グループディスカッションにおいて現れるコミュニケーション能力のマルチモーダル分析

A multimodal analysis for estimating communication skills in group discussions

岡田 将吾[1*]　松儀 良広[2]　中野 有紀子[3]　林 佑樹[4]
黄 宏軒[5]　高瀬 裕[3]　新田 克己[1]
Shogo Okada[1] Yoshihiro Matsugi[2] Yukiko Nakano[3] Yuki Hayashi[4]
Hung-Hsuan Huang[5] Yutaka Takase[3] Katsumi Nitta[1]

[1] 東京工業大学 情報理工学院
[1] School of Computing, Tokyo Institute of Technology
[2] 東京工業大学 総合理工学研究科　知能システム科学専攻
[2] Dept. of Computational Intelligence and Systems Science, Tokyo Institute of Technology
[3] 成蹊大学 理工学部情報科学科
[3] Department of Computer and Information Science, Seikei University
[4] 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
[4] College of Sustainable System Sciences, Osaka Prefecture University
[5] 立命館大学 情報理工学部情報コミュニケーション学科
[5] Department of Information and Communication Science, College of Information Science and Engineering, Ritsumeikan University

Abstract: This paper presents an computational analysis for estimating communication skills of each participant in a group. For this purpose, we use a multimodal group meeting corpus including audio, visual, head motion sensor data and communication skill indices of participants. The communication skills of each participant is assessed by external observers. We extract multimodal features including spoken utterances, acoustic features, speaking turns and head activity for estimating communication skills. In the experiment, we modeled the relationship between the skill indices and the multimodal features using a machine learning technique: support vector machine. The result reports effective features in predicting the level of communication skill.

1　はじめに

近年，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・訓練基盤整備が必要とされている [JAVADA 05]，このような背景において，我々は，アイディア創出や，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目指している．その初期検討として，マルチモーダル情報に基づくコミュニケーション能力の分析を行う．本研究では，[JAVADA 05] で定義された「意思疎通能力」を，コミュニケーション能力と定義する．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（非言語情報）と発言内容（言語情報）のマルチモーダル情報から，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．機械学習によるモデリングを通じて，参加者の評定値の高・低を識別する上で有効な特徴量を特定し，人事採用経験者より高・低評価をうけた参加者の言語・非言語特徴を明らかにする．

*連絡先：東京工業大学 情報理工学院
〒 226-8502 横浜市緑区長津田町 4259
E-mail:okada@c.titech.ac.jp

2 関連研究

社会心理学・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]．なかでも，参加者の個人特性を分析・モデル化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．

Nguyen ら [Nguyen 14] はアルバイト面接場面において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．[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 者対話に関する分析・モデル化であり，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．

Sanchez-Cortes らは ELEA コーパス [Sanchez-Cortes 13] 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て，問題解決を行う過程でリーダシップを発揮した参加者を推定した [Sanchez-Cortes 12]．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（**G**roup **D**iscussion）コーパスと呼称する．GD コーパスには，質問紙調査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，加速度センサからの頭部加速度，Kinect からの深度情報，指向性マイクから集音した音声情報，Web カメラからの顔映像，光学式モーションキャプチャを使用して取得した頭部位置座標，ビデオカメラを使用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，加速度センサから観測した頭部加速度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 人の参加者を 1 つのグループとして，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．

3.2 コミュニケーション能力値のアノテーション

[JAVADA 05] で定義される意思疎通能力は「傾聴する姿勢」,「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 5 つの要素項目で構成される．本研究では，その全ての要素項目を考慮して，人事採用経験者に評定を行ってもらい，参加者の「総合的なコミュニケーション能力」を決定した．

人事採用経験者は人材派遣会社を通じて募集した．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 つの課題毎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，5 つの要素項目と総合的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（クローンバック α 値）を計算し，各項目に関する信頼性を確認した．「傾聴する姿勢」($\alpha = 0.66$) 以外のすべての項目で α は 0.8 を上回った．分析対象である「総合的なコミュニケーション能力」の α は 0.91 と最大であった．この結果より評定値の信頼性は確保されたと考える．以降では，21

名による「総合的なコミュニケーション能力」の評定値の平均値を分析・教師付き学習の教師データに用いる．

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，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．

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．

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．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，発話区間の 0.1 秒ごとの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．

最大シンテンシティ，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．インテンシティは声の大きさを示す値である．

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．上記と同様に発話中の動作量の平均，標準偏差を計算する．

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[1] 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，会話内容は多岐に渡る．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．

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，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，参加者ごとに計算した発言回数を，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．

[1] MeCab:http://taku910.github.io/mecab/

5 評価実験

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分類モデルの評価実験を行った．

5.1.1 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 σ 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．

5.1.2 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 9 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．以下の 9 種類の特徴量セットを用意した．

4 種類の単一モダリティの特徴量セット：
(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 種類のマルチモーダル特徴量セット：
(5) $A+S+L$: 韻律と発話ターンと言語,
(6) $A+S+M$: 韻律と発話ターンと動作,
(7) $A+L+M$: 韻律と言語と動作,
(8) $S+L+M$: 発話ターンと言語と動作,
(9) All: 全ての特徴量 $(A+S+L+M)$

5.2 実験結果

表 1 に SVM を用いた分類タスクの認識精度結果を示す．単一のモダリティ A, S, L, M（(1)〜(4)）を用いたモデルの性能を比較すると,「言語（L)」のモデルの精度が最大であり 0.85 であった．マルチモーダル特徴量を用いたモデルも含め，全モデルの性能を比較すると,「韻律＋言語＋動作（$A+L+M$)」のモデルの精度が最大であり 0.93 であった．

表 1: SVM を用いた分類タスクの結果

特徴量	分類精度
(1) A:	0.68
(2) S:	0.82
(3) L:	0.85
(4) M:	0.64
(5) $A+S+L$:	0.81
(6) $A+S+M$:	0.82
(7) $A+L+M$:	**0.93**
(8) $S+L+M$:	0.81
(9) All:	0.90

この推定精度は全ての単一モダリティ特徴量を用いた場合の精度を 5% 水準で有意に上回っている．この結果よりマルチモーダル特徴の統合は，コミュニケーション能力の高低の識別に有効であることが示された．

先行研究である [Okada 15] ではリーダシップ（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，最大で 0.93 と上回ることを示した．

5.3 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（表 1）が得られたモデルの重みを分析する．表 2 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．

表 2: 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．)

分類精度最大のモデルの特徴量セット	$A+L+M$
	SVM の w
最大ピッチ	0.483
最小ピッチ	**1.210**
ピッチ平均	**-1.040**
最大インテンシティ	**1.519**
最小インテンシティ	-0.203
音圧の幅	**1.722**
抑揚	**-0.437**
話速	-0.003
名詞数	0.610
動詞数	0.084
感動詞	0.759
フィラー数	0.106
新規名詞数	0.556
既出名詞数	0.542
新規名詞数 / 発話回数	**0.849**
動作量の平均	**-1.166**
動作量の偏差	**-1.787**
発話中の動作量の平均	**-1.228**
発話中の動作量の偏差	**2.056**

韻律特徴に関して，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，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，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．

6 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・評価を通じて，グループディスカッション中に現れるコミュニケーション能力の分析を行った．機械学習の結果，高群・低群の 2 クラスの分類タスクで最大 0.93 の分類精度を得た．また，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．

本論文ではマルチモーダル分析の初期検討を述べたが，コミュニケーション能力を構成する要素項目に関する分析や，コミュニケーション能力値を推定する回帰学習を通じた分析は本研究では行われていない．これらについては，別途報告を行う予定である．

謝辞

本研究は JSPS 科研費 25280076，15K00300，15H02746 の助成を受けたものです．

参考文献

[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 *LREC* (2004)

[Gatica-Perez 09] Gatica-Perez, D.: Automatic nonverbal analysis of social interaction in small groups: A review, *Image Vision Computing*, Vol. 27, No. 12, pp. 1775–1787 (2009)

[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)

[Knapp 13] Knapp, M., Hall, J., and Horgan, T.: *Nonverbal communication in human interaction*, Cengage Learning (2013)

[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 Trans. on Multimedia*, Vol. 16, No. 4, pp. 1018–1031 (2014)

[Okada 15] Okada, S., Aran, O., and Gatica-Perez, D.: Personality Trait Classification via Co-Occurrent Multiparty Multimodal Event Discovery, Proc. of ACM ICMI, pp. 15–22 (2015)

[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 *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pp. 17–27 (2015)

[林 15] 林 佑樹, 二瓶 芙巳雄, 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)