓		✓		0.343*	M67	✓	✓	✓				✓	0.516	M110	✓		✓	✓	✓		✓	0.497
M25				✓			✓	0.354*	M68	✓	✓		✓	✓			0.507	M111	✓		✓	✓		✓	✓	0.497
M26					✓	✓		0.477*	M69	✓	✓		✓		✓		0.389*	M112	✓		✓		✓	✓	✓	0.489
M27					✓		✓	0.482*	M70	✓	✓		✓			✓	0.385*	M113	✓			✓	✓	✓	✓	0.5
M28						✓	✓	0.342*	M71	✓	✓			✓	✓		0.486*	M114		✓	✓	✓	✓	✓		0.509
M29	✓	✓	✓					0.556	M72	✓	✓			✓		✓	0.494	M115		✓	✓	✓	✓		✓	0.487*
M30	✓	✓		✓				0.407*	M73	✓	✓				✓	✓	0.38*	M116		✓	✓	✓		✓	✓	0.498
M31	✓	✓			✓			0.502	M74	✓		✓	✓	✓			0.53	M117		✓	✓		✓	✓	✓	0.488*
M32	✓	✓				✓		0.388*	M75	✓		✓	✓		✓		0.519	M118		✓		✓	✓	✓	✓	0.49
M33	✓	✓					✓	0.39*	M76	✓		✓	✓			✓	0.488	M119			✓	✓	✓	✓	✓	0.493
M34	✓		✓	✓				0.541	M77	✓		✓		✓	✓		0.529	M120	✓	✓	✓	✓	✓	✓		0.508
M35	✓		✓		✓			0.508	M78	✓		✓		✓		✓	0.505	M121	✓	✓	✓	✓	✓		✓	0.507
M36	✓		✓			✓		0.514	M79	✓		✓			✓	✓	0.52	M122	✓	✓	✓	✓		✓	✓	0.51
M37	✓		✓				✓	0.492	M80	✓			✓	✓	✓		0.494	M123	✓	✓	✓		✓	✓	✓	0.5
M38	✓			✓	✓			0.506	M81	✓			✓	✓		✓	0.485*	M124	✓	✓		✓	✓	✓	✓	0.471*
M39	✓			✓		✓		0.379*	M82	✓			✓		✓	✓	0.354*	M125	✓		✓	✓	✓	✓	✓	0.506
M40	✓			✓			✓	0.381*	M83	✓				✓	✓	✓	0.483*	M126		✓	✓	✓	✓	✓	✓	0.502
M41	✓				✓	✓		0.472*	M84		✓	✓	✓	✓			0.504	M127	✓	✓	✓	✓	✓	✓	✓	0.52
M42	✓				✓		✓	0.473*	M85		✓	✓	✓		✓		0.524									

*M29 との間に有意差（p <.05）が認められたモデル.

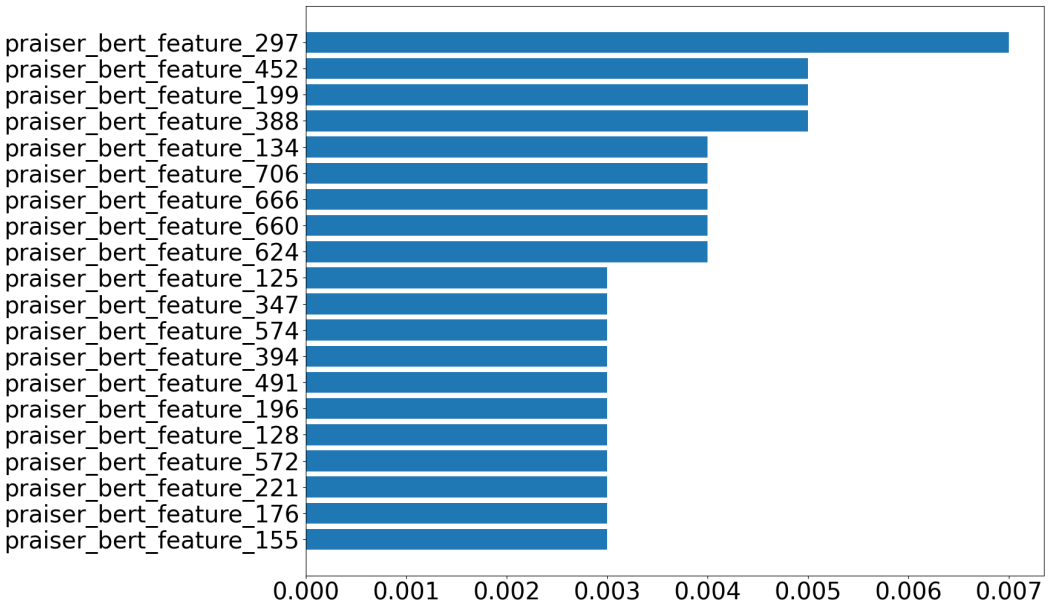

図 5.6: M29 における重要度の高い特徴量上位 20 件

第6章　遠隔対話コーパスにおける上手く褒めるための振る舞いの分析

本章では，褒め方の上手さを予測する機械学習モデルの構築について述べる．具体的には，4 章で構築した対話コーパスをから，言語・非言語行動に関する特徴量を抽出し，抽出した特徴量を用いて褒め方の上手さを評価した値を推定する機械学習モデルの構築を行った．構築する褒め方の上手さ推定モデルの概要を図 6.1 に示す．

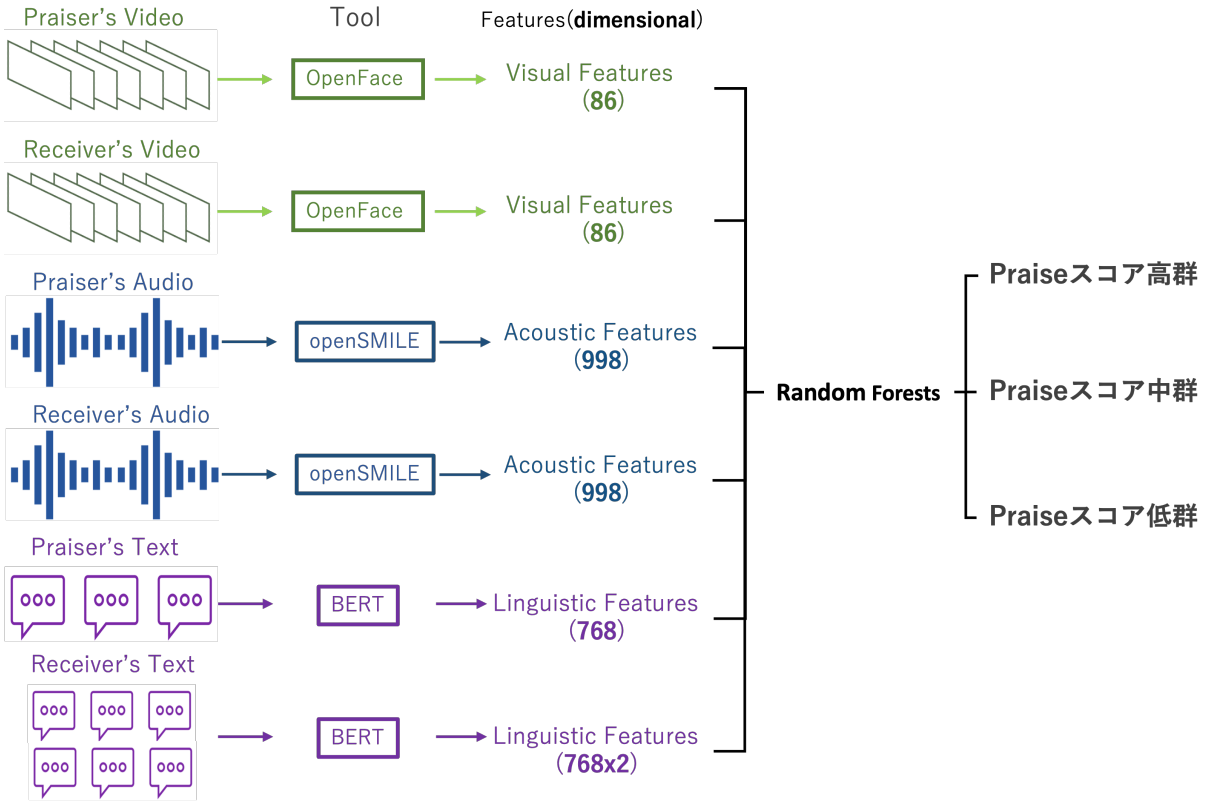

図 6.1: モデルの概要

6.1　特徴量の抽出

本節では，機械学習モデルを構築するために使用する特徴量の抽出について述べる．本研究では，2.2 節にて述べた研究を参考に，視覚的，韻律的，言語的な行動に関する特徴量を用いた．

6.1.1　視覚的特徴量

本項では，視覚的行動に関する特徴量の抽出について述べる．遠隔対話コーパスにおける視覚的特徴量は，対面対話コーパスと同様の手法で抽出された．顔画像処理ツールである OpenFace[42] を用いて，テーブルに設置された PC に内蔵されているカメラで撮影した映像データから，各話者の頭部・顔部の振舞いに関連する特徴量と AU に関する

特徴量の抽出を行った．頭部の向きに関しては，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（図 5.2）における褒める人と褒められる人の頭部の 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，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．AU 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 AU（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．

6.1.2 韻律的特徴量

本項では，韻律的行動に関する特徴量の抽出について述べる．遠隔対話コーパスにおける韻律的特徴量は，対面対話コーパスと同様の手法で抽出された．音声情報処理ツールである openSMILE[44] を用いて，テーブルに設置された PC に内蔵されている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．Praise シーンの範囲（図 5.3）における褒める人と褒められる人の表 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 [45][46] として用いた．

6.1.3 言語的特徴量

本項では，言語的行動に関する特徴量の抽出について述べる．遠隔対話コーパスにおける言語的特徴量は，対面対話コーパスと同様の手法で抽出された．日本語事前学習済みの BERT モデル [47] を利用して Praise シーンの褒める人の発話内容と Praise シーンの直前の褒められる人の発話内容（B)，Praise シーンの直後の褒められる人の発話内容（A)（図 5.4）をそれぞれ 768 次元のベクトル（文章埋め込み）に変換する方法を用いた．BERT とは，大規模コーパスを用いて事前学習したモデルを転移学習により様々なタスクに応用できる言語モデルである．具体的な抽出方法は，入力した文章を BERT が処理した最終層の先頭にある特殊トークンである [CLS] の埋め込み表現を取得し言語的特徴量とした（図 5.5）．

6.2 褒め方の上手さ予測モデル構築

本節では，機械学習モデルの構築について述べる．6.1 節にて抽出した特徴量を用いて 4.2.2 項にて評価した褒め方の上手さの評価値を予測する機械学習モデルを構築した．本研究では機械学習モデルを構築することで，褒め方の上手さの予測に寄与する特徴量の組

み合わせから，褒め方の上手さに重要な行動を明らかにする．機械学習モデルを構築するにあたり，既存研究 [48][49][50] を参考に Praise シーンを Praise スコア低群，中群，高群の 3 クラスに分割した．遠隔対話コーパスにおける各群に属する Praise シーン数（全 236 シーン）ができるだけ均等になるように，下記のように Praise スコア低〜高群を定義した* ．

Praise スコア低群 Praise スコアが 3.8 点以下の Praise シーン（計 86 シーン）

Praise スコア中群 Praise スコアが 3.8 点より大きく 4.6 点未満の Praise シーン（計 62 シーン）

Praise スコア高群 Praise スコアが 4.6 点以上の Praise シーン（計 88 シーン）

本論文ではモデルを構築するアルゴリズムとしてアンサンブル学習のバギング手法である Random Forest[51] を利用し，分類モデルを構築した．具体的には，目的変数を Praise スコア低群，中群，高群の 3 クラスとし，説明変数を 6.1 節で抽出した特徴量とする分類モデルを構築した．決定木の木の本数や木の深さといったハイパーパラメータは Hyperopt[52] を用いて最適化した．ハイパーパラメータのチューニングは 90%のトレーニングデータと 10%のテストデータに分割されたデータを使用して 1000 回行った．1000 回の試行結果に基づいて，最も性能の良いパラメータを用いて機械学習モデルを構築した．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，褒め方の上手さを推定するために寄与するモダリティを明らかにするために，各モダリティの組み合わせを変えて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）〜（2）を 100 回ずつ行った．

(1) 特徴量が含まれている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．

(2) 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．

このとき，ベースライン（チャンスレベル）は，データセットにおける各群の割合にあわせて，37%，26%，37%の確率で Praise スコアの低，中，高群を出力するモデルを用いた．

6.3 遠隔対話の分析

本節では遠隔対話コーパスから 6.1 節にて抽出した特徴量を用いて 6.2 節にて構築した機械学習モデル結果について述べる．構築した分類モデルの推定結果を表 6.1 に示す．加えて，各モデルの F 値の組に対して有意水準 5%のマン・ホイットニーの U 検定を行い，多重検定の問題を回避するために p 値をホルム法で補正した．

*各群に属するシーン数は等しいことが理想ではあるが，スコアが同じシーンが多数存在したため，各群のシーン数を等しくすることができなかった．

単一の特徴量のみを用いた機械学習モデル（M1-M7）の結果を確認する．M6，M7 を除く全てのモデルとベースラインとの間に 5%水準で有意な差が確認できた．また，M4，M6，M7 を除く全てのモデルがベースラインよりも高い F 値となった．このことは，褒め方の上手さを予測するためには言語・非言語行動が重要であることを示唆している．最も F 値の高いモデルは，褒める人の韻律的特徴量を用いたモデル（M2）であった．M2 と M1，M3，M4，M6，M7 との間に 5%水準で有意な差が確認できた．このことから，褒める人の韻律的行動が他の行動に比べて褒め方の上手さに関係していると考えられる．

複数の特徴量を用いた機械学習モデル（M8-M127）の結果を確認する．M12，M13，M24，M25，M28，M39，M40，M43，M62，M82，M113 を除く全てのモデルがベースラインとの間に 5%水準で有意な差が確認できた．また，M24，M25，M28，M62 を除くすべてのモデルがベースラインよりも高い F 値となった．このことは，褒め方の上手さを予測するためには複数の言語・非言語行動を組み合わせることが重要であることが示唆している．最も F 値の高いモデルは，褒める人の韻律的特徴量と褒められる人の視覚的特徴量を用いたモデル (M15) であった．M15 との間に 5%水準で有意な差が確認できたモデルに共通していることは，褒める人の韻律的特徴量を使用していないという点である．褒める人の韻律的行動は，F 値が高い機械学習モデル上位 5 位の全てにおいて使用されている．さらに，褒める人の韻律的行動と他の行動を組み合わせて構築した機械学習モデル 64 件中，14 件のモデルが褒める人の韻律的行動のみを用いた機械学習モデルよりも高い F 値であった．褒め方の上手さを予測するためには，褒める人の韻律的行動と他の行動を組み合わせることが重要であると考えられる．加えて，褒められる人の視覚的行動は，単一で用いた場合の機械学習モデル（M4）の F 値は 9 位であったが，褒める人の韻律的行動と組み合わせることで最高性能となった．また，M15 と M4 との間に有意な差が確認できた．このことより，褒められる人の視覚的特徴量は褒める人の韻律的特徴量と組み合わせることで褒め方の上手さを予測するための F 値を向上させることが確認できた．加えて，M15 における重要度の高い特徴量上位 20 件を図 6.2 に示す．図 6.2 より，重要度の高い特徴量上位 20 件は全て褒める人の韻律的特徴量であった．そのため，褒め方の上手さの予測において特に重要な特徴量は褒める人の韻律的特徴量であると考えられる．

表 6.1: 遠隔対話コーパスを用いた機械学習モデルの性能

	褒める人			褒められる人				F 値		褒める人			褒められる人				F 値		褒める人			褒められる人				F 値
	非言語		言語	非言語		言語				非言語		言語	非言語		言語				非言語		言語	非言語		言語		
	視覚	韻律	言語	視覚	韻律	言語 (B)	言語 (A)			視覚	韻律	言語	視覚	韻律	言語 (B)	言語 (A)			視覚	韻律	言語	視覚	韻律	言語 (B)	言語 (A)	
Baseline								0.346†	M43	✓					✓	✓	0.358†	M86		✓	✓	✓			✓	0.519
M1	✓							0.453†	M44		✓	✓	✓				0.529	M87		✓	✓		✓	✓		0.508
M2		✓						0.532	M45		✓	✓		✓			0.521	M88		✓	✓		✓		✓	0.513
M3			✓					0.454†	M46		✓	✓			✓		0.531	M89		✓	✓			✓	✓	0.519
M4				✓				0.363†	M47		✓	✓				✓	0.499	M90		✓		✓	✓	✓		0.492
M5					✓			0.473†	M48		✓		✓	✓			0.538	M91		✓		✓	✓		✓	0.492
M6						✓		0.32†	M49		✓		✓		✓		0.519	M92		✓		✓		✓	✓	0.512
M7							✓	0.327	M50		✓		✓			✓	0.515	M93		✓			✓	✓	✓	0.512
M8	✓	✓						0.553	M51		✓			✓	✓		0.532	M94			✓	✓	✓	✓		0.469†
M9	✓		✓					0.47†	M52		✓			✓		✓	0.516	M95			✓	✓	✓		✓	0.464†
M10	✓			✓				0.446†	M53		✓				✓	✓	0.533	M96			✓	✓		✓	✓	0.465†
M11	✓				✓			0.493	M54			✓	✓	✓			0.49	M97			✓		✓	✓	✓	0.454†
M12	✓					✓		0.386†	M55			✓	✓		✓		0.458†	M98				✓	✓	✓	✓	0.447†
M13	✓						✓	0.391†	M56			✓	✓			✓	0.464†	M99	✓	✓	✓	✓	✓			0.532
M14		✓	✓					0.548	M57			✓		✓	✓		0.455†	M100	✓	✓	✓	✓		✓		0.531
M15		✓		✓				0.555	M58			✓		✓		✓	0.453†	M101	✓	✓	✓	✓			✓	0.521
M16		✓			✓			0.538	M59			✓			✓	✓	0.44†	M102	✓	✓	✓		✓	✓		0.532
M17		✓				✓		0.524	M60				✓	✓	✓		0.433†	M103	✓	✓	✓		✓		✓	0.509
M18		✓					✓	0.521	M61				✓	✓		✓	0.444†	M104	✓	✓	✓			✓	✓	0.505
M19			✓	✓				0.449†	M62				✓		✓	✓	0.338†	M105	✓	✓		✓	✓	✓		0.533
M20			✓		✓			0.501	M63					✓	✓	✓	0.452†	M106	✓	✓		✓	✓		✓	0.532
M21			✓			✓		0.434†	M64	✓	✓	✓	✓				0.545	M107	✓	✓		✓		✓	✓	0.528
M22			✓				✓	0.444†	M65	✓	✓	✓		✓			0.535	M108	✓	✓			✓	✓	✓	0.498
M23				✓	✓			0.491	M66	✓	✓	✓			✓		0.513	M109	✓		✓	✓	✓	✓		0.454†
M24				✓		✓		0.337†	M67	✓	✓	✓				✓	0.518	M110	✓		✓	✓	✓		✓	0.458†
M25				✓			✓	0.311†	M68	✓	✓		✓	✓			0.524	M111	✓		✓	✓		✓	✓	0.442†
M26					✓	✓		0.436†	M69	✓	✓		✓		✓		0.522	M112	✓		✓		✓	✓	✓	0.466†
M27					✓		✓	0.441†	M70	✓	✓		✓			✓	0.523	M113	✓			✓	✓	✓	✓	0.42†
M28						✓	✓	0.343†	M71	✓	✓			✓	✓		0.51	M114		✓	✓	✓	✓	✓		0.517
M29	✓	✓	✓					0.546	M72	✓	✓			✓		✓	0.521	M115		✓	✓	✓	✓		✓	0.536
M30	✓	✓		✓				0.549	M73	✓	✓				✓	✓	0.526	M116		✓	✓	✓		✓	✓	0.516
M31	✓	✓			✓			0.546	M74	✓		✓	✓	✓			0.497	M117		✓	✓		✓	✓	✓	0.511
M32	✓	✓				✓		0.526	M75	✓		✓	✓		✓		0.465†	M118		✓		✓	✓	✓	✓	0.505
M33	✓	✓					✓	0.534	M76	✓		✓	✓			✓	0.457†	M119			✓	✓	✓	✓	✓	0.466†
M34	✓		✓	✓				0.468†	M77	✓		✓		✓	✓		0.469†	M120	✓	✓	✓	✓	✓	✓		0.516
M35	✓		✓		✓			0.501	M78	✓		✓		✓		✓	0.449†	M121	✓	✓	✓	✓	✓		✓	0.51
M36	✓		✓			✓		0.466†	M79	✓		✓			✓	✓	0.452†	M122	✓	✓	✓	✓		✓	✓	0.527
M37	✓		✓				✓	0.449†	M80	✓			✓	✓	✓		0.463†	M123	✓	✓	✓		✓	✓	✓	0.502
M38	✓			✓	✓			0.491	M81	✓			✓	✓		✓	0.439†	M124	✓	✓		✓	✓	✓	✓	0.492
M39	✓			✓		✓		0.372†	M82	✓			✓		✓	✓	0.375†	M125	✓		✓	✓	✓	✓	✓	0.486†
M40	✓			✓			✓	0.374†	M83	✓				✓	✓	✓	0.447†	M126		✓	✓	✓	✓	✓	✓	0.51
M41	✓				✓	✓		0.449†	M84		✓	✓	✓	✓			0.527	M127	✓	✓	✓	✓	✓	✓	✓	0.524
M42	✓				✓		✓	0.438†	M85		✓	✓	✓		✓		0.521									

† M15 との間に有意差（p <.05）が認められたモデル.

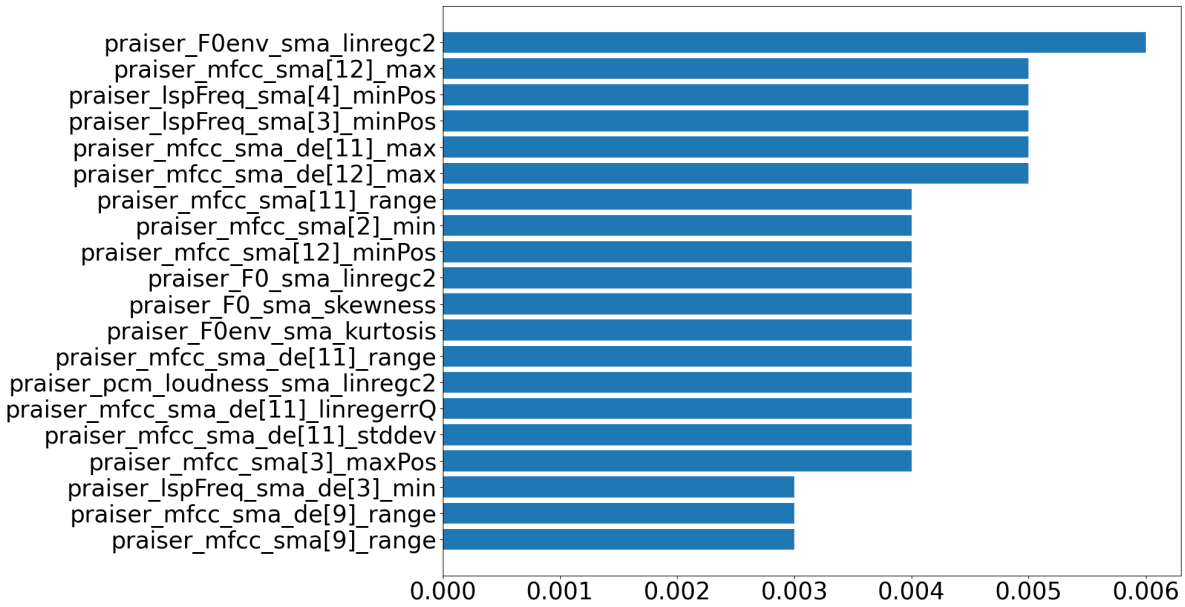

図 6.2: M15 における重要度の高い特徴量上位 20 件

第７章　対面対話と遠隔対話の振舞いの比較

本章では，5 章と 6 章で構築した機械学習モデルの結果から対面対話と遠隔対話における，上手く褒めるために重要な振舞いの相違点の分析を行う．

7.1　機械学習モデルの性能

本節では，対面・遠隔対話における構築した機械学習モデルの性能について分析する．対面・遠隔対話における最も性能の良かったモデルの confusion matrix を表 7.1 表 7.2 に示す．　表 7.1 表 7.2 より，低群と高群については，6 割を超える正答率となり，中群は 2 割を超える正答率であった．本論文の目的は，褒め方の上手さの分析であり，より正しく分類する必要があるのは低群と高群である．それらの群の正答率が 6 割を超えているため，本論文で構築した機械学習モデルの性能は十分な性能であると考える．また，各対話コーパスの最も性能良かったモデルの F 値は対面対話コーパスが F 値=0.556，遠隔対話コーパスが F 値=0.555 であった．これらの性能は，2.2 節で述べた関連研究で構築された機械学習モデルと同程度の性能である．マルチモーダルインタラクションの分野では，分析対象が人の能力やスキルであり，人間であっても正しく予測することは困難であるとされている．そのため本論文で構築した機械学習モデルの性能は，本分野の他の研究と比較しても妥当な性能であると考えられる．

表 7.1: 対面対話コーパスにおいて最も性能の良かったモデル (M29) の confusion matrix

		Pred		
		低群	中群	高群
Actual	低群	0.708	0.104	0.188
	中群	0.310	0.249	0.441
	高群	0.145	0.166	0.689

表 7.2: 遠隔対話コーパスにおいて最も性能の良かったモデル (M15) の confusion matrix

		Pred		
		低群	中群	高群
Actual	低群	0.695	0.172	0.133
	中群	0.335	0.289	0.375
	高群	0.134	0.196	0.670

低群と高群で正しく予測できなかった実際のデータを分析することで，各群の正答率を向上できるか考察する．高群であるが低群と予測したシーンについては，発話数が少ないが感情を込めて褒めているシーンが多く存在していた．一方で，低群であるが高群と予測

したシーンについては，発話数が多く，一定の声の高さで褒めているシーンが多く存在していた．これらから，上手く褒めるためには実際に賞賛したいという感情を込めて褒めているかどうかという要素も重要であると考えた．本研究で使用した，特徴量抽出方法ではこの要素を上手く捉えることができておらず，上記のような結果となったと考える．今後は，感情を込めた話し方ができているかという要素を考慮した褒め方の上手さの分析手法を検討していく必要がある．

7.2　振舞いの比較

本節では，対面・遠隔対話における褒め方の上手さの振舞いを比較する．5.3 節と 6.3 節より，重要な特徴量が異なっていることが確認できる．対面対話において褒め方の上手さと関連があると考えられる行動は，褒める人の視覚的，韻律的，言語的行動である．遠隔対話において褒め方の上手さと関連があると考えられる行動は，褒める人の韻律的行動と褒められる人の視覚的行動である．対面・遠隔対話のどちらにおいても褒める人の韻律的行動が重要であることが確認できた．このことから，上手く褒めるためには，対面・遠隔対話において褒める人の韻律的行動が重要であると考えられる．

一方で褒める人の視覚的，言語的行動と褒められる人の視覚的行動は，対話シーンによって褒め方の上手さと関連性が低い可能性がある．このような結果となった理由として，遠隔対話における言語・非言語情報の伝達の違いが考えられる．まず，視覚的行動については，遠隔対話では表情の表出が乏しくなる可能性が考えられる．その理由として，遠隔対話の際には対面対話とことなり対話相手が実空間上の正面に存在していないことが挙げられる．正面に対話相手がいないことから，自身が他人から視認されているという認識が低下する．それによって，自身の表情を変化させることやアイコンタクトをとろうとする行動が，減ってしまった可能性が考えられる．次に，対面対話において最も重要な特徴量であると考えられる言語的行動について遠隔対話において重要でないとされた理由の分析を行う．まず，非線形な関係を低い計算量で捉えることができる次元削減手法である UMAP[53] を用いることで，言語的特徴量の可視化を行った．言語的特徴量の次元を 768 次元から 2 次元に次元削減を行った．このとき UMAP のハイパーパラメータのチューニングは optuna[54] を用いた．次元削減した結果を図 7.1 に示す．図 7.1 より，言語的特徴量が各クラスをどの程度分割できているか確認する．対面対話コーパスでは，高群が中心から右下，低群が左上に分布する傾向がみえるが，中群は満遍なく分布しているようにみえる．遠隔対話コーパスでは，低群が右上に分布する傾向がみえるが，高群と中群は満遍なく広く分布しているようにみえる 次に，Praise スコアと各言語的特徴量との間の相関係数を算出した（表 7.3）．表 7.3 より，対面対話コーパスの方が相関のある言語的特徴量が多いことが確認できた．このことから，対面対話コーパスにおける言語的特徴量はある程度各クラスの特徴を捉えていると考えられる．一方で，遠隔対話コーパスにおける言語的特徴量は各クラスの特徴を上手く捉えられていなかったと考えられる．そして，実際の対話内容を確認することでより詳しい分析を行う．対面・遠隔対話における Praise スコア高群と低群の発話内容の単語数を算出した（図 7.2）．その結果，遠隔対話より対面対

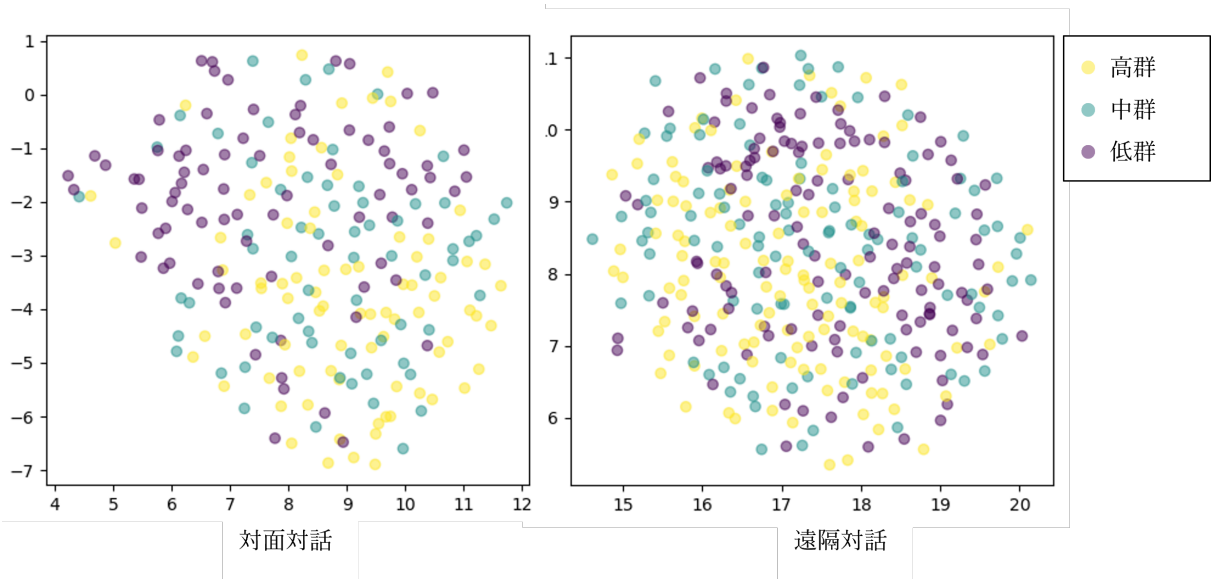

図 7.1: 対面・遠隔対話コーパスの褒める人の発話内容から抽出した言語的特徴量を UMAP を用いて 2 次元に削減した結果

表 7.3: 対面・遠隔対話コーパスにおける各言語的特徴量と Praise スコアとの間に相関関係が確認できた個数

	$0.2 \leq r < 0.4$	$0.4 \leq r$
対面対話コーパス	230	6
遠隔対話コーパス	192	1

話の方が発話内容の単語数が高群，低群どちらも多い傾向にあることがわかった．特に，遠隔対話の低群における発話内容の単語数は，ほとんどが 1，2 語程度あることがわかる．発話内容の単語数が少ないということは，褒める発話に含まれる情報が少なかったことが考えられる．このことは，遠隔対話では褒める際に短く端的に褒める傾向がある可能性がある．

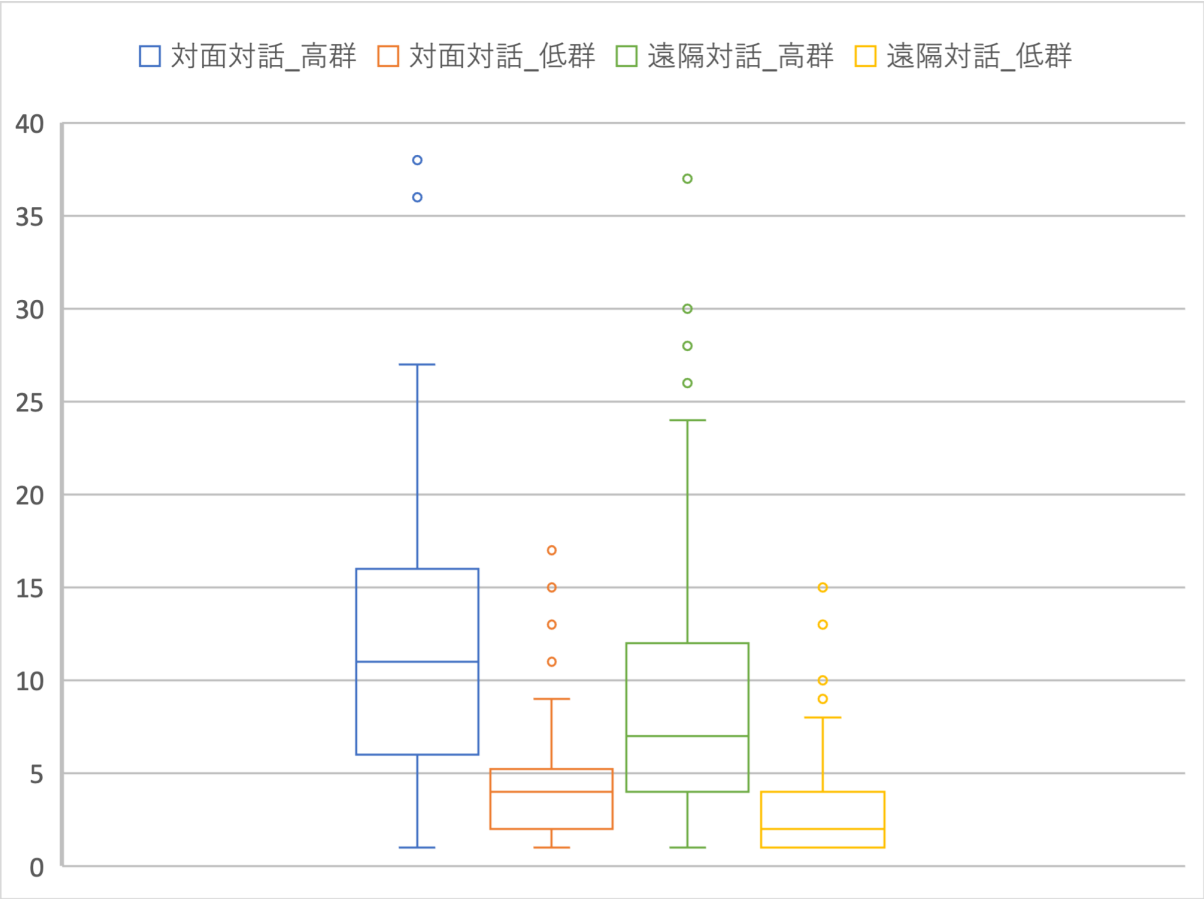

図 7.2: 対面・遠隔対話における Praise スコア高群と低群の発話内容の単語数

第8章　結論

本論文では，対面・遠隔対話における話者（褒める人，褒められる人）の言語・非言語行動から褒め方の上手さを推定することができるか明らかにする取り組みを行った．具体的には，対面・遠隔対話における話者（褒める人，褒められる人）の視覚的，韻律的，言語的特徴量を抽出して機械学習モデルを構築した．対面対話コーパスを用いた機械学習モデルは，褒める人の視覚的，韻律的，言語的特徴量を用いたモデルが最も性能が高く F 値 $= 0.556$ となった．特に上手く褒めるために重要であると考えられた行動は，褒める人の言語的な行動であった．遠隔対話コーパスを用いた機械学習モデルは，褒める人の韻律的特徴量と褒められる人の視覚的特徴量を用いたモデルが最も性能が高く F 値 $= 0.555$ となった．特に上手く褒めるために重要であると考えられた行動は，褒める人の韻律的な行動であった．

また，対面対話と遠隔対話における振舞いの比較も行った．対面対話と遠隔対話では重要な振舞いに相違点が確認できた．まず，共通して重要であると考えられた行動は，褒める人の韻律的な行動である．韻律的な行動とは，発話の際の声の大きさや声色といった話し方に関する行動である．次に，共通して重要であると考えられない行動は，褒める人の視覚的，言語的な行動と褒められる人の視覚的な行動である．対面対話においては，褒める人の行動が全て重要であり，褒める人の各行動を組み合わせた相互作用による影響も考えられた．一方で，遠隔対話においては，褒める人の行動と褒められる人の行動が部分的にだが重要であると考えられた．このことは，対面対話と遠隔対話では褒める行為に相違点があることを示唆している．対面対話と遠隔対話では，情報の伝達のされ方に差異があるとされており，それぞれの場面によって褒める行為に変化が生じているのではないかと考えられる．今後は，より詳細に対面対話と遠隔対話での上手く褒めるための振舞いの比較を行いたい．

本論文にはいくつかの限界がある．1 つ目が，対話コーパスの性質である．本論文では，2 種類の対話コーパスを使用した．これらの対話コーパスの参加者属性は異なっている．具体的には，対面対話コーパスでは 20 代の男女（男性 28 名，女性 6 名）であり，遠隔対話コーパスでは，20〜50 代の男女（男性 20 名，女性 20 名）であった．年齢や性別によって上手く褒める行動に相違点が生じている可能性が考えられる．これらの影響を排除するためには，年齢や性別を統制したコーパスを使用した上での分析を行う必要がある．その一方で，年齢や性別といった属性の一部を変化させたコーパスを複数用いることで，特定の年齢や性別固有の振舞いも明らかにできる可能性があると考えられる．加えて，参加者の褒め方の上手さに偏りがあった可能性がある．参加者の募集を行う際の条件に褒めが上手いかどうかといった，褒め方に関する条件を設定しなかった．そのため，褒め方が上手いとされている人だけを集めた場合の結果と本論文で得られた知見では異なる結果となる可能性がある．今後は，参加者自身の褒め方の上手さがどの程度なのかを考慮したコーパスを作成することで，本論文との知見の比較をすることが必要である．2 つ目が，使用した特徴量である．本論文では，既存研究を参考に，視覚的，韻律的，言語的な行動に関する特徴量を使用した．しかし，既存研究では，手や腕，体全体の動きを特徴量として使用している研究も存在している．日本人はジェスチャーが少ないとされているが，上手く褒めるための行動にはこれらも重要になってくる可能性がある．3 つ目が，特徴量の抽出

方法である．本論文で使用した特徴量の抽出方法は低次元のものと高次元のものがどちらも含まれている．また，現在は特徴量は発話シーンを最低単位としている．そのため，発話シーン中のダイナミズムを上手く取り扱えていない可能性がある．4 つ目が．時系列の考慮である．本論文では，時系列情報として，Praise シーンの直前と直後の褒められる人の発話内容に着目した．しかし，褒める人の Praise シーン以外の情報は取り扱えていない．特に Praise シーン同士の情報は褒め方の上手さと関係がある可能性が考えられる．具体的には，Praise シーンまでに何回 Praise シーンがあったのか，前回の Praise シーンとの振舞いの相違点といった情報である．同じような褒め方を連続して行うよりも，バラエティに富む褒め方を出来ているかが，褒め方の上手さに関連していると考えたからだ．

参考文献

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

[2] L.N. Jenkins, M.T. Floress, and W. Reinke. Rates and types of teacher praise: A review and future directions. *Psychology in the Schools*, Vol. 52, No. 5, pp. 463–476, 2015.

[3] V. M. Catano. Relation of improved performance through verbal praise to source of praise. *Perceptual and Motor Skills*, Vol. 41, No. 1, pp. 71–74, 1975.

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

[5] M. H. Nguyen, J. Gruber, J. Fuchs, W. Marler, A. Hunsaker, and E. Hargittai. Changes in digital communication during the covid-19 global pandemic: Implications for digital inequality and future research. *Social Media + Society*, 2020.

[6] S.G. Doherty, A. Anderson, C. O'malley, S. Langton, S. Garrod, and V. Bruce. 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.

[7] G. Sears, H. Zhang, W. Wiesner, R. Hackett, and Y. Yuan. A comparative assessment of videoconference and face-to-face employment interviews. *Management Decision*, Vol. 51, p. 1733–1752., 2013.

[8] J. Morton, M. Mikolajczak, and O. Luminet. New perspectives on the praise literature: towards a conceptual model of compliment. *Current Psychology*, Vol. 41, pp. 6038–6050, 2022.

[9] W. Reinke, T. Lewis, and K. Merrell. The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School psychology review*, Vol. 37, pp. 315–332, 2008.

[10] J. K. Luiselli and N. D. Joan. Improving a student's arithmetic performance using feedback and reinforcement procedures. *Education and Treatment of Children*, Vol. 3, No. 1, pp. 45–49, 1980.

[11] T. F. Margaret, L. B. Shelby, L. M. Kari, and M. R. Wendy. Praise research trends and future directions: Characteristics and teacher training. *Behavioral Disorders*, Vol. 43, pp. 227–243, 2017.

[12] M. Floress and S. Beschta. An analysis of general education teachers' use of diverse praise: Floress and beschta. Vol. 55, pp. 1188–1204, 2018.

[13] Robert K. H. Sex-based differences in compliment behavior. *Language in Society*, Vol. 19, pp. 201–224, 1990.

[14] Z. Tan, A. Goel, T. Nguyen, and D. Ong. A multimodal lstm for predicting listener empathic responses over time. In *14th IEEE International Conferenceon Automatic Face and Gesture Recognition (FG '19)*, pp. 1–4, 2019.

[15] V. P. Rosas, R. Mihalcea, and L. P. Morency. Multimodal sentiment analysis of spanish online videos. *IEEE Intelligent Systems*, Vol. 28, No. 3, pp. 38–45, 2013.

[16] V. Lin, J. M. Girard, M. A. Sayette, and L. Morency. Toward multimodal modeling of emotional expressiveness. In *Proceedings of the 22nd ACM International Conference on Multimodal Interaction (ICMI '20)*, pp. 548–557, 2020.

[17] C. Mawalim, S. Okada, Y. Nakano, and M. Unoki. Personality trait estimation in group discussions using multimodal analysis and speaker embedding. *Journal on Multimodal User Interfaces*, Vol. 17, pp. 1–17, 2023.

[18] S. Park, H.S. Shim, M. Chatterjee, K. Sagae, and L.P. Morency. Computational analysis of persuasiveness in social multimedia: A novel dataset and multimodal prediction approach. In *Proceedings of the 16nd ACM International Conference on Multimodal Interaction (ICMI '14)*, pp. 50–57, 2014.

[19] M. Soleymani, K. Stefanov, H.S. Kang, J. Ondras, and J. Gratch. Multimodal analysis and estimation of intimate self-disclosure. In *Proceedings of the 21st ACM International Conference on Multimodal Interaction (ICMI '19)*, pp. 59–68, 2019.

[20] I. Naim, M.I. Tanveer, D. Gildea, and M. E. Hoque. Automated prediction and analysis of job interview performance: The role of what you say and how you say it. In *11th IEEE International Conferenceon Automatic Face and Gesture Recognition (FG '15)*, pp. 1–6, 2015.

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

[22] S. Rasipuram, P. Rao, and D.B. Jayagopi. Asynchronous video interviews vs. face-to-face interviews for communication skill measurement: A systematic study. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16)*, pp. 370–377, 2016.

[23] V. Ramanarayanan, C. W. Leong, L. Chen, G. Feng, and D. Suendermann-Oeft. Evaluating speech, face, emotion and body movement time-series features for automated multimodal presentation scoring. In *Proceedings of the 17nd ACM International Conference on Multimodal Interaction (ICMI '15)*, pp. 23–30, 2015.

[24] Y. Yagi, S. Okada, S. Shiobara, and S. Sugimura. Predicting multimodal presentation skills based on instance weighting domain adaptation. *Journal on Multimodal User Interfaces*, Vol. 16, pp. 1–16, 2021.

[25] J. Schneider, D. Börner, P. V. Rosmalen, and M. Specht. Presentation trainer, your public speaking multimodal coach. In *Proceedings of the 17nd ACM International Conference on Multimodal Interaction (ICMI '15)*, pp. 539–546, 2015.

[26] L. Chen, G. Feng, J. Joe, C. W. Leong, C. Kitchen, and C. M. Lee. Towards automated assessment of public speaking skills using multimodal cues. In *Proceedings of the 16nd ACM International Conference on Multimodal Interaction (ICMI '14)*, pp. 200–203, 2014.

[27] T. Saga, H. Tanaka, Y. Matsuda, T. Morimoto, M. Uratani, K. Okazaki, Y. Fujimoto, and S. Nakamura. Automatic evaluation-feedback system for automated social skills training. *Scientific Reports*, 2023.

[28] T. Saga, H. Tanaka, H. Iwasaka, Y. Matsuda, T. Morimoto, M. Uratani, K. Okazaki, Y. Fujimoto, and S. Nakamura. Multimodal dataset of social skills training in natural conversational setting. In *Proceedings of the 23nd ACM International Conference on Multimodal Interaction (ICMI '21)*, pp. 395–399, 2021.

[29] T.M. Kalis, K.J. Vannest, and R Parker. Praise counts: Using self-monitoring to increase effective teaching practices. *Preventing School Failure*, Vol. 51, No. 3, pp. 20–27, 2007.

[30] T. Onishi, A. Yamauchi, R. Ishii, Y. Aono, and A. Miyata. Analyzing nonverbal behaviors along with praising. In *Proceedings of the 22nd ACM International Conference on Multimodal Interaction (ICMI '20)*, pp. 609–613, 2020.

[31] T. Onishi, A. Yamauchi, A. Ogushi, R. Ishii, A. Fukayama, T. Nakamura, and A. Miyata. Modeling japanese praising behavior by analyzing audio and visual behaviors. *Frontiers in Computer Science*, 2022.

[32] T. Onishi, A. Ogushi, Y. Tahara, R. Ishii, A. Fukayama, T. Nakamura, and A. Miyata. A comparison of praising skills in face-to-face and remote dialogues. In *Proceedings of the 13th International Conference on Language Resources and Language Evaluation (LREC '22)*, pp. 5805–5812, 2022.

[33] 山内愛里沙, 大西俊輝, 武藤佑太, 石井亮, 青野裕司, 宮田章裕. 表情・音声を用いた褒め方の上手さを評価するシステムの基礎検討. 情報処理学会インタラクション 2020 論文集, 第 2020 巻, pp. 247–252, 2020.

[34] 山内愛里沙, 大西俊輝, 呉健朗, 武藤佑太, 石井亮, 青野裕司, 宮田章裕. 音声および視線・表情・頭部運動に基づく上手い褒め方の評価システムの検討. 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 '20), 第 2020 巻, pp. 98–106, 2020.

[35] H. Brugman and A. Russel. Annotating multimedia / multi-modal resources with elan. In *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC '04)*, pp. 2065–2068, 2004.

[36] H. Mori, T. Satake, M. Nakamura, and H. Kasuya. Constructing a spoken dialogue corpus for studying paralinguistic information in expressive conversation and analyzing its statistical/acoustic characteristics. *Speech Communication*, Vol. 53, pp. 36–50, 2011.

[37] The UniDic Consortium. unidic-csj-3.0.1.1.

[38] K. Komatani, R. Takeda, and S. Okada. Analyzing differences in subjective annotations by participants and third-party annotators in multimodal dialogue corpus. In *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL '23)*, pp. 104–113, 2023.

[39] P. E. Shrout and J. L. Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, Vol. 86, No. 2, pp. 420–428, 1979.

[40] T Koo and M Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, Vol. 15, pp. 155–163, 2016.

[41] Zoom Video Communications Inc. Zoom cloud meetings. `https://zoom.us/`.

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

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

[44] F. Eyben, M. Wöllmer, and B. Schuller. opensmile - the munich versatile and fast open-source audio feature extractor. In *Proceedings of ACM Multimedia*, pp. 1459–1462, 2010.

[45] B. Schuller, S. Steidl, and A. Batliner. The interspeech 2009 emotion challenge. In *Proceedings of Interspeech '09*, pp. 312–315, 2009.

[46] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, and S. Narayanan. The interspeech 2010 paralinguistic challenge. In *Proceedings of Interspeech '10*, pp. 2794–2797, 2010.

[47] J. Devlin, M. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT '19)*, pp. 4171–4186, 2019.

[48] B. Lepri, J. Staiano, R. Subramanian, F. Pianesi, K. Kalimeri, and N. Sebe. Employing social gaze and speaking activity for automatic determination of the extraversion trait. In *International conference on multimodal interfaces and the workshop on machine learning for multimodal interaction*, pp. 1–8, 2010.

[49] A. Kindiroglu, L. Akarun, and O. Aran. Multi-domain and multi-task prediction of extraversion and leadership from meeting videos. *EURASIP Journal on Image and Video Processing*, Vol. 2017, pp. 1–14, 2017.

[50] I. Naim, M. Tanveer, D. Gildea, and E. Hoque. Automated prediction and analysis of job interview performance: The role of what you say and how you say it. In *11th IEEE International Conferenceon Automatic Face and Gesture Recognition (FG '15)*, pp. 1–6, 2015.

[51] L. Breiman. Random forests. *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001.

[52] J. Bergstra, D. Yamins, and D.D. Cox. Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. In *Proceedings of the 12th Python in Science Conferences (SciPy '13)*, pp. 13–20, 2013.

[53] M. Leland, H. John, and M. James. Umap: Uniform manifold approximation and projection for dimension reduction, 2020.

[54] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework, 2019.

研究業績

査読付き論文誌

(1) Kenro Go, Shunsuke Tokuda, Haruna Niiyama, Toshiki Onishi, Asahi Ogushi, and Akihiro Miyata: Evaluation of Prosodic Features Suitable for Conversational Agents Replying with a Joke. Journal of Information Processing, Vol.32 (2024 年 1 月).

(2) Kenro Go, Toshiki Onishi, Asahi Ogushi, Shunsuke Tokuda, and Akihiro Miyata: Evaluation of Motivation for Conversational Agents Replying to Manzai-Style Jokes. Transactions of the Virtual Reality Society of Japan, Vol. 28. No. 1, pp.43–53 (2023 年 3 月).

(3) Toshiki Onishi, Arisa Yamauchi, Asahi Ogushi, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata: Modeling Japanese Praising Behavior by Analyzing Audio and Visual Behaviors. Frontiers in Computer Science (2022 年 3 月).

(4) 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・顔部の振舞いの分析, 情報処理学会論文誌, Vol.62, No.9, pp.1620–1628 (2021 年 9 月).

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).

(3) Kenro Go, Toshiki Onishi, Asahi Ogushi and Akihiro Miyata:Conversational Agents Replying with a Manzai-style Joke. Proceedings of the 31st Australian Conference on Human-computer-interaction (OzCHI '21), pp.221–231 (2021 年 12 月).

査読付き国内会議

(1) 新山はるな, 呉健朗, 大西俊輝, 大串旭, 宮田章裕: 2 体の柔らかい物体の動きによる共感表現方法の検証. 情報処理学会コラボレーションとネットワークサービスワークショップ 2023 論文集, Vol.2023, pp.24–32 (2023 年 11 月).

(2) 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, Vol.2022, pp36–43 (2022 年 11 月).

(3) 呉健朗, 大西俊輝, 大串旭, 宮田章裕: ノリツッコミを行う対話型エージェント. 情報処理学会インタラクション 2022 論文集, pp.39–47 (2022 年 2 月).

(4) 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).

(5) 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) 大串旭, 大西俊輝, 石井亮, 深山篤, 宮田章裕: 表情特徴量に着目した褒め方の上手さフィードバックシステムの基礎検討. 情報処理学会インタラクション 2024 論文集 (2024 年掲載予定).

(2) 渡辺好汰, 丸山葉, 大西俊輝, 大串旭, 呉健朗, 大澤正彦, 宮田章裕: 対話型エージェントが意見伝達を仲介する際に適した音声の高さに関する調査. 情報処理学会インタラクション 2024 論文集 (2024 年掲載予定).

(3) 渡貫健太, 呉健朗, 新山はるな, 大西俊輝, 大串旭, 宮田章裕: 語尾でボケて返す対話型エージェントの実装. 情報処理学会インタラクション 2024 論文集 (2024 年掲載予定).

(4) 丸山葉，大西俊輝，大串旭，木下峻一，石井亮，深山篤，大澤正彦，宮田章裕: デジタルツインを用いた自己効力感向上システムの基礎検討. 情報処理学会コラボレーションとネットワークサービスワークショップ 2023 論文集, Vol.2023, pp.88–89 (2023 年 11 月).

(5) 呉健朗，渡貫健太，新山はるな，大西俊輝，大串旭，宮田章裕: 語尾でボケて返す対話型エージェントの基礎検討. 情報処理学会コラボレーションとネットワークサービスワークショップ 2023 論文集, Vol.2023, pp.152–153 (2023 年 11 月).

(6) 新山はるな, 呉健朗, 大西俊輝, 大串旭, 宮田章裕: ユーモラスに話題提供を行うエージェントの検証. 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 ’23), Vol.2023, pp.700–707 (2023 年 7 月).

(7) 大西俊輝, 新山はるな, 丸山葉, 大串旭, 呉健朗, 宮田章裕: 柔らかい物体の動きによる共感表現方法に及ぼす共感スキルの影響調査. 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 '23), Vol.2023, pp.708–7015 (2023 年 7 月).

(8) 新山はるな, 得田舜介, 大串旭, 大西俊輝, 呉健朗, 大澤正彦, 宮田章裕: ユーモラスに話題提供を行うエージェントの基礎検討, 情報処理学会インタラクション 2023 論文集, pp.447–450 (2023 年 3 月).

(9) 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: 遠隔対話における上手い褒め方のモデリングの検討, 情報処理学会インタラクション 2023 論文集, pp.749–754 (2023 年 3 月).

(10) 丸山葉, 大西俊輝, 大串旭, 呉健朗, 大澤正彦, 宮田章裕: 意見伝達を仲介する対話型エージェントに対する利用意欲の調査. 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.369-–373 (2022 年 7 月).

(11) 得田舜介, 呉健朗, 大西俊輝, 大串旭, 宮田章裕: 対話型エージェントのユーモア表現に適した韻律的特徴の調査. 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.570–575 (2022 年 7 月).

(12) 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).

(13) 丸山葉, 大西俊輝, 大串旭, 呉健朗, 大澤正彦, 宮田章裕: 意見伝達を仲介するエージェントの基礎検討. HAI シンポジウム 2022 予稿集, G-17 (2022 年 3 月).

(14) 得田舜介, 呉健朗, 田中柊羽, 大西俊輝, 大串旭, 宮田章裕: 英語でボケる発言を行うエージェントの基礎検討, 情報処理学会インタラクション 2022 論文集, pp.291–294 (2022 年 3 月).

(15) 大串旭, 大西俊輝, 山内愛里沙, 石井亮, 杵渕哲也, 青野裕司, 宮田章裕: 言葉づかいに着目した褒め方の上手さの推定モデルの基礎検討, 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 '21), Vol.2021, pp.791–797 (2021 年 7 月).

(16) 田中柊羽, 呉健朗, 武藤祐太, 大西俊輝, 大串旭, 宮田章裕: たとえツッコミを行う対話型エージェントの基礎検討, 情報処理学会インタラクション 2021 論文集, pp.639–641 (2021 年 3 月).

(17) 大串旭, 大西俊輝, 山内愛里沙, 石井亮, 杵渕哲也, 青野裕司, 宮田章裕: 上手く褒めるために効果的な言葉づかいの調査, 情報処理学会インタラクション 2021 論文集, pp.714–718 (2021 年 3 月).

受賞

(1) 情報処理学会コラボレーションとネットワークサービスワークショップ 2023 ベストペーパー賞: 2 体の柔らかい物体の動きによる共感表現方法の検証, 受賞書: 新山はるな，呉健朗，大西俊輝，大串旭，宮田章裕 (2023 年 11 月).

(2) 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).

(3) 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 (DICOMO '22) 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, 大西俊輝, 大串旭, 石井亮, 深山篤, 中村高雄, 宮田章裕 (2022 年 7 月).

(4) 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).

(5) 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 (DICOMO '21) 優秀プレゼンテーション賞, 言葉づかいに着目した褒め方の上手さの推定モデルの基礎検討, 受賞者: 大串旭 (2021 年 7 月).