

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

大串 旭¹ 大西 俊輝² 田原 陽平¹ 石井 亮³ 深山 篤³ 中村 高雄³ 宮田 章裕^{1,a)}

概要: 日常生活において褒める行為は大切なコミュニケーションである。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。これより、我々は褒め方の上手さを評価するシステムの構築を目指す。この目標を達成するためには、褒め方の上手さを推定できる機械学習モデルの構築が必要である。そこで本稿では、話者（褒める人、褒められる人）の発話内容から、褒め方の上手さを推定できるか明らかにする取り組みを行う。はじめに、話者の発話内容としてBERTでベクトル変換したものを言語特徴量として抽出した。次に、抽出した言語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の構築を行った。最後に、話者の発話内容をどのように利用すれば上手い褒め方を推定できるのか分析を行った。その結果、分類モデルにおいては褒める人の発話のみを利用したモデルの精度が最もよく、回帰モデルにおいては褒める人の発話とその後の褒められる人の発話を利用したモデルの精度が最もよくなることが確認できた。

Empirical Study on Estimation of Praising Skills by Focusing on Linguistic Features

1. はじめに

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。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は困難である。そこで我々は、褒め方の上手さを自動的に評価するシステムの構築を目指す。この目標を達成するためには、褒め方の上手さを推定できる機械学習モデルを構築し、褒め方の上手さを自動的に評価する必要があると

考えられる。我々はこれまで、対話において話者が褒めているときの発話に着目し、言語・非言語行動の分析を行ってきた [1][2][3]。しかし、褒める行為には褒めているときの発話だけでなく、褒められる人の発話も重要である。関連研究より、褒める行為は対象の行動や性格に向けられたものであること [4]、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ことから [5]、褒めているときの発話は褒められる人の発話内容に基づいていることが考えられるため、褒められる人の発話にも着目すべきである。事前検討 [3][6] では、発話中の品詞の出現頻度に着目してきたが、本稿では推定精度向上を図りBERTによるベクトル変換をして推定を行う。本稿での貢献は下記の通りである。

- 褒め方の上手さを推定するためには、褒める人だけでなく褒められる人の発話も重要であるのか明らかにした。
- 言語情報に着目した褒め方の上手さを推定するために適切な特徴量の抽出方法を検証した。

¹ 日本大学文理学部

² 日本大学大学院総合基礎科学研究科

³ 日本電信電話株式会社 NTT 人間情報研究所

a) miyata.akihiro@acm.org

2. 関連研究

2.1 褒める行為に関する研究事例

褒める行為に関する研究は、心理学、社会学の分野で多く行われている。特に、教育心理学の観点で重要視されており、教師と生徒の関係性や生徒の能力を最大限発揮させるという観点から注目されている [7][8]。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・非言語行動と考えられている [4][9]。加えて、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 [5]。Mortonらは、褒める行為の一連の流れをモデル化し、褒める行為にはどういった要因や反応が行われているのか定義している [10]。褒める行為の対象として“外見”、“所有物”、“業績”、“褒められる人の努力”に分類できるとされており、褒める人は褒められる人の“業績”や“褒められる人の努力”を聞くことで褒め言葉を形成すると考えられている。褒める人の褒め言葉を受け取った後の褒められる人の発話には、褒められたことへの反応が含まれていると考えられている。

2.2 言語、非言語情報を用いた対話中のスキルや評価を評価する研究

人間の言語・非言語行動を利用して、コミュニケーションスキル、プレゼンテーションスキル、共感スキル、自己開示、話者の説得力などの行動・能力を分析する研究は数多く行われている。本研究は、人間の言語・非言語情報を利用して、特定のタスクやシーンにおける行動・能力を分析する研究と深く関連している。

コミュニケーションスキルを分析する研究事例として Okada ら [11]、Rasipuram ら [12] の研究が挙げられる。Okada らは、コミュニケーションスキルを人事管理経験者の評価をもとに言語・非言語情報を利用して、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 [11]。Rasipuram らは、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 [12]。

プレゼンテーションスキルを分析する研究事例として Ramanarayanan ら [13]、Yagi ら [14] の研究が挙げられる。Ramanarayanan らは、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 [13]。Yagi らは、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能力を推定するモデルの構築・評価を行っている [14]。

共感・自己開示・説得力などのスキルを分析する研究事

例として Park ら [15]、Ishii ら [16]、Tan ら [17]、Soleymani ら [18]、Naim ら [19] の研究が挙げられる。Park らは、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予測するために、複数のコミュニケーション・モダリティからの言語・非言語行動を利用したアプローチを提案し、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していると示している [15]。Ishii らは、Davis の対人反応性指標を用いて測定した共感スキルレベルに応じて、話者交替/継続における視線行動と対話行為について分析している [16]。Tan らは、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 [17]。Soleymani らは、対話中の言語・非言語行動を相関分析することで言語行動の言語内容が自己開示に関連していることを明らかにしている。加えて、自己開示のレベルを推定するために、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 [18]。Naim らは、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推定するための回帰モデルを構築し、重要な特徴量の相関を基に就職面接での重要な行動を分析している [19]。

これらの研究事例では、コーパスデータから分析の対象となるシーンの言語・非言語行動を抽出し、機械学習モデルの構築や分析を行っている。言語特徴量を利用している研究として、単語の品詞や使用回数、単語のベクトル化といった1単語ごとに特徴量を抽出する研究 [14]、[17]、[19] や文章のベクトル化 [18] といった1文章ごとに特徴量を抽出した研究がある。Soleymani らは、事前学習済みBERTを用いて発話を768次元のベクトルに変換し特徴量として利用したモデルが最も精度が良く自己開示レベルの推定には言語が最も適した特徴量であることを報告している [18]。我々は事前検討として、単語ごとに着目して特徴量の抽出を行ってきた [3] が、文章ごとの特徴量の抽出を行っていない。

3. 研究課題

これまでに我々は対話における言語・非言語情報に着目して上手く褒めるための行動の分析を行ってきた [1][2][3]。言語行動については、褒めているときの発話内容（以下、褒める発話とする）に着目して褒め方の上手さの推定と分析を行ってきた。しかし、褒める人の発話は褒められる人の発話と密接に関係していると考えられる。2.1節より、褒められる人の発話内容に起因して褒める発話が起り、その後褒められる人が褒める発話への反応となる発話を行っていると考えられる。このことから、褒める発話は褒められる人の発話と密接な関係にあることがわかる。褒める発話だけでなく褒められる人の発話を利用することで、褒め方の上手さの推定精度の向上が期待できる。これより、褒



図 1 2者対話の様子

める人、褒められる人の発話内容の両方に着目することで褒め方の上手さを推定できるのかを検証を行う。

加えて、事前検討 [3] における現在の推定精度 (F 値 = 0.523, $R^2 = 0.173$) では褒め方の上手さの評価システムを構築するためには十分ではないと考えられる。そこで 2.2 節を参考に特徴量の抽出方法を変更し、推定精度の向上を図る。具体的には、単語数や品詞といった単語単位から特徴量の抽出を行うのではなく、文章全体の意味を考慮できる BERT[20] を利用することで最適な特徴量の抽出方法の比較・検討を行う。BERT を利用することで文章全体の意味を考慮することが可能となり、褒める人、褒められる人の発話との関連を考慮できるといったことから褒め方の上手さの推定精度の向上が期待できる。上記をふまえ、本稿では褒める人、褒められる人の発話内容から、褒め方の上手さを推定することができるのか明らかにすることを研究課題とする。

4. 対話コーパス

4.1 2者対話の収録

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の作成を行った [1]。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 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 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。各組の参加者 (参加者 A, 参加者 B) は、撮影者の合図に従い、次の (1) ~ (3) を行った。

- (1) 自己紹介 (5 分間)
- (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)
- (3) 参加者 B が褒める人となり、参加者 A が褒められる

表 1 発話シーンのシーン数, 平均値, 最大値, 最小値

役割	シーン数	平均値	最大値	最小値
褒める人	2701 件	1.324 秒	23.117 秒	0.062 秒
褒められる人	3413 件	2.040 秒	26.234 秒	0.018 秒

人となる対話 (5 分間)

(1) ~ (3) の対話を 17 組分, 計 255 分間収録した。なお, (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。(2) と (3) の対話において, 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するエピソードを話すように指示した。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 アノテーション

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[21] を利用して人手で発話シーンのタグ付けを行った (表 1)。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。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, アノテータ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 [22] 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。

4.3 褒め方の上手さの評価

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, 褒め方の上手さの評価を行った。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

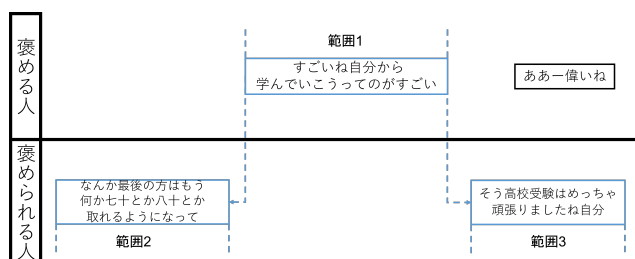


図 2 発話の抽出範囲

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 (228 件, 平均値 2.018 秒, 最大値 9.127 秒, 最小値 0.368 秒) 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) [23] 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があり、信頼性のあるデータであることが示唆された。

5. モデル構築

5.1 特徴量抽出

言語特徴量を抽出するにあたり、次に示す 3 つの抽出範囲を利用した (図 2)。

範囲 1 Praise シーンの褒める人 (Praiser) の発話

範囲 2 Praise シーン開始前の褒められる人 (Receiver) の発話

範囲 3 Praise シーン終了後の褒められる人の発話

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。発話内容を単語ごとに分析する方法として、範囲 1 の発話内容から MeCab[24] を利用して 13 種類の品詞の出現頻度と単語数 (表 2) を算出する方法を用いた [3]。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 モデルを利用して範囲 1~3 の発話内容をそれぞれ 768 次元のベクトルに変換する方法を用いた。

5.2 モデル構築

褒め方の上手さを推定するために回帰モデルと分類モデルの構築を行った。分類モデルを構築するにあたり、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。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

表 2 抽出した品詞

大分類	小分類
感動詞	感動詞
形容詞	形容詞
助詞	助詞 (格助詞), 助詞 (副助詞), 助詞 (連体化)
助動詞	助動詞
動詞	動詞 (自立), 動詞 (非自立)
名詞	名詞 (サ変接続), 名詞 (数), 名詞 (接尾), 名詞 (非自立),
連体詞	連体詞

ア低~高群を定義した*1。

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)

本稿ではモデルを構築するアルゴリズムとして Random Forest[25] を利用し、分類モデルと回帰モデルを構築した。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と、目的変数を Praise スコアとし、説明変数を 5.1 節で抽出した特徴量とする回帰モデルを構築した。決定木の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[26] を用いて最適化した。hold-out 法を利用して、トレーニングデータとテストデータを 9:1 に分割した上で機械学習モデルを構築するタスクを行った。なお、回帰モデルと分類モデルのそれぞれで上記のタスクを 100 回行った。

5.3 褒め方の上手さの推定結果

5.3.1 分類モデル

分類モデルの推定結果を図 3, 表 3 に示す。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。構築した全てのモデルに対して、Bonferroni 検定を行い、精度が向上しているのか確認した。ベースラインと Model A ($t(99) = 18.480, p < .01$), Model B ($t(99) = 18.212, p < .01$), Model C ($t(99) = 19.584, p < .01$), Model D ($t(99) = 15.847, p < .01$), Model E ($t(99) = 20.290, p < .01$) の間に 1% 水準で有意差が認められることが確認できた。これより、本稿で構築したすべてのモデルがベースラインよりも推定精度を向上させることができたと言える。Model A と Model B ($t(99) = 8.962, p < .01$), Model C

*1 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。

表 3 各分類モデルで利用した特徴量と推定精度の各指標の平均値 (N=100)

	単語数, 品詞の出現頻度 (範囲 1)	BERT (範囲 1)	BERT (範囲 2)	BERT (範囲 3)	適合率	再現率	F 値
Baseline					0.358	0.340	0.357
Model A	✓				0.602	0.519	0.523*
Model B		✓			0.659	0.554	0.620* †
Model C		✓	✓		0.649	0.631	0.609*†
Model D		✓		✓	0.632	0.624	0.590*†
Model E		✓	✓	✓	0.661	0.695	0.612*†

表 4 各回帰モデルで利用した特徴量と推定精度の各指標の平均値 (N=100)

	単語数, 品詞の出現頻度 (範囲 1)	BERT (範囲 1)	BERT (範囲 2)	BERT (範囲 3)	MSE	R ²
Baseline					0.669	-0.044
Model F	✓				0.507	0.173*
Model G		✓			0.483	0.237*‡
Model H		✓	✓		0.437	0.295*‡
Model I		✓		✓	0.390	0.336* ‡§
Model J		✓	✓	✓	0.439	0.319*‡§

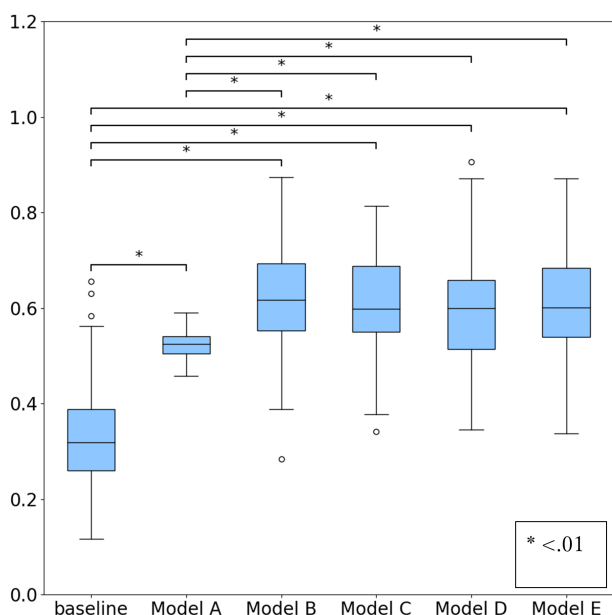
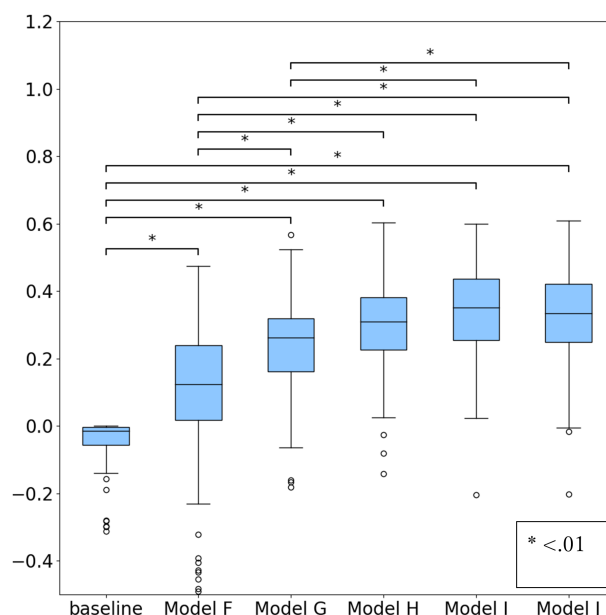


図 3 各分類モデルの推定精度 (F 値)

図 4 各回帰モデルの推定精度 (R²)

($t(99) = 8.714, p < .01$), Model D ($t(99) = 6.017, p < .01$), Model E ($t(99) = 8.300, p < .01$) の間に 1%水準で有意差が認められることが確認できた。このことから、分類モデルにおいて単語数や品詞の出現頻度を特徴量としたモデルよりも BERT を用いて抽出した文章ベクトルを特徴量としたモデルの方の精度が向上することが明らかになった。

5.3.2 回帰モデル

回帰モデルの推定結果を図 4, 表 4 に示す。ベースラインは、Praise スコアの平均値をテストデータとして推定するモデルを用いた。構築した全てのモデルに対して、Bonferroni

検定を行い、精度が向上しているのか確認した。ベースラインと Model F ($t(99) = 6.170, p < .01$), Model G ($t(99) = 16.012, p < .01$), Model H ($t(99) = 23.182, p < .01$), Model I ($t(99) = 24.655, p < .01$), Model J ($t(99) = 25.635, p < .01$) の間に 1%水準で有意差が認められることが確認できた。これより、本稿で構築したすべてのモデルがベースラインよりも推定精度を向上させることができたと言える。Model F と Model G ($t(99) = 4.976, p < .01$), Model H ($t(99) = 7.677, p < .01$), Model I ($t(99) = 10.468, p < .01$), Model J ($t(99) = 8.796, p < .01$) の間に 1%水準で有意差が認められることが確認できた。このことから、回帰モデルにおいて単語ごとに着目するよりも BERT を用いて特徴量を抽出して構築したモデルの方の精度が向上することが明らかになった。Model G と Model I ($t(99) = 4.680, p < .01$),

* Baseline との間で $p < .01$ † Model A との間で $p < .01$ ‡ Model F との間で $p < .01$ § Model G との間で $p < .01$

表 5 各分類モデルが正しく分類できた個数

Model	高群	中群	低群	合計
Model B	58	20	61	139
Model C	62	12	62	136
Model D	59	16	60	135
Model E	62	9	62	133

Model J ($t(99) = 3.730, p < .01$) の間に 1%水準で有意差が認められることが確認できた。このことから、回帰モデルにおいては、褒める人の発話だけでなく褒められる人の発話を利用することで精度が向上することが明らかになった。

5.4 考察

5.4.1 分類モデル

事前検討 [3] と比較して本稿で提案した分類モデルは推定精度を向上させることができたのか分析を行う。事前検討で提案した特徴量の抽出方法を用いた分類モデル (Model A) の推定精度は F 値 = 0.523 であった。本稿で提案した特徴量の抽出方法を用いた分類モデル (Model B~C) で最も推定精度が良かったのは Model B であり F 値 = 0.620 であった。構築した全ての分類モデルに対して Bonferroni 検定を行ったところ、Model A と Model B の間に 1%水準で有意差が認められた。事前検討と比較して本稿で提案した分類モデルが推定精度を向上させたことがわかった。加えて、Model B はある程度妥当に褒め方の上手さを推定できたと考えられる。そのため、本稿で構築した分類モデルを利用して褒め方の上手さの評価システムを構築することは有用であると考えられる。

本稿で構築した分類モデルでは Praise スコア高群をどのように分類できているのか分析を行う。分析を行うために、各分類モデルの文章ごとの分類結果を集計した。まず、100 回推定した際の各文章を予測したクラスの結果を集計した。次に、文章ごとの予測結果の中で最も多く予測できたクラスを求め、その文章の予測クラスとした。各分類モデルが正しく分類できた個数を表 5 に示す。高群を最もよく分類できていたのは Model C (範囲 1 + 範囲 2)、Model E (範囲 1 + 範囲 2 + 範囲 3) であった。この結果から上手い褒め方を判断するためには褒められる人の直前の発話が有用であると考えられる。Model B (範囲 1) では不正解であったが Model C で正解となった実際の対話をいくつか示す。褒められる人の“来年は大学院筑波のほうに”という発言に対して、褒める人が“えーっめっちゃ賢いじゃないっすか”と発言する。褒められる人の“二人の協力もあったんですか”という発言に対して、褒める人が“なるほどえ力作じゃないですか”と発言する。このように褒められる人の発話内容について言及しながら褒めていることがわかる。Model C と Model E がより高群を分類できたのは、褒められる人の発話内容と褒める発話との

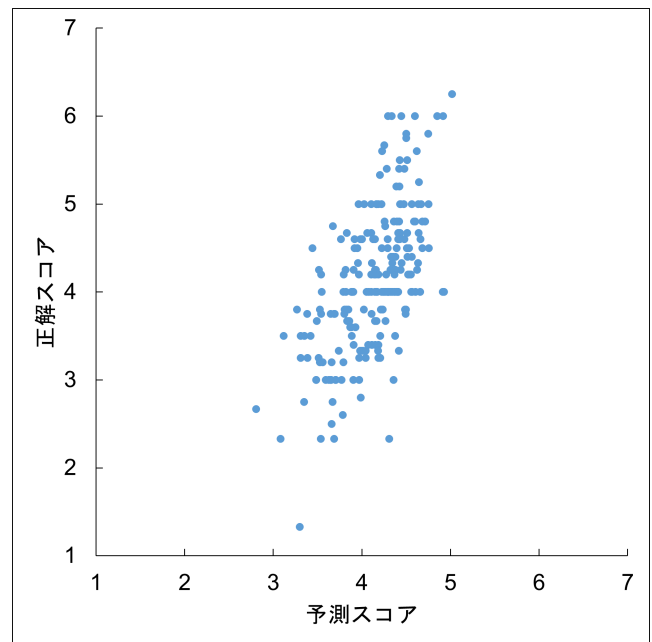


図 5 Praise スコアの正解値・予想値

表 6 各回帰モデルの予測値と Praise スコアの差が他のモデルと比べて最小となった個数

Model	高群	中群	低群	合計
Model G	17	15	14	46
Model H	14	13	23	50
Model I	35	12	22	69
Model J	15	25	23	63

関係性を抽出できたためだと考えられる。これより上手い褒め方するためには褒められる人の発話内容と関係した褒める発話をするのが有用であると考えられる。

5.4.2 回帰モデル

事前検討と比較して本稿で提案した回帰モデルは推定精度を向上させることができたのか分析を行う。事前検討で提案した特徴量の抽出方法を用いた回帰モデル (Model G) の推定精度は F 値 = 0.173 であった。本稿で提案した特徴量の抽出方法を用いた回帰モデル (Model H~J) で最も推定精度が良かったのは Model I であり F 値 = 0.336 であった。構築した全ての回帰モデルに対して Bonferroni 検定を行ったところ、Model G と Model I の間に 1%水準で有意差が認められた。事前検討と比較して本稿で提案した回帰モデルが推定精度を向上させたことがわかった。また、Model I はある程度褒め方の上手さを推定できたと考えられる。しかし、褒め方の上手さの評価システムを構築するために本稿で構築した回帰モデルを利用するためにはさらなる推定精度向上が必要であると考えられる。

本稿で構築した回帰モデルが Praise スコア高群をどの程度推定できていたのか分析を行う。分析を行うために、各モデルの文章ごとの予測値を算出した。まず、100 回推定した際の各文章の予測値を集計した。次に、文章ごとの

予測値の平均値を求め、その文書の予測値とした。正解スコア (Praise スコア) と Model I の予測値を図5に示す。図5より、回帰モデルの推定結果は3~5の間に収まっていることがわかる。Praise スコアが高い部分や低い部分への当てはまりはいま一步であった。各回帰モデルの予測値と Praise スコアの差が他のモデルと比べて最小となった個数を表6に示す。表6より、最も高群に近似できていたのは Model I であった。Model I の推定結果と Praise スコアの差が他のモデルと比べて最小となった高群の実際の発話をいくつか示す。褒める人が“あっもう済んでるんですね ああ凄い”と発言するのに対して、褒められる人が“いやもうほんとにびっくりしましたねあれは”と発言する。褒める人が“それで入れでも全然筑波に入れるってだけでも相当すごい気がするんですけどね”と発言するのに対して、褒められる人が“そうじゃないんですよ実は”と発言する。このように褒められる人の発話では褒める発話の内容への返答が行われていることがわかる。Model I がより上手い褒め方を推定できていたのは、褒める発話の内容と褒められる人の発話内容との関係性を抽出できたためだと考えられる。これより上手い褒め方をした際には、褒められる人が褒める発話についての反応をするのではないかと考えられる。

6. おわりに

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。そこで我々は、褒め方の上手さを評価するシステムの構築を目指す。この目標を達成するためには、褒め方の上手さを推定できる機械学習モデルの構築が必要である。本稿では、話者 (褒める人、褒められる人) の発話内容から、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、話者 (褒める人、褒められる人) の発話内容から BERT を用いて抽出した文章ベクトルを特徴量として機械学習モデルを構築した。分類モデルにおいては、褒める発話のみを利用したモデルが最も推定精度が良く F 値 = 0.620 となった。回帰モデルにおいては、褒める発話とその直後の褒められる人の発話を利用したモデルが最も推定精度が良く R^2 = 0.336 となった。各モデルの推定結果より、上手い褒め方を推定するためには褒められる人の発話が有用であることを明らかにした。さらに、言語特徴に着目した褒め方の上手さの推定を行う際に適切な特徴量の種類を検証した。特徴量として、単語ごとに着目したモデルとベクトルに変換したモデルを構築し褒め方の上手さの推定精度の比較を行った。その結果、単語ごとに着目して特徴量を抽出するのではなく、BERT を利用して文章を 768 次元に変換し特徴量として抽出する方が適切であった。

本稿では褒める人と褒められる人の両者の発話に着目して褒め方の上手さを推定する取り組みを行った。だが、褒められる人の発話の抽出範囲が適切であったのか疑問が残る。具体的には、褒められる人の発話中に褒める人が発話をするシーンが確認できた。褒められる人の発話中の褒める人の発話はその発話中の内容に向けたものであると考えることができる。今後、褒められる人の発話を抽出する適切な範囲を検討したい。加えて、非言語行動である表情や音声と合わせてマルチモーダルな分析を行う予定である。さらに、Random Forest 以外のアルゴリズムを利用することで褒め方の上手さを推定するために最適なアルゴリズムの調査を行いたい。

参考文献

- [1] 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).
- [2] 山内愛里沙, 大西俊輝, 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), pp. 98-106 (2020).
- [3] 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), pp. 791-797 (2021).
- [4] Henderlong, J. and Lepper, M. R.: The effects of praise on children's intrinsic motivation: A review and synthesis, *Psychological Bulletin*, Vol. 128, No. 5, pp. 774-795 (2002).
- [5] Jenkins, L., Floress, M. 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).
- [6] 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ティブ 2021 論文集, pp. 714-718 (2021).
- [7] Catano, V. M.: Relation of improved performance through verbal praise to source of praise, *Perceptual and Motor Skills*, Vol. 41, No. 1, pp. 71-74 (1975).
- [8] Brophy, J.: Teacher praise: A functional analysis, *Review of Educational Research*, Vol. 51, No. 1, pp. 5-32 (1981).
- [9] Kalis, T., Vannest, K. and Parker, R.: Praise Counts: Using Self-Monitoring to Increase Effective Teaching Practices, *Preventing School Failure*, Vol. 51, No. 3, pp. 20-27 (2007).
- [10] Morton, J., Mikolajczak, M. and Luminet, O.: New perspectives on the praise literature: towards a conceptual model of compliment, *Current Psychology* (2020).
- [11] Okada, S., Ohtake, Y., Nakano, Y., Hayashi, Y., Huang, H., Takase, Y. and Nitta, K.: Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets, *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16)*, pp. 169-176 (2016).
- [12] Rasipuram, S., Rao, P. and Jayagopi, D.: Asynchronous

- video interviews vs. Face-To-face interviews for communication skill measurement: A systematic study, *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16)*, pp. 370–377 (2016).
- [13] Ramanarayanan, V., Leong, C. W., Chen, L., Feng, G. and Suendermann-Oeft, D.: Evaluating Speech, Face, Emotion and Body Movement Time-series Features for Automated Multimodal Presentation Scoring, *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*, pp. 23–30 (2015).
- [14] Yagi, Y., Okada, S., Shiobara, S. and Sugimura, S.: Predicting multimodal presentation skills based on instance weighting domain adaptation, *Journal on Multimodal User Interfaces*, pp. 1–16 (2021).
- [15] Park, S., Shim, H., Chatterjee, M., Sagae, K. and Morency, L.: Computational Analysis of Persuasiveness in Social Multimedia: A Novel Dataset and Multimodal Prediction Approach, *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*, pp. 50–57 (2014).
- [16] 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, *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*, pp. 31–39 (2018).
- [17] Tan, Z., Goel, A., Nguyen, T. and Ong, D.: A Multimodal LSTM for Predicting Listener Empathic Responses Over Time, *14th IEEE International Conference on Automatic Face and Gesture Recognition (FG '19)*, pp. 1–4 (2019).
- [18] Soleymani, M., Stefanov, K., Kang, H., Ondras, J. and Gratch, J.: Multimodal Analysis and Estimation of Intimate Self-Disclosure, *Proceedings of the 21st ACM International Conference on Multimodal Interaction (ICMI '19)*, pp. 59–68 (2019).
- [19] Naim, I., Tanveer, M., Gildea, D. and Hoque, M. E.: Automated prediction and analysis of job interview performance: The role of what you say and how you say it, *11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG '15)*, pp. 1–6 (2015).
- [20] 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).
- [21] Brugman, H. and Russel, A.: Annotating Multimedia / Multi-modal resources with ELAN, *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC '04)*, pp. 2065–2068 (2004).
- [22] Consortium, T. U.: unidic-csj-3.0.1.1.
- [23] Shrout, P. E. and Fleiss, J. L.: Intraclass correlations: uses in assessing rater reliability, *Psychological bulletin*, Vol. 86, No. 2, pp. 420–428 (1979).
- [24] Kudo, T.: MeCab: Yet Another Part-of-Speech and Morphological Analyzer.
- [25] Breiman, L.: Random Forests, *Machine Learning*, Vol. 45, No. 1, pp. 5–32 (2001).
- [26] Bergstra, J. and Yamins, D. and Cox, D.: Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms, *Proceedings of the 12th Python in Science Conferences (SciPy '13)*, pp. 13–20 (2013).