

# マルチモーダル情報に基づくグループ会話におけるコミュニケーション能力の推定

## Estimating Communication Skills based on Multimodal Information in Group Discussions

岡田 将吾  
Shogo Okada

東京工業大学 情報理工学院 情報工学系  
School of Computing, Tokyo Institute of Technology  
okada@c.titech.ac.jp

松儀 良広  
Yoshihiro Matsugi

東京工業大学大学院総合理工学研究科知能システム科学専攻  
Department of Computational Intelligence and Systems Interdisciplinary, Graduate School of Science and Engineering, Tokyo Institute of Technology  
matsugi\_y@ntt.dis.titech.ac.jp

中野 有紀子  
Yukiko Nakano

成蹊大学理工学部情報科学科  
Department of Computer and Information Science, Seikei University  
y.nakano@st.seikei.ac.jp

林 佑樹  
Yuki Hayashi

大阪府立大学現代システム科学域知識情報システム学類  
College of Sustainable System Sciences, Osaka Prefecture University  
hayashi@kis.osakafu-u.ac.jp

黄 宏軒  
Hung-Hsuan Huang

立命館大学情報理工学部情報コミュニケーション学科  
Department of Information and Communication Science, College of Information Science and Engineering, Ritsumeikan University  
huang@fc.ritsumeit.ac.jp

高瀬 裕  
Yutaka Takase

成蹊大学理工学部情報科学科  
Department of Computer and Information Science, Seikei University  
yutaka-takase@st.seikei.ac.jp

新田 克己  
Katsumi Nitta

東京工業大学 情報理工学院 情報工学系  
School of Computing, Tokyo Institute of Technology  
nitta@dis.titech.ac.jp

**keywords:** multimodal interaction, social signal processing, communication skills, group conversation analysis

### Summary

This paper focuses on developing a model for estimating communication skills of each participant in a group from multimodal (verbal and nonverbal) features. For this purpose, we use a multimodal group meeting corpus including audio signal data and head motion sensor data of participants observed in 30 group meeting sessions. The corpus also includes the communication skills of each participant, which is assessed by 21 external observers with the experience of human resource management. We extracted various kinds of features such as spoken utterances, acoustic features, speaking turns and the amount of head motion to estimate the communication skills. First, we created a regression model to infer the level of communication skills from these features using support vector regression to evaluate the estimation accuracy of the communication skills. Second, we created a binary (high or low) classification model using support vector machine. Experiment results show that the multimodal model achieved 0.62 in  $R^2$  as the regression accuracy of overall skill, and also achieved 0.93 as the classification accuracy. This paper reports effective features in predicting the level of communication skill and shows that these features are also useful in characterizing the difference between the participants who have high level communication skills and those who do not.

### 1. はじめに

人工知能分野において「インタラクションを通じて行動主体が他者との社会的な関係を構築し、利用することで問題解決を行う能力」と定義された社会知 [西田 09] を

モデル化し、活用する研究が重要視されている [Nishida 08]。社会知を構築するための基盤として、人間同士の対面コミュニケーションをモデル化することは重要な課題の一つとして位置づけられている。本研究では対面コミュ

コミュニケーションのモデル化の 1 つの側面として、「コミュニケーション能力」に関する個人特性を計算モデルにより捉えることに焦点を当てる．人材育成の現場では，若年者のコミュニケーション能力向上のために，教育・訓練基盤の整備が必要であるとの認識が高まっており [JAVADA 05]，コミュニケーション能力の推定モデルの構築と評価はそれを支える技術となりうる．

このような背景において，我々は，アイデア創出や，意思決定等，多くの場面で行われるグループディスカッションにおける個人のコミュニケーション能力に着目し，コミュニケーション能力の改善・向上を支援するシステムに応用可能な技術を提案することを目指している．本研究で対象とするコミュニケーション能力を [JAVADA 05] に基づき定義する．[JAVADA 05] によると，若年者の習得すべきコミュニケーション能力は，双方向のコミュニケーションを円滑に行い，情報伝達を適切かつ的確に行う「意思疎通」，組織に適応する「協調性」，プレゼンテーション能力に関する「自己表現」の 3 つの項目から構成される．3 項目の中で，グループディスカッションでは，他者との円滑なコミュニケーション，他者への情報伝達が重要であるため，参加者の評価として「意思疎通」の項目を採用する．

意思疎通に関する能力の向上支援に向け，本研究は，ディスカッションを通じて観測できる参加者個人の発言の仕方／聞き方と発言内容から，意思疎通に関するコミュニケーション能力値を推定するモデルを機械学習により構築し，評価することを目的とする．推定モデルの構築には [林 15] で収集された，グループディスカッションコーパスを用いる．このコーパスは会話参加者の発話内容に関する言語情報，音声情報，頭部動作といった複数の非言語情報，会話参加者の「意思疎通」に関するコミュニケーション能力の評定値が含まれる．さらに，コミュニケーション能力の高い（もしくは低い）と評定された参加者の言語・非言語（マルチモーダル）特徴量を分析し，能力の高い／低い参加者に顕著な特徴を明らかにする．なお，本研究の対象は個人の意思疎通能力であるため，グループディスカッションの結果や，その良し悪しに関する評価は行わない．以上，まとめると，本研究の成果は以下の 2 点である．

#### (1) コミュニケーション能力の推定モデルの構築

40 名の参加者による計 9 時間を超える会話データから，人事採用経験者 21 名により評価した会話参加者のコミュニケーション能力を推定するモデルを学習する．評価実験の結果より，総合的なコミュニケーション能力に関して 9 割以上の精度で能力値の高・低を推定できることを示す．

#### (2) 能力値に寄与する特徴量の分析

各参加者の評定値の高・低を識別する上で有効な特徴量を分析し，人事採用経験者より高・低評価を受けた参加者の言語・非言語特徴を明らかにする．

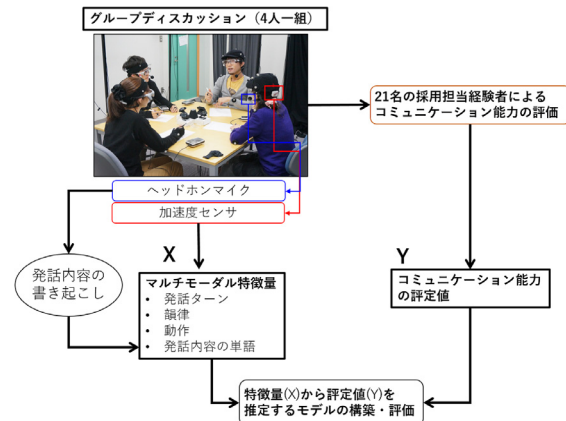


図1 コミュニケーション能力を推定する枠組み

本研究の概要を図1に示す．論文の構成は以下の通りである．2章で関連研究と本研究の位置づけに関して述べ，3章では本研究で用いるコーパスの概要に関して述べる．4章ではマルチモーダル特徴量の抽出方法に関して述べる．5章では評価実験の設定，実験目的を述べ，6章では評価実験の結果を述べる．7章でコミュニケーション能力値と各特徴量の関係を明らかにし，推定モデルの実用可能性に関して議論し，8章で本研究をまとめる．

## 2. 関連研究

社会学・コミュニケーション科学における会話分析の知見から，対面会話中に交わされる言語情報だけでなく，発話，韻律，視線，ジェスチャ，姿勢，表情といった非言語情報の役割が重要であることが明らかになった [Knapp 13]．上記の知見に基づき，カメラ，マイク，モーションセンサといったセンシングデバイスを利用して，会話中の対面インタラクション・参加者の状態を自動認識するための技術に注目が集まっている [Gatica-Perez 09]．

上記に関する研究を「会話中の非言語行動の認識・マイニング」(2.1節)と「会話参加者の個人特性の認識」(2.2節)の研究に分けて概観し，本研究の位置づけを2.3節で述べる．

### 2.1 会話中の非言語行動の認識・マイニング

グループ会話における注視状態，注視方向 (VFOA) を認識することで，受話者の自動推定が行われている．Jovanovic ら [Jovanovic 07] はグループ会話において，受話者が一人，複数人，全員の場合を認識する枠組みを提案した．Otsuka ら [Otsuka 07] は Dynamic Bayesian Network を用いて，グループ会話における全メンバーの頭部方向，頭部動作，発話状態から，視線方向，視線状態とインタラクションの構造を推定する枠組みを提案した．他にも話者交代のメカニズムをモデル化し，発話交代場面を認識したり [Chen 09]，次話者の発話タイミングを予測する枠組みが提案されている [Ishii 13]．McCowan

ら [McCowan 05] は音声・画像特徴量を利用して、ミーティングにおけるイベント（一人の発話、議論、プレゼン）を認識し、シーンを分割することを試みた。中田ら [中田 11] は、発話区間、あいづち、指さし、視線、頷きに関するアノテーションのラベル系列データから、頻出ラベル列を抽出し、分析した。上記の研究は、会話中の非言語パターンや、会話イベントを推定、もしくはマイニングするものである。これに対し、非言語パターンや会話イベントの特徴量を統合して、参加者の個人特性を認識する研究が行われており、本研究もその研究分野に属する。次節では、これについて述べる。

## 2.2 会話参加者の個人特性の認識

Nguyen ら [Nguyen 14] はアルバイト面接場面における面接官と応募者の会話を収録し、観測されるマルチモーダル情報から、応募者のコミュニケーション能力や、採用に値するか否かの評定値を推定するモデルを構築し、高評価を得る応募者の所作を分析した。Tanaka ら [Tanaka 15] はバーチャルエージェントとユーザの対話における、ユーザの発話中の韻律情報、発話単語数を手掛かりとして、エージェントがソーシャルスキルを向上させるためのアドバイスを行うシステムを開発した。前者の研究は、コミュニケーション能力に関する特徴量の分析を、後者の研究は、コミュニケーション能力の向上支援のためのシステムを構築している。[Nguyen 14, Tanaka 15] の目的は本研究の目的に類似する点はあるが、いずれも 2 者対話を対象としており、本研究のようにグループディスカッションを対象としていない。

Pianesi ら [Pianesi 08] は、心理学の分野で提案された Big Five と呼ばれる個人の性格特性を、グループ会話で観測される非言語情報から推定した。Aran ら [Aran 13] は外部観察者によって参加者の性格に対する印象をアノテーションし、その値を推定した。これらの研究は、性格特性を推定することが目的であり本研究のようにコミュニケーションスキルの推定を目的としていない。

Sanchez-Cortes らは ELEA コーパス [Sanchez-Cortes 13] と呼ばれるグループディスカッションコーパスを収集した。また同著者は ELEA コーパスを用いて、頭部方向、発話区間、韻律、上半身・頭部動作の非言語特徴量を組み合わせることで、問題解決を行う過程で、リーダーシップを発揮した参加者を推定した [Sanchez-Cortes 12]。この研究では、グループ内の会話参加者が相互評価を行い、各参加者のリーダーシップを発揮したレベルを決定する。[Sanchez-Cortes 12] では、リーダーシップの分析のために、個人の Dominance（優位性）、Competence（能力、力量）と呼ばれる項目の優劣も推定しており、それらの項目は、一部本研究の分析対象であるコミュニケーション能力と類似している。一方で、本研究はリーダーシップではなく、意思疎通に関するコミュニケーション能力の分析に焦点を当てている点が異なっている。

## 2.3 本研究の位置づけ

本研究が属する「会話参加者の個人特性の認識」（2.2 節）に関する研究と、本研究との相違点をまとめる。グループにおける個人のコミュニケーション能力の分析 [Aran 13, Pianesi 08] ではグループにおける個人の性格特性を、[Sanchez-Cortes 12] ではグループにおける個人のリーダーシップを分析している。[Nguyen 14, Tanaka 15] では対面 2 者対話における、コミュニケーション能力の分析・モデル化を試みている。これらの研究との差異として、本研究はグループにおける個人のコミュニケーション能力を分析する。ディスカッションコーパスの持つ多様性 [Aran 13, Sanchez-Cortes 12] で用いられた ELEA コーパス [Sanchez-Cortes 13] や、[Pianesi 08] で用いられたコーパスでは、参加者はインバケット型と呼ばれる問題解決タスク<sup>\*1</sup>を 1 種類行っているのに対し、本研究で用いるコーパス [林 15] では、参加者は 1 つの「インバケット型」課題、2 つの「ケーススタディ型」課題の計 3 種類の問題解決課題を行っているため、研究対象のデータは多様性を有しており、多様な課題における参加者のマルチモーダル情報と評定値データを含んでいる。客観的な評価データの取得 [Sanchez-Cortes 12] では会話に参加したメンバがお互いのリーダーシップ性を評定した。本研究は、参加メンバではなく、21 名の人事採用経験者による評定を基にコミュニケーション能力値を付与しており、より客観的かつ信頼性の高い評定値を得ている。言語・非言語特徴量の利用 グループ対話を対象とした [Aran 13, Pianesi 08, Sanchez-Cortes 12] では、主に会話者の非言語情報に基づき分析を行っている。一方で、グループディスカッションにおける各個人の発言内容は重要であることから、本研究では非言語特徴量だけでなく、新規に発言された名詞単語数などの、言語特徴量も用いて、能力値を推定するモデルを学習する。

## 3. グループディスカッションコーパス

本実験で使用するグループディスカッションコーパス [林 15] の概要を説明する。以降ではこのコーパスを GD（Group Discussion）コーパスと呼称する。GD コーパスには、質問紙調査により取得した性格特性（Big Five）、アイトラッカーのログデータ、加速度センサから得られた頭部加速度、Kinect から得られた深度情報、指向性ヘッドセットマイクから集音した音声情報、Web カメラから取得した顔映像、光学式モーションキャプチャから取得した頭部位置座標、ビデオカメラから取得した俯瞰映像が含まれる。このうち、本研究ではヘッドセットマイク

<sup>\*1</sup> [Pianesi 08, Sanchez-Cortes 13] で参加者は「Mission survival task」と呼ばれる課題に取り組んでおり、[林 15] のコーパスの課題内容とは異なる。



から観測した音声データ，加速度センサから観測した頭部加速度データより以降のマルチモーダル特徴抽出を行う．実験には 40 名の大学生が参加し，4 名の参加者を 1 つのグループ（図 1 の左図）として，合計 10 グループの会話データを収集した．各参加者は以下で述べる 3 つの課題に取り組み，その過程で得られるデータセットを収集した．結果として，合計 120 名分（4 名×10 グループ×3 課題）の参加者のデータセットが GD コーパスに含まれている．

### 3.1 ディスカッション課題概要

課題による得手・不得手が原因で各参加者のコミュニケーション能力が過大・過小評価されることを懸念し，各グループで 1 つの課題だけでなく，計 3 つの課題に取り組むよう設定した．グループ内の参加者同士は初対面であり，各グループは学生が身近に経験する機会があるテーマとして「学園祭に呼ぶべき有名人をランク付けする課題（15 分）」、「学園祭における出店計画を作る課題（20 分）」、「外国人の友人を日本に呼んでもてなす計画を作る課題（20 分）」の計 3 つの問題解決課題に取り組む．

各参加者のコミュニケーション能力の差異を，人事採用経験者に評価させるために，会話参加者には，設定した課題に真剣に取り組んでもらう必要がある．そこで，参加者に以下の点を指示した．

- ディスカッションを通じて問題解決を行い，グループでの結果を発表すること
- ディスカッションの様子は，後日人事採用経験者に閲覧され，ディスカッションの過程が評価されること
- 人事採用経験者の評価において発言の仕方／聞き方と発言の内容が重視されること

### 3.2 コミュニケーション能力項目の定義

[JAVADA 05] で定義される意思疎通能力は 5 つの要素項目で構成される．

傾聴する姿勢：

相手の言動をよく観察することができる．相手の主張を正確に聞き取ることができる．相手の立場に立って真意を聞き取ったうえで，的確な質問により更に話を聞き出すことができる．

双方向の円滑なコミュニケーション：

相手の意見を受け入れることができる．自分の価値観と異なる意見・考え方を否定しない．相手の考え方を総合的に理解したうえで，意見交換を円滑に行う．

意見集約力：

相手の意見や複数の異なる意見を分類・整理したうえで，要旨を整理し要約ができる．

情報伝達力：

タイミングを外すことなく，相手にとって必要な情報を正確に伝えることができる．

論理的で明瞭な主張：

表 1 各評定者間での評定値の一致度

能力項目	クローンバック $\alpha$
傾聴する姿勢	0.66
双方向の円滑な コミュニケーション	0.86
意見集約力	0.90
情報伝達力	0.90
論理的で明瞭な主張	0.87
コミュニケーション能力	0.91

発言の道筋が明確で論理的な主張ができる．場面に応じて，適切かつ明瞭な表現方法で主張ができる．

総合的なコミュニケーション能力：

本項目は，5 つの要素項目を考慮して評定される．

### 3.3 人事採用経験者による評定

人材派遣会社を通じて「人事採用経験者」の募集を行った．その結果，様々な業種（小売業，人材派遣，IT 関連等）と企業規模（中小企業から大企業まで）の採用面接担当を経験した計 21 名が集まった．3 つの課題ごとに参加者のコミュニケーション能力を，上記の 21 名の人事採用経験者が評定した．

人事採用経験者（評定者）が会話の俯瞰映像を閲覧し，3.2 節で定義した 5 つの要素項目と総合的なコミュニケーション能力の評定を行う．初めに各セッションのビデオデータを 3 分割（約 5～7 分）し，各 7 名が各分割区間のビデオデータを閲覧し，各参加者に点数を付ける．参加者の能力評価を行うために，ディスカッションの過程で行われる，個人のコミュニケーション行動に関して評価を行うよう指示した．各要素項目を最低 1～最大 5 の 5 段階で，総合的なコミュニケーション能力値を最低 1～最大 10 の 10 段階で評定した．

各評定者による評定値の一致度を計算し，各項目に関する信頼性を確認した．一致度の計算に関して，信頼性分析の項目を評定者として各項目ごとにクローンバック  $\alpha$  値を算出した．ある項目に関してクローンバック  $\alpha$  が 0.8 以上であれば各評定者による参加者への評定値は概ね一致しており，その項目の評定値は信頼性が高いと言える．一方でクローンバック  $\alpha$  が低い場合，ある参加者に対して各評定者が異なる評定値を付与したことを示し，その項目の評定値の信頼性は低い．

表 1 の 2 列目に項目ごとのクローンバック  $\alpha$  の値を示す．図中では小数点以下第三位を四捨五入した値を記載している．「傾聴する姿勢」( $\alpha = 0.66$ ) 以外のすべての項目で  $\alpha$  は 0.8 を上回っており，項目としての信頼性は高いと考える．また 21 名は異なる業種に属する採用経験者であるにもかかわらず，参加者への評定一致度が高いことが示された．以降では，21 名の評定者が回答した評定値の平均値を，各参加者のコミュニケーション能力値と定義する．また一致度が低い「傾聴する姿勢」の項目を本研究では能力項目から除外する．

表 2 コミュニケーション能力推定に用いた特徴量のまとめ

特徴量グループ	特徴量	説明
発話ターン <i>S</i>	合計発話長	1 セッション中の発話時間
	合計発話回数	1 セッション中の発話回数
	合計発話長 (1 秒以上)	1 秒以上の発話の合計時間
	合計発話回数 (1 秒以上)	1 秒以上の発話の合計回数
韻律 <i>A</i>	最大ピッチ	1 セッション中の各発話におけるピッチの最大の平均値
	最小ピッチ	1 セッション中の各発話におけるピッチの最小の平均値
	ピッチ平均	1 セッション中の各発話におけるピッチの平均の平均値
	最大インテンシティ	1 セッション中の各発話の最大音圧の平均値
	最小インテンシティ	1 セッション中の各発話の最小音圧の平均値
	音圧の幅	1 セッション中の発話の最大音圧と最小音圧の差
	抑揚	最大ピッチと最小ピッチの差
動作 <i>M</i>	話速	発話の音節数を発話長で割った値
	動作量の平均	セッション中の動作量の平均
	動作量の偏差	セッション中の動作量の標準偏差
	発話中の動作量の平均	発話中の動作量の平均
言語 <i>L</i>	発話中の動作量の偏差	発話中の動作量の標準偏差
	名詞数	1 セッション中の発話に含まれる名詞数
	動詞数	1 セッション中の発話に含まれる動詞数
	感動詞数	1 セッション中の発話に含まれる感動詞数
	フィラー数	1 セッション中の発話に含まれるフィラー数
	新規名詞数	1 セッション中の発話に含まれる新しく出現した名詞数
	既出名詞数	1 セッション中の発話に含まれる新規名詞以外の名詞数
	新規名詞数/発話回数	新規名詞数を発話回数で割った値

## 4. マルチモーダル特徴量の抽出

[Sanchez-Cortes 12] を参考に発話ターン特徴量、発話中の韻律特徴量、動作特徴量を抽出する．さらにコミュニケーション能力を構成する要素のうち「意見集約力」, 「論理的で明瞭な主張」といった項目は、主に発話内容に関連すると考え、発話内容の言語特徴量を抽出する．アノテーションツール ELAN[Brugman 04] を用いて、人手により発話内容を書き起こし、発話区間と発話内容を取得する．以下で述べる特徴量を表 2 にまとめる．

### 4.1 発話ターン特徴量

発話ターンの特徴量を以下にまとめる．  
合計発話長：

発話区間から算出した発話時間を発話長とし、1 セッション単位（各タスクの開始時間から終了時間まで）で発話長の合計を計算する．

合計発話回数：

1 つの発話区間を発話断片と定義する．発話断片の回数を発話回数とし、1 セッション単位で発話回数の合計を計算する．

合計発話長 (1 秒以上)：

1 秒以上の発話断片を抽出し、その長さの総和を計算する．[Jayagopi 10] ではあいづちなどの発話区間が短い、1 秒未満の発話断片を除き、それ以外の発話断片を用いた．本研究でも 1 秒未満の発話断片を除いた発話区間が長い発話断片のみを使用した発話長も計算する．

合計発話回数 (1 秒以上)：

1 秒以上の発話断片を抽出し、その発話回数を計算する．

### 4.2 韻律特徴量

各発話断片中の韻律情報の特徴量を抽出する．4.1 節で得られた発話区間情報を用いて、各発話断片の特徴量を抽出する．特徴抽出には音声分析ソフトウェア Praat [Boersma 13] を用いる．音声のピッチ特性と高い相関を有し、音声分析の分野で実質的に声の高さとして扱われる基本周波数を特徴量として計算する．厳密にピッチレベルと基本周波数は一致しないが、以降ではこの特徴量を便宜上ピッチ特徴量と呼称する．また声の大きさの特徴を捉えるため、インテンシティレベルに関する特徴量を抽出する．韻律特徴量を以下にまとめる．

最大ピッチ、最小ピッチ：

各発話断片の最大ピッチ、最小ピッチを計算し、1 セッション中の全発話断片に対して平均値を計算する．

ピッチ平均：

各発話断片に関して、発話区間でピッチの平均値を計算し、1 セッション中の全発話断片に対して平均値を計算する．

最大インテンシティ、最小インテンシティ：

各発話データの最大、最小音圧（インテンシティ）を計算し、1 セッション中の全発話断片に対して平均値を計算する．

音圧の幅：

最大、最小インテンシティの差を計算する．

抑揚：

最大ピッチと最小ピッチの差を計算する．

話速：

発話のシラブル数を各発話長で割った値を計算する．

### 4.3 動作特徴量

参加者の後頭部に加速度センサ (ATR-Promotions: WAA-010) を取り付け、 $x, y, z$  の 3 軸方向における加速度を 30fps で計測した後得られる時系列データから 4 つの頭部動作特徴量を抽出する。

動作量の平均, 標準偏差:

セッション中に参加者の頭部が動いた量を抽出するために, ある時刻  $t$  における加速度の 3 次元ベクトル:  $a_t = \{x_t, y_t, z_t\}$  のノルム  $|a_t|$  を計算し,  $|a_t|$  のセッション間における, 平均と標準偏差を計算する。

発話中の動作量の平均, 標準偏差:

4.1 節で得た発話区間情報を用いて, 発話区間中の動作量に関する特徴量を抽出する。ある参加者から  $M$  回の発話断片:  $UT = \{ut_1 \dots ut_m \dots ut_M\}$  が観測されたとする。 $ut_m$  の区間中の動作量の平均  $|\bar{a}|_m$ , 標準偏差  $std|a|_m$  を計算する。 $M$  個の発話断片に関して  $|\bar{a}|_m, std|a|_m$  の平均を計算し, これらの値を発話中の動作量の平均, 標準偏差と, それぞれ定義する。

### 4.4 発話内容の言語特徴量

書き起こしを行った発話内容のテキストデータを入力として, 形態素解析を行い, 発話内容を単語ベクトルに変換し, 単語の品詞情報を得る。形態素解析には MeCab<sup>\*2</sup> を用いた。

GD コーパスに含まれる 3 種類のディスカッションタスクは性質が異なるうえ, グループごとに自由な提案を行うことを許容しているため, 会話内容は多岐にわたる。この理由より, 語彙の共起特徴 (bag of words) や語彙順序 (n-gram) の頻度特徴量を抽出せず, 各単語の品詞情報に関する特徴量を抽出する。具体的には以下の品詞の種類ごとに, 単語をカウントして特徴量を抽出する。言語特徴量を以下にまとめる。

名詞数, 動詞数, 感動詞数, フィラー数:

1 セッション中の発話に含まれる各品詞の数を合計する。明瞭な意見を発話するためには名詞と動詞が伴った発話が必要になると考えるため, 名詞数と動詞数を特徴量として抽出する。感動詞は「ああ」「なるほど」「うん」などを含み, 相手の発言に対するあいづちや反応をするときによく観測される。フィラーは「えーっと」「なんか」などを含み, 発話の合間を埋めるときによく観測される。

新規名詞数, 既存名詞数:

新規名詞に関して, 4.1 節で得た発話区間情報を利用して, 各名詞が初めて発言された時間, 発言した参加者を特定し, 参加者ごとに数え上げた回数を新規名詞数とする。セッション中の新規名詞は他の参加者に対して新しい提案をした際に使用される傾向にある。また名詞数から新規名詞数を引くことで既存名詞数を計算する。

新規名詞数/発話回数:

新規名詞数を発話回数で割った値を計算する。

## 5. 評価実験

回帰予測タスクを行い, 各参加者の評定値 (連続値) の推定精度を評価する (6.1 節)。次に参加者の評定値を高群・低群に分類し, 2 クラス分類タスクを行う (6.2 節)。学習された分類モデルのパラメータを分析することで, コミュニケーション能力の高い (または低い) と評価された会話参加者の特徴を明らかにする (7.1 節)。

### 5.1 実験手順

参加者の特徴量セットを入力  $X$ , その会話を閲覧した評定者が採点した, その参加者の評定値を出力  $Y$  として, この入出力関係を学習する。合計で 120 個のデータを利用できる予定であったが, 13 データの特徴量がセンサの不具合などで一部欠損していたため, これらを除外し合計 107 個のデータを用いた。評価には交差検定法を利用した。あるグループ  $i$  に属する参加者から得られたデータをテスト, それ以外の 9 グループに属する参加者から得られたデータを訓練に用いて回帰, 分類モデルの評価実験を行った。

#### §1 回帰学習の概要

回帰モデルの推定性能の評価指標にはテストデータに対する決定係数  $R^2$  を用いた。テストデータに対応する評定値 (正解) と推定値の二乗誤差を, テストデータに対応する評定値の分散で割った値を  $E$  とすると, 決定係数  $R^2$  は  $1 - E$  で計算される。したがって, 誤差が小さいほど  $R^2$  は大きくなり, 良い結果を示す。

回帰学習には線形のサポートベクトル回帰モデル (Support Vector Regression: SVR) を用いた。SVR における誤差関数には  $\epsilon$  許容誤差を用いた。誤差の許容度を調整する  $\epsilon$ , 損失関数とマージンの大きさの間のトレードオフを調整するパラメータである  $C$  に関して, それぞれ  $[0, 0.01, 0.1, 1]$ ,  $[0.01, 0.1, 1, 5, 10]$  の探索域を設け, 訓練データを用いた交差検定を通じて, グリッドサーチにより最適パラメータを決定した。

#### §2 分類学習の概要

評価値の高群と低群を分類するタスクを行うために, 平均値付近に対応するデータを除外し, 予め連続値である評定値を高・低の二値ラベルに変換した。具体的には, 全データの評定値の平均値  $m$  と標準偏差  $\sigma$  を算出し,  $m + \beta\sigma$  以上を高群,  $m - \beta\sigma$  以下を低群, に分類し, 上記の条件を満たさない平均値に近い値をもつデータを除外した。データ数のバランスを考慮し,  $\beta = 0.1$  と決定した。分類モデルの推定性能には全テストデータの正答率を用いた。分類学習には, 線形のサポートベクトルマシン (Support Vector Machine: SVM) を用いた。SVM における損失と, マージンの大きさの間のトレードオフ

\*2 MeCab: <http://taku910.github.io/mecab/>



を調整するパラメータである  $C$  を  $[0.01, 0.1, 1, 5, 10]$  の範囲で探索しテストに用いた。

## 5.2 比較対象の特徴量セット

各モダリティの推定性能への寄与を検証するため、以下の15種類の特徴量セットを準備し、推定精度の比較を行う。韻律、発話ターン、言語、動作の特徴量セットを  $A, S, L, M$  とそれぞれ略記する。

- (1)  $A$ : 韻律特徴量
- (2)  $S$ : 発話ターン特徴量
- (3)  $L$ : 言語特徴量
- (4)  $M$ : 動作特徴量
- (5)  $A+S$ : 韻律と発話ターン
- (6)  $A+L$ : 韻律と言語
- (7)  $A+M$ : 韻律と動作
- (8)  $S+L$ : 発話ターンと言語
- (9)  $S+M$ : 発話ターンと動作
- (10)  $L+M$ : 言語と動作
- (11)  $A+S+L$ : 韻律と発話ターンと言語
- (12)  $A+S+M$ : 韻律と発話ターンと動作
- (13)  $A+L+M$ : 韻律と言語と動作
- (14)  $S+L+M$ : 発話ターンと言語と動作
- (15)  $All$ : 全ての特徴量 ( $A+S+L+M$ )

## 6. 実験結果

コミュニケーション能力を構成する4つの要素項目と、総合的なコミュニケーション能力の項目の計5つの評定値を推定する回帰タスク、分類タスクの結果を図2、図3に示す。図2、図3は各項目に対応する5つのグラフで構成されており、1つのグラフは15本の棒グラフの束で構成されている。この15本は5.2節で述べた15種類の特徴量セットに対応している。各棒グラフの色と、特徴量の対応を図2の上部に記載している。左から4本の棒グラフ（濃い青色）は単一モダリティの特徴量セットを用いた結果を示しており、 $A, S, L, M$  ((1)~(4))を用いたモデルの性能をそれぞれ示している。5本目から15本目までは、マルチモーダル特徴量セット((5)~(15))の結果を示している。

### 6.1 回帰モデルの評価実験結果

図2において、縦軸はテストデータへの決定係数  $R^2$  を示している。 $A, S, L, M$  の各特徴量セットを用いた場合の推定精度と  $A, S, L, M$  のうち、2つ以上の特徴量セットを組み合わせたマルチモーダル特徴量セットを用いた場合の推定精度の間で差が存在するかどうかを  $t$  検定により検証した。各テストデータの推定値と対応する正解の評定値との間の二乗誤差を検定に用いた。 $\square$  は  $p < 0.05$  で有意な差が認められたペアを示す。

訓練データに回帰モデルが過剰適合した場合や、特徴量が推定に有効でない場合、テストデータへの  $R^2$  は負の値をとる。つまり  $R^2 < 0$  の場合、評定値を推定できな

かったことを意味している。縦軸の最小値を  $-0.1$  とし、 $R^2 < -0.1$  の場合は  $-0.1$  として表示した。

#### §1 単一モダリティ特徴量セットのモデルの精度

単一モダリティの特徴量セットを用いたモデルに関して全要素項目で「発話ターン ( $S$ )」のモデルの精度が最大であった。「双方向の円滑なコミュニケーション」、「意見集約力」、「情報伝達力」、「論理的で明瞭な主張」の最大推定精度はそれぞれ 0.43, 0.50, 0.46, 0.49 であった。「総合的なコミュニケーション能力値」に関しては「言語 ( $L$ )」のモデルの精度が最大であり、0.52 であった。

[Sanchez-Cortes 12] では「発話ターン ( $S$ )」に含まれる発話長や発話回数などの特徴量はリーダーシップの印象推定にも有効であることが報告されており、本研究における個人のコミュニケーション能力の印象においても同様の結果が得られた。一方で、[Sanchez-Cortes 12] では言語特徴量を分析対象としていなかったが、本研究では発話中に含まれる各品詞単語の頻度特徴量も、非言語特徴と同様にコミュニケーション能力の推定に有効であることを示した。

#### §2 マルチモーダル特徴量セットのモデルの精度

全項目においてマルチモーダル特徴量セットを用いた場合に、最も良好な推定精度が得られた。全要素項目で「発話ターン+動作 ( $S+M$ )」のモデルの精度が最大であった。「双方向の円滑なコミュニケーション」、「意見集約力」、「情報伝達力」、「論理的で明瞭な主張」の最大推定精度は、それぞれ 0.47, 0.54, 0.55, 0.52 であった。

最後に「総合的なコミュニケーション能力値」の推定精度について、最大値は「発話ターン+言語+動作 ( $S+L+M$ )」で学習したモデルの推定精度であり 0.62 であった。この推定精度は  $A, S, L, M$  の各特徴量を用いた場合の精度を有意に上回っている。

要素項目に関して、マルチモーダル情報の統合は、一部の単一モダリティ特徴量セットのモデルの精度を向上させるために役立つ。また「総合的なコミュニケーション能力」を推定する場合に、マルチモーダル情報の利用が最も有効であることが示された。

推定精度が最大であった特徴量セットを以下にまとめる。「総合的なコミュニケーション能力」の推定には  $S+L+M$ 、「双方向の円滑なコミュニケーション」、「意見集約力」、「情報伝達力」、「論理的で明瞭な主張」の推定には  $S+M$  が有効であった。発話ターン特徴量 ( $S$ )、動作特徴量 ( $M$ ) は全ての項目を推定するために有効であった。

### 6.2 分類モデルの評価実験結果

図3の縦軸は、テストデータの正答率 (Accuracy) を示しており、各棒グラフは、各モダリティの特徴量セットの組み合わせごと (15種類) の精度を比較している。6.1節と同様に分類精度を報告する。いずれの特徴量セットを用いた場合も 0.5 を上回っていたため、縦軸の最小値を 0.5 とした。図3において正答率の小数点第3位を四

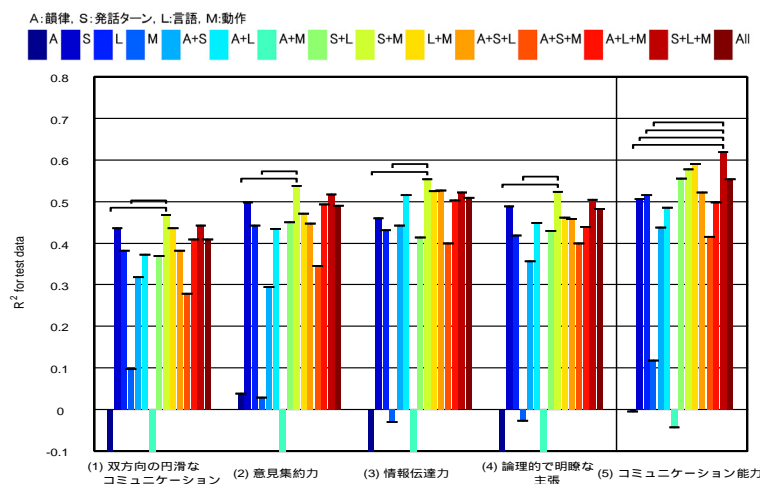


図2 SVRを用いた回帰タスクの結果(1)~(5)の棒グラフの束は各コミュニケーション能力項目の推定精度を示しており、各棒グラフは、各特微量セットを用いた場合の精度を示している。各棒グラフと特微量セットの対応関係は図の上に記載している。縦軸の精度はテストデータに対する決定係数  $R^2$  を示す。

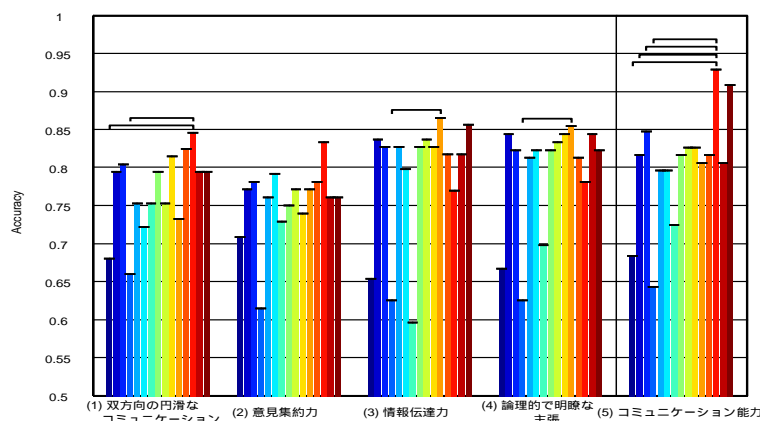


図3 SVMを用いた分類タスクの結果(1)~(5)の棒グラフの束は各コミュニケーション能力項目の推定精度を示しており、各棒グラフは、各特微量セットを用いた場合の精度を示している。各棒グラフと特微量セットの対応関係は図2の上に記載している。縦軸の精度はテストデータの正答率を示す。

捨五入した値を報告する。

### §1 単一モダリティの特微量セットのモデルの精度

「情報伝達力」、「論理的で明瞭な主張」に関して、「発話ターン(S)」のモデルの精度が最大であり、両方0.84であった。また「双方向の円滑なコミュニケーション」、「意見集約力」、「総合的なコミュニケーション能力」に関して、「言語(L)」のモデルの精度が最大であり0.80, 0.78, 0.85であった。

### §2 マルチモダリティ特微量セットを用いたモデルの精度

「双方向の円滑なコミュニケーション」の評定値の推定精度について、最大値は  $A+L+M$  の0.85であり、韻律、動作特微量を単体で用いるよりも、精度が有意に高かった。「意見集約力」の評定値の推定精度についても、同様に最大値は  $A+L+M$  のモデルで得られ、0.83であった。「情報伝達力」の評定値の推定精度に関して、最大値は  $A+S+L$  の0.87であった。「論理的で明瞭な主張」の評定値の推定精度に関しても、最大値は  $A+S+L$  のモデルで得られ、0.85であり、動作特微量を単体で用いるよりも、精度が有意に高かった。

上記の要素項目の分類精度に関して、単一モダリティ特微量を用いた場合の精度と比較して、0.01~0.05(最大5%)改善した。また「双方向の円滑なコミュニケーション」、「情報伝達力」、「論理的で明瞭な主張」において有意差が認められた。「総合的なコミュニケーション能力」の推定精度について、最大値は「韻律+言語+動作( $A+L+M$ )」で学習したモデルの0.93であった。この推定精度は全ての単一モダリティ特微量を用いた場合の精度を有意に上回っている。回帰タスクと同様に、「総合的なコミュニケーション能力」に関して、マルチモダリティ特徴の統合が特に有効であることが示された。

### §3 関連研究との分類精度の比較

先行研究である[Okada 15]ではリーダーシップ(Leadership)のレベルの高低を最大0.75の精度で、コミュニケーション能力(Competence)を最大0.66で分類可能であることが示された。[Okada 15]でも使用されたELEAコーパス[Sanchez-Cortes 13]における参加者の多くは、日本以外の国籍を持ち、英語やフランス語で会話されているのに対し、GDコーパスの参加者は全て日本人母語話者であった。コーパス参加者が異なるため、本研究の結



果と上記の結果とを直接比較することはできないが、同様に評定値の高低の2クラスを分類するタスクで、最低で0.83, 最大で0.93と上回ることを示した。

## 7. 考 察

### 7.1 能力値の推定に有効な特徴量の分析

学習後の線形SVMの超平面  $f(x) = w^T x$  における, 正規化済みの重み  $w$  を分析することで, 各項目の推定に寄与した特徴量を明らかにする.  $w$  の各成分は各特徴量に係る重みを示しており, 正の値である場合は評定値の高群(正例のクラス)を, 負の値である場合は評定値の低群(負例のクラス)を識別するためにそれぞれ重要である.

分類実験より, 最大の精度(図3)が得られたモデルの重みを分析する. 表3に各項目の分類モデルの重みを記載する. 「総合的コミュニケーション能力」および, 要素項目である「双方向の円滑なコミュニケーション」, 「意見集約力」, 「情報伝達力」, 「論理的で明瞭な主張」の最大分類精度はそれぞれ0.93, 0.85, 0.83, 0.87, 0.85であり, 分類精度が最大であった特徴量セットはそれぞれ, 「韻律+言語+動作 ( $A+L+M$ )」, 「韻律+言語+動作 ( $A+L+M$ )」, 「韻律+言語+動作 ( $A+L+M$ )」, 「韻律+発話ターン+言語 ( $A+S+L$ )」, 「韻律+発話ターン+言語 ( $A+S+L$ )」であった.

いずれの項目の推定結果も, 2クラス分類のチャンスレベルである0.5を大幅に上回っており, これらの高精度な推定モデルの重みを分析することで推定に寄与する特徴量を明らかにできると考える. また評定者が高く(または低く)評価した参加者のマルチモーダル特徴量を分析することが同時に可能となる. 表3において, 各列は項目ごとの重みを示し, 各行は特徴量ごとの重みを示す. 分類実験では計10回のテストを試行したため, 10回の学習で構築されたモデルの重みの平均を表3に記載している. 正の重み, 負の重みの値の上位5位までを, 太字, 下線付き太字でそれぞれ示している. 各項目で有効な特徴量を分析する.

「双方向の円滑なコミュニケーション」

評定値の推定には韻律, 言語, 動作特徴量の統合が有効であった. 韻律特徴量に関して, 高群を推定するために有効であった韻律特徴量は, 最小ピッチ, 最大インテンシティ, 音圧の幅であり, 特徴量の  $w$  はそれぞれ1.26, 1.26, 1.75であった. また低群の推定に有効な特徴量はピッチ平均, 最少インテンシティ, 抑揚であり, 特徴量の  $w$  はそれぞれ-1.49, -0.49, -0.15であった. インテンシティに関して, 声の大きさの最大値, また音圧の幅が大きく, 最小値が小さいほど高群に分類される傾向にある. つまりある場面では大きい声で発話を行い, ある場面では小さい声で発話を行うといった, 声の大小の使い分けがコミュニケーション能力の高低の推定に有効に働くことが推察される. ピッチに関して, 平均ピッチが

小さく, 最小ピッチが大きく, 抑揚(最大・最小ピッチの差)が小さいほど高群に分類される傾向にある.

言語特徴量に関して, 感動詞が発話内容に多く含まれる参加者は評定値の高群に分類される傾向にある. 一方で, 動詞が発話内容に多く含まれる参加者は評定値の低群に分類される傾向にある. 感動詞, 動詞の特徴量  $w$  はそれぞれ0.95, -0.22であった.

動作特徴量に関して, セッション全体を通じた動作量の偏差が大きいくほど, 低群に分類される( $w = -1.33$ ), 一方で, 発話中の動作量の標準偏差が大きいくほど, 高群に分類される( $w = 1.40$ ). 動作量の標準偏差は, 動作の強弱の大きさと相関していると考えられる.

上記をまとめると, 声は大きく, また発言中の声の平均は高すぎない, また発話中に動作の強弱をつけ, 発言中に感動詞(あいづちや応答)を含むほど, 双方向の円滑なコミュニケーションを行っている印象を受ける傾向にあることが明らかになった.

「意見集約力」

「双方向の円滑なコミュニケーション」と同様に, 評定値の推定には韻律, 言語, 動作特徴量の統合が有効であった. 韻律特徴量, 動作特徴量に関して, 概ね「双方向の円滑なコミュニケーション」と同様の傾向を示した. 特徴として動作量平均が, 小さい負の重みをもつ( $w = -0.07$ ).

言語特徴量に関して, 重みの大きい特徴量が「双方向の円滑なコミュニケーション」のそれと異なっていた. 名詞数・新規名詞数が発話内容に多く含まれる参加者は評定値の高群に分類される傾向にある. 一方で, 発言中のフィラーの特徴量が負の重み( $w = -0.44$ )を有しており, フィラーの多い参加者は意見集約力の低群に分類される傾向にある. フィラー特徴の重みが上位にランクされたのは「意見集約力」のみであり, この評定値を推定するために特有の特徴量であった.

「双方向の円滑なコミュニケーション」の高群参加者の発言にはあいづちを含む感動詞の回数が多い. これに対し「意見集約力」では, 名詞や新規名詞の回数が多く, フィラーが少ないという特徴があることが示された.

「情報伝達力」

評定値の推定には韻律, 発話ターン, 言語特徴量の統合が有効であった. 発話ターンに関して, 全発話回数が高群の推定に有効である( $w = 0.15$ ). 言語特徴に関して名詞数, 動詞数, 新規名詞数, 新規名詞数を発話回数で正規化した特徴量(正規化新規名詞数)が発話内容に多く含まれる参加者は評定値の高群に分類される傾向にある. 参加者は情報を十分に伝達するためには多く発言する必要があるため発話回数が多く, 発話内容に名詞が多く含まれる参加者が高い評価を得たという結果は妥当であると考えられる.

「論理的で明瞭な主張」

評定値の推定に韻律, 発話ターン, 言語特徴量の統合が有効であった. 韻律特徴に関して, 低群の推定に有効な

表 3 各特徴量の識別への寄与を示す識別境界面の重み係数 (3 列目の重み係数に関して, 分類精度が最大であった特徴量セットで学習した線形 SVM の識別境界面における各特徴量の重み  $w$  を記載する. 10 回の試行で学習されたモデルの  $w$  を平均した値を比較する. 各項目に, 上位 5 位の正の重み係数を太字で, 同じく負の重み係数を下付き太字で示す.)

	要素項目				総合的なコミュニケーション能力
	双方向の円滑なコミュニケーション	意見集約力	情報伝達力	論理的で明瞭な主張	
分類精度最大のモデルの特徴量セット	$A+L+M$	$A+L+M$	$A+S+L$	$A+S+L$	$A+L+M$
平均分類精度 (図 3 における数値)	0.85	0.83	0.87	0.85	0.93
	SVM の $w$	SVM の $w$	SVM の $w$	SVM の $w$	SVM の $w$
最大ピッチ	0.674	0.494	0.028	0.037	0.483
最小ピッチ	<b>1.258</b>	0.556	0.032	0.012	<b>1.210</b>
ピッチ平均	<b>-1.488</b>	<b>-1.306</b>	0.016	0.012	<b>-1.040</b>
最大インテンシティ	<b>1.257</b>	<b>1.483</b>	0.068	0.031	<b>1.519</b>
最小インテンシティ	<b>-0.494</b>	<b>-0.669</b>	0.012	<b>-0.017</b>	-0.203
音圧の幅	<b>1.751</b>	<b>2.152</b>	0.056	0.048	<b>1.722</b>
抑揚	<b>-0.154</b>	0.276	0.015	0.052	<b>-0.437</b>
話速	0.042	-0.001	0.004	<b>-0.006</b>	-0.003
合計発話長			0.134	<b>0.121</b>	
合計発話回数			<b>0.150</b>	0.118	
合計発話長 (1 秒以上)			0.119	0.114	
合計発話回数 (1 秒以上)			0.137	0.117	
名詞数	0.252	<b>0.796</b>	<b>0.156</b>	<b>0.133</b>	0.610
動詞数	<b>-0.215</b>	0.355	<b>0.140</b>	<b>0.127</b>	0.084
感動詞	<b>0.945</b>	0.516	0.022	0.024	0.759
フィラー数	0.154	<b>-0.440</b>	0.049	0.055	0.106
新規名詞数	0.055	<b>1.120</b>	<b>0.180</b>	<b>0.139</b>	0.556
既出名詞数	0.509	0.065	0.078	0.089	0.542
新規名詞数 / 発話回数	0.853	0.071	<b>0.179</b>	<b>0.149</b>	<b>0.849</b>
動作量の平均	0.333	<b>-0.073</b>			<b>-1.166</b>
動作量の偏差	<b>-1.330</b>	<b>-1.574</b>			<b>-1.787</b>
発話中の動作量の平均	0.118	0.173			<b>-1.228</b>
発話中の動作量の偏差	<b>1.404</b>	<b>1.461</b>			<b>2.056</b>

特徴量は最少インテンシティ, 話速であった. また発話ターン特徴に関して, 合計発話長は高群の推定に有効であった ( $w = 0.12$ ). 言語特徴に関しては「情報伝達力」の傾向と類似しており, 動詞数, 名詞数, 新規名詞数が発話内容に多く含まれる参加者は評定値の高群に分類される傾向にある. また話速が遅く, 最少インテンシティが低く, 音圧, 最大インテンシティが正の値であるといった特徴から, ゆっくり発話を行い, 声の強弱を付けながら発言する参加者が高群に分類される傾向にある. この傾向は「明瞭な主張」という評価項目に合致していると考えられる.

#### 「総合的なコミュニケーション能力」

総合的なコミュニケーション能力を推定するために評定値の推定には韻律, 言語, 動作特徴量の統合が有効であった. また全体的な傾向としては要素項目の有効な特徴量の性質を統合した内容となっている.

韻律特徴に関しては「双方向の円滑なコミュニケーション」「意見集約力」の傾向と類似する. 高群の推定に最小ピッチ, 最大インテンシティ, 音圧の幅が有効であり, 低群の推定に, ピッチ平均, 抑揚が有効である. 音圧の幅の  $w$  は 1.72 であり韻律特徴の中で最大であった.

言語特徴に関して「情報伝達力」や「論理的で明瞭な主張」で重要であった, 正規化新規名詞数は正の重みを有

しており, 高群を推定するために有効であった. 上位には現れなかったが, 名詞数 ( $w = 0.61$ ), 感動詞 ( $w = 0.76$ ) といった特徴量も正の重みを有しており, 言語特徴量が評価値の高群推定に有効であることを示している.

動作に関しては「双方向の円滑なコミュニケーション」, 「意見集約力」と同様の傾向でありセッション中の動作量の平均, 標準偏差, 発話中の動作量の平均共に, 低群の推定に有効である. 一方で, 発話中の動作量の標準偏差は高群の推定に有効であった.

#### 7.2 実用化に向けた推定モデルの頑健性の検証

本節では SVM の出力値から ROC 曲線を計算し, モデルの識別における頑健性を考察する. 図 4 に総合的なコミュニケーション能力に関して SVM から ROC 曲線をプロットした. 図 4 では低群を正解ラベル (Positive) として, 単一モダリティの特徴量を用いて学習したモデルの ROC 曲線と, 最大の分類精度をもつ全特徴量を用いて学習したモデルの ROC 曲線を比較した. 横軸は False Positive (FP) で, 低群のデータを誤って高群に分類する割合であり, 縦軸は True Positive (TP) で, 低群のデータを正しく分類する割合を示す. また図中には曲線と横軸の間の面積である AUC (Area Under the Curve) を記載した. 識別に頑健なモデルでは False Positive が低く,

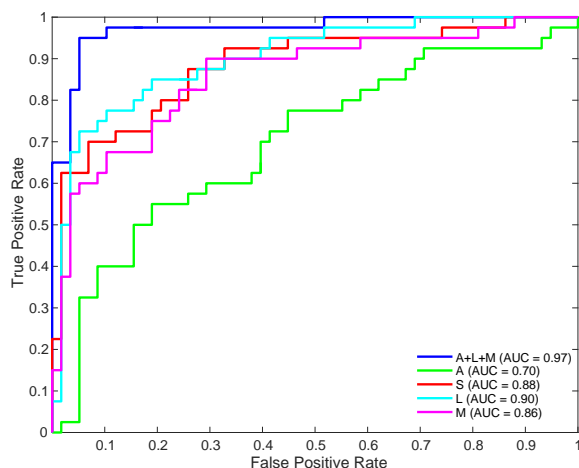


図4 コミュニケーション能力値の分類タスクのテストデータより計算されたROC曲線(単一モダリティ(A, S, L, M)のモデルによる曲線と, 最大分類精度を得た(A+L+M)のモデルによる曲線を比較している.)

True Positive が高く, 結果として AUC の値が大きくなる(1 に近づく). 図より,  $A+L+M$  を用いたモデルの AUC は 0.97 であり, 他の単一モダリティのモデルのそれを上回っている. さらに,  $FP=0.05$  において, TP は 0.95 に達する. これは, 高群のデータの誤分類精度を低く保ったまま, 95% の精度で低群のデータを正しく分類できることを示している. この分析より, 提案するマルチモーダル特徴量を統合したモデルは, ユーザの能力を推定するモデルとして有用であることを示唆した.

### 7.3 モデルの実用性の検証

7.1 節では, コミュニケーション能力の各項目の高・低を推定するために有効な特徴量を明らかにした. この特徴量は, 能力値が高い, または低いと評定された参加者の会話における所作を捉えており, コミュニケーション行動の可視化システムに利用できると考える. 例えば, 推定モデルに基づき能力値の低いと判定されたユーザには, 高いユーザの特徴を可視化・提示することで, 自身の会話行動への気づきを与えられると考える.

7.2 節では, 今回構築した分類モデルの識別に関する頑健性を示した. 推定モデルを支援システムなどに導入することを想定する場合, 本研究で定義した特徴量を全て自動抽出する必要がある.

発話ターン, 韻律特徴量について, 今回のように4名の着座会話がノイズ音の少ないクリーンな音環境で収録される場合であれば, 各参加者が装着した指向性マイクを用いて各話者の音声区間を十分な精度で検出可能であるため, 自動抽出を行える可能性が高い.

動作特徴量について頭部に装着した加速度センサを利用して取得したが, 今後はユーザの装着負荷をなくするために, カメラなどの非接触センサを用いて, 動作特徴量を抽出することが望ましい. RGB カメラや深度センサが

ら抽出した動作特徴量で代用可能かを検証することは今後の課題である.

言語特徴量について, 議論タスクや議論内容に極力依存しないよう, 品詞の頻度を特徴量に採用した. この品詞特徴量は, 能力値の推定に有用であった. 実用化に向けては, 音声認識結果から各参加者の発話内容に含まれる品詞頻度を正確に計算する必要があり, これについても今後の課題である.

## 8. 結 論

本研究では, 会話参加者の表出するマルチモーダル情報から「コミュニケーション能力」を推定するモデルの構築・評価を行った. 発話内容に含まれる単語の品詞, 発話ターン, 韻律情報, 頭部動作量をマルチモーダル特徴量として抽出し, 人事採用経験者より評定された各参加者のコミュニケーション能力値を推定するモデルを機械学習により構築した. 評価実験の結果, 総合的なコミュニケーション能力に関する評定値の推定に関して, 回帰タスクで最大 0.62 の決定係数  $R^2$ , 高群・低群の2クラス分類タスクで最大 0.93 の精度を得た. また, 各要素項目と総合的なコミュニケーション能力を識別するために有効な特徴量を明らかにし, コミュニケーション能力の高・低に分類される参加者に見られる特徴を明らかにした. 今後, 7.3 節で述べた課題を解決し, 推定システムの実用化を目指す.

また, 本研究ではディスカッションの結果や質の評価を行っていない. したがって, 本研究でモデル化した意思疎通に関する能力と, ディスカッションの結果や質の関連性は明らかになっていない. ディスカッションの収束性や結果の評価に関する分析は, ディスカッションの良し悪しを判断する上でも重要である. よって, グループディスカッションの結果や質を評価する方法を検討し, 個人のコミュニケーション能力との関連性を分析することも今後の課題である.

## 謝 辞

本研究は JSPS 科研費 25280076, 15K00300 の助成を受けたものです.

## ◇ 参 考 文 献 ◇

- [Aran 13] Aran, O. and Gatica-Perez, D.: One of a Kind: Inferring personality impressions in meetings, in *Proc. of ACM ICMI*, pp. 11–18, ACM (2013)
- [Boersma 13] Boersma, P. and Weenink, D.: *Praat: Doing Phonetics by Computer*, [Computer program], Version 5.3.51 (2013)
- [Brugman 04] Brugman, H., Russel, A., and Nijmegen, X.: Annotating multi-media/multi-modal resources with ELAN, in *Proc. of LREC* (2004)
- [Chen 09] Chen, L. and Harper, M. P.: Multimodal floor control shift detection, in *Proc. of ACM ICMI*, pp. 15–22, ACM (2009)
- [Gatica-Perez 09] Gatica-Perez, D.: Automatic nonverbal analysis of social interaction in small groups: A review, *Image Vision Comput-*



- ing, Vol. 27, No. 12, pp. 1775–1787 (2009)
- [林 15] 林 佑樹, 二瓶 英巳雄, 中野 有紀子, 黄 宏軒, 岡田 将吾: グループディスカッションの構築および性格特性との関連性の分析, 情報処理学会論文誌, Vol. 56, No. 4, pp. 1217–1227 (2015)
- [Ishii 13] Ishii, R., Otsuka, K., Kumano, S., Matsuda, M., and Yamato, J.: Predicting next speaker and timing from gaze transition patterns in multi-party meetings, in *Proc. of ACM ICMI*, pp. 79–86, ACM (2013)
- [JAVADA 05] JAVADA 中央職業能力開発協会: 若年者就職基礎能力修得のための目安策定委員会報告書, 厚生労働省 (2005)
- [Jayagopi 10] Jayagopi, D. and Gatica-Perez, D.: Mining group non-verbal conversational patterns using probabilistic topic models, *IEEE Transactions on Multimedia*, Vol. 12, No. 8, pp. 790–802 (2010)
- [Jovanovic 07] Jovanovic, N.: To whom it may concern: address identification in face-to-face meetings, *CTIT Ph.D. thesis series 07-96* (2007)
- [Knapp 13] Knapp, M., Hall, J., and Horgan, T.: *Nonverbal Communication in Human Interaction*, Cengage Learning (2013)
- [McCowan 05] McCowan, I., Gatica-Perez, D., Bengio, S., Lathoud, G., Barnard, M., and Zhang, D.: Automatic analysis of multimodal group actions in meetings, *IEEE Transactions on Pattern Analysis Machine Intelligence*, Vol. 27, No. 3, pp. 305–317 (2005)
- [中田 11] 中田 篤志, 角 康之, 西田 豊明: 非言語行動の出現パターンによる会話構造抽出, 電子情報通信学会論文誌, Vol. 94, No. 1, pp. 113–123 (2011)
- [Nguyen 14] Nguyen, L., Frauendorfer, D., Mast, M., and Gatica-Perez, D.: Hire me: computational inference of hirability in employment interviews based on nonverbal behavior, *IEEE Transactions on Multimedia*, Vol. 16, No. 4, pp. 1018–1031 (2014)
- [Nishida 08] Nishida, T.: *Conversational informatics: an engineering approach*, Vol. 9, Wiley, com (2008)
- [西田 09] 西田 豊明, 角 康之, 松村 真宏: 社会知デザイン, 知の科学, オーム社 (2009)
- [Okada 15] Okada, S., Aran, O., and Gatica-Perez, D.: Personality trait classification via co-occurrent multiparty multimodal event discovery, in *Proc. of ACM ICMI*, pp. 15–22, ACM (2015)
- [Otsuka 07] Otsuka, K., Sawada, H., and Yamato, J.: Automatic inference of cross-modal nonverbal interactions in multiparty conversations: “Who Responds to Whom, When, and How?” from gaze, head gestures, and utterances, in *Proc. of ACM ICMI*, pp. 255–262, ACM (2007)
- [Pianesi 08] Pianesi, F., Mana, N., Cappelletti, A., Lepri, B., and Zancanaro, M.: Multimodal recognition of personality traits in social interactions, in *Proc. of ACM ICMI*, pp. 53–60, ACM (2008)
- [Sanchez-Cortes 12] Sanchez-Cortes, D., Aran, O., Mast, M. S., and Gatica-Perez, D.: A nonverbal behavior approach to identify emergent leaders in small groups, *IEEE Trans. on Multimedia*, Vol. 14, No. 3, pp. 816–832 (2012)
- [Sanchez-Cortes 13] Sanchez-Cortes, D., Aran, O., Jayagopi, D. B., Mast, M. S., and Gatica-Perez, D.: Emergent leaders through looking and speaking: From audio-visual data to multimodal recognition, *Journal on Multimodal User Interfaces*, Vol. 7, No. 1-2, pp. 39–53 (2013)
- [Tanaka 15] Tanaka, H., Sakti, S., Neubig, G., Toda, T., Negoro, H., Iwasaka, H., and Nakamura, S.: Automated social skills trainer, in *Proc. of IUI*, pp. 17–27, ACM (2015)

〔担当委員: 栗原聡, 山川宏, 矢入健久〕

2015 年 12 月 10 日 受理

ム科学専攻助教。2014 年 IDIAP Research Institute 滞在研究員。2016 年より東京工業大学大学院情報理工学専攻助教。博士(工学)。ACM, 電子情報通信学会各会員。FIT2012 ヤングリサーチ賞。2015 年 本会研究会優秀賞を受賞。マルチモーダルインタラクション, 人間行動解析, 社会的信号処理, パターン認識の研究に従事。

松儀 良広

2015 年 成蹊大学理工学部情報科学科卒業, 現在, 東京工業大学大学院知能システム科学専攻修士課程在学中。



中野 有紀子(正会員)

1990 年 東京大学大学院教育学研究科修士課程修了。同年, 日本電信電話(株)入社。2002 年 MIT Media Arts & Sciences 修士課程修了。2002–2005 年(独)科学技術振興機構社会技術研究開発センター専門研究員。2005–2008 年 東京農工大学大学院工学府特任准教授, 成蹊大学理工学部情報科学科准教授を経て, 現在, 成蹊大学理工学部情報科学科教授。知的で自然なユーザインタフェースの実現に向けて, 人との言語・非言語コミュニケーションが可能な会話エージェントの研究に従事。博士(情報理工学)。ACL, ACM, 電子情報通信学会各会員。



林 佑樹(正会員)

2007 年 名古屋大学工学部電気電子情報工学科卒業。2012 年 同大学院情報科学研究科博士課程後期課程修了。博士(情報科学)。2009–2012 年 日本学術振興会特別研究員。2012–2014 年 成蹊大学理工学部情報科学科助教を経て, 2014 年より大阪府立大学現代システム科学域知識情報システム学類助教。人の知的活動の高度化を目指して, 多人数コミュニケーションや思考スキルに関する研究に従事。情報処理学会, ヒューマンインタフェース学会, 教育システム情報学会各会員。



黄 宏軒(正会員)

2009 年 京都大学大学院情報科学研究科知能情報学専攻博士後期課程修了。博士(情報学)。同年, 成蹊大学理工学部情報科学科ポスドク研究員。2010 年より立命館大学情報理工学部情報コミュニケーション学科助教, 2013 年より同准教授, 現在に至る。主に会話エージェント, マルチモーダル・ヒューマン・マシン・インタラクションの研究に従事。情報処理学会, 電子情報通信学会, ヒューマンインタフェース学会, 台湾人工知能学会, ACM 各会員。



高瀬 裕

2006 年 東京都立大学理学部物理学科卒業, 2008 年 総合研究大学院大学物理科学研究科天文科学専攻修了。2014 年 東京工業大学知能システム科学専攻博士課程修了。同年, 成蹊大学理工学部情報科学科助教, 現在に至る。ヒューマンインタフェースの研究に従事。博士(工学)。



新田 克己(正会員)

1980 年 東京工業大学大学院博士課程修了。同年, 電子技術総合研究所に入所。1989–1993 年まで(財)新世代コンピュータ技術開発機構に出向。1994 年 東京工業大学大学院総合理工学研究科教授。工学博士。2012 年本会論文賞, 2013 年本会業績賞を受賞。法的推論システム, ヒューマンインタフェース, マルチエージェントシステムの研究に従事。情報処理学会, 電子情報通信学会, ヒューマンインタフェース学会, 情報ネットワーク法学会, IEEE 各会員。



## 著者紹介

岡田 将吾(正会員)

2008 年 東京工業大学大学院知能システム科学専攻修士課程修了。同年, 京都大学情報科学研究科知能情報専攻特定助教, 2011 年 東京工業大学大学院総合理工学研究科知能システ

