

**科学研究費助成事業 研究成果報告書**

平成 27 年 6 月 8 日現在

機関番号：12608

研究種目：若手研究(B)

研究期間：2013～2014

課題番号：25730132

研究課題名(和文)系列マイニングを用いた説明状況下に出示する非言語情報の機能解析・モデル化

研究課題名(英文) Analysis and modeling of nonverbal behaviors in explanation using time-series multimodal analysis

研究代表者

岡田 将吾 (Okada, Shogo)

東京工業大学・総合理工学研究科(研究院)・助教

研究者番号：00512261

交付決定額(研究期間全体)：(直接経費) 2,200,000円

研究成果の概要(和文)：様々な事象を説明することは、プライベート・ビジネス問わず、情報共有を行うために重要なプロセスである。本研究は会話者の説明の上手さ(スキル)を測定するための計算機モデルを構築する。このモデルを用いることで、各会話者の説明スキルや良質な説明シーンを自動検出したり、「ジェスチャを頻繁に行いながら説明を行う」といった説明スキルが高い人の特徴を抽出出来る。本研究では説明者・聞き手を含む会話者間の対面会話において観測される、発話言語・非言語情報を抽出し、外部観察者により評価された説明者の説明スキル値を関連付ける。複数の会話データを学習したモデルを用いて、説明者のスキル値の高・低を82%の精度で推定できた。

研究成果の概要(英文)：This research focuses on modeling the explanation-performance of participants in group conversation. We present a multimodal analysis of explanation performance in group conversation as evaluated by external observers. A new multimodal data corpus, including the performance score of participants, is collected through group storytelling task. We extract multimodal features regarding explanators and listener from a manual description of spoken dialog and from various nonverbal patterns, including speaking turn, utterance prosody, head gesture, hand gesture, and head direction of each participant. We also extract multimodal co-occurrence features, such as utterance with head gestures. In the experiment, we modeled the relationship between the performance indices and the features using machine learning techniques. Experimental results show that the highest accuracy is 82% for the total explanation performance obtained with a combination of these features in a binary classification task.

研究分野：マルチモーダルインタラクション

キーワード：社会的信号処理 パターン認識 会話分析 データマイニング

## 1. 研究開始当初の背景

多人数インタラクションにおける、各参加者の言語・非言語を含むマルチモーダル情報からコミュニケーション行為を理解・モデル化する研究が行われてきた。マルチモーダル情報をマイク・カメラ・モーションセンサ等でコーパスとして集積し、非言語パターンの自動認識を行うシステムの開発がプロジェクトで行われてきた。視線、表情、頭部ジェスチャ、ハンドジェスチャといった多チャンネルの非言語情報のセンシングを通じて、会話者の態度・個人特性・感情などを認識する技術の開発が盛んに行われてきた。

## 2. 研究の目的

本研究の目的は、説明者が聞き手に説明する場面において、説明者の説明スキルなどの説明行為の質に関わる要因と、説明者・聞き手の表出する非言語パターンの間にある関係性を分析の結果明らかにすることである。説明者の非言語行為・聞き手の発話頻度やうなづきといった非言語行為を抽出し、それらの特徴量と説明スキルの関係性との関係性を解析する。本研究では各種センサから抽出した視線・うなづき・ハンドジェスチャなどの非言語系列を自動アノテーションするデータマイニング・機械学習の手法に基づいて、説明者・聞き手の多様な非言語情報を抽出することで、説明スキル値に相関の高い特徴量を列挙すると共に、説明スキルを予測出来るモデルを構築する。

## 3. 研究の方法

### (1) センシング環境の準備

会話依存データの収集まずは若手Bのプロジェクトで収集した、三人会話データセットを用いて、データマイニングを行う。ただしデータの量が十分でない場合、各会話形態の会話データを取得する。

### (2) データの収集

説明タスクにおける非言語・言語インタラクションの活動を記録する。非言語情報：手のジェスチャに関しては、地磁気・加速度・ジャイロセンサを実験協力者に装着してもらいデータを取得した。頭の動作の縦方向成分よりうなづきパターンを検出し、頭の動作の横方向成分からどの話者に視線を向けているかを近似的に求める。

### (3) 説明タスクの概要

説明タスクについて説明者2人が1人の聞き手にアニメーションの内容を説明するタスクを設定した。本タスクでは、説明内容情報を複数人が共有している状況で、協調的に説明する様子を観測できる。取得した10セッションのデータをデータマイニングに用いて取得する非言語モダリティについては頭部の動き・手の動きを動作センサから、韻律・声の大きさ・発話の有無を説話マイクから取得する。言語については、書き起こしコーパスから単語特徴量を抽出する。

(4) 会話構造に依存したイベントを抽出できるマイニング手法の開発・マイニングの実施  
グラフマイニングに基づき、各参加者の非言語パターンとの共起関係を列挙した。一方で、マイク・動作センサ・顔認識モジュールから推定した顔方向データから、取得したローデータより各話者の時間軸に沿って表出された非言語パターンを時系列データマイニングシステムにより半自動的にアノテーションした

### (5) 抽出された非言語パターンと質問紙調査から得られた説明スキル指標の間の相関分析、回帰分析、機械学習

会話参加者の質問紙調査だけでは主観的な評価が行われるため、同時に第三者に説明タスクのビデオを見せ、説明の流暢さ・わかりやすさに関し調査を行う。これらの質問紙調査から得られた、各セッションの説明者の説明に関する評価値とセンサから抽出された非言語インタラクションイベントの間の分析・モデリングを行った。

## 4. 研究成果

本研究の成果は以下の5点である

- (1) 説明会話における会話者全員の言語・非言語のマルチモーダル情報、説明スキル評価値を含むデータコーパスを新規に構築した。
- (2) マイク、モーションキャプチャ、加速度センサから、音声区間、韻律情報、ハンドジェスチャ、頭部ジェスチャ、顔向き方向のラベルを自動的にアノテーションする枠組みを構築し、「ジェスチャをしながら発話する」といった共起パターンを多次元時系列データから効率的に列挙する系列マイニングアルゴリズムを構築した。
- (3) 抽出したマルチモーダル情報と説明スキルの間の関係を多変量解析と機械学習技術によりモデル化、評価を行った。
- (4) 発表会場におけるポスタ説明のように、多数の聞き手が流動する中で、説明者と聞き手、某参加者などの非言語行動をセンシングするために複数台の深度センサを利用したセンシング環境を構築した。
- (5) (2,3)で構築したマルチモーダル非言語情報のラベリング・モデリング技術を、患者の状態推定に応用した。

### (1) 説明会話データコーパスの収集

本研究では、研究代表者がJSPS若手B(課題番号22700146, 2010-2012)において収集した、会話データセットを基に、複数のアノテーションデータを追加し、形で、説明スキルを評価するための新規データセットを構築した。収録した8セットの説明タスク会話に対して、3名の外部アノデータに、説明スキルの評価を依頼し、説明者の説明スキル指標

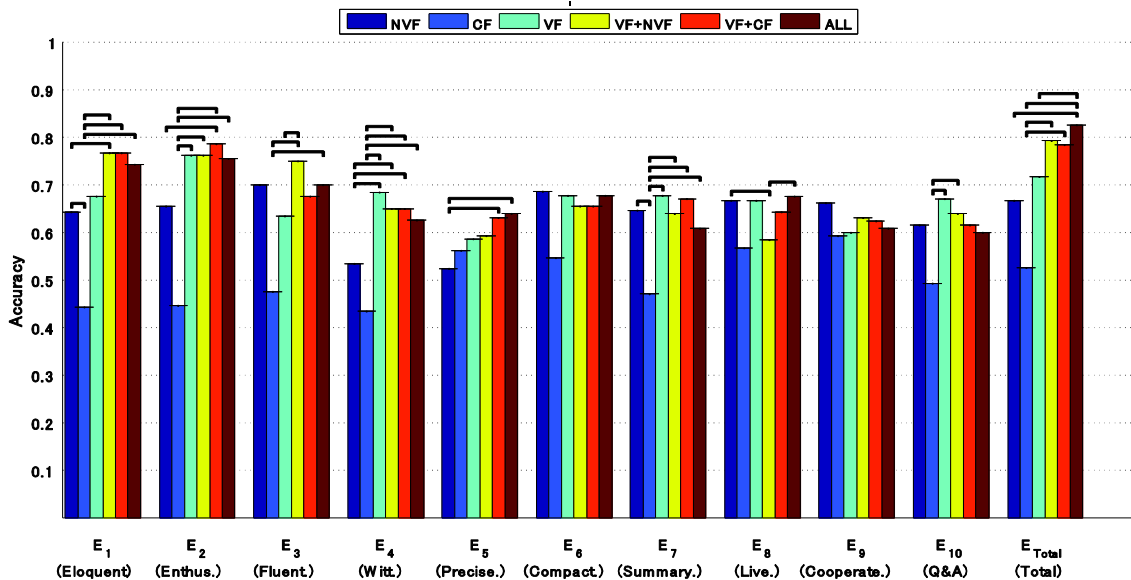


図1：説明評価値の予測精度（□は t-test により有意差が認められたペアを指す）

として10種類の変数を用意した。

- (1) "Eloquence" (雄弁さ),
  - (2) "Enthusiasm" (熱意),
  - (3) "Fluentness" (流暢さ),
  - (4) "Wittiness" (機知に富んだ説明),
  - (5) "Preciseness" (情報の正確さ),
  - (6) "Compactness" (端的な説明),
  - (7) "Summarization" (説明内容の要約),
  - (8) "Liveliness" (活気のある説明),
  - (9) "Cooperation" (協調的な説明),
  - (10) "Q&A" (質疑応答が適切に出来たか)
- 上記の10項目を10段階で評価した結果、3人の評価値の一致率(ICC(1, k))は0.71~0.85と十分な値が得られた。また「フィラー」、 「笑い」を含む書き起こしデータを収集し、アノテーションセットに加えた。

### (2) マルチモーダル特徴量セットの構築

書き起こしデータセットに対して、形態素解析を行い、各説明者の発話内容に関する特徴量を抽出した。また音声区間、韻律情報、ハンドジェスチャ、頭部ジェスチャ、顔向け方向を、種々の信号処理・機械学習の手法を用いてアノテーションした。発話内容、音声、韻律、頭部ジェスチャ、ハンドジェスチャ、顔向け方向といった特徴量を全て含むデータコーパスは今までに存在しないため、豊富な非言語情報を含むコーパスを作成することが出来たと考える。

また、マルチモーダルアノテーションラベル列に対して時間共起頻度の高い共起パターンを抽出する系列マイニング手法を、グラフクラスタリングを応用して開発した。この手法を用いて、「説明中に聞き手がうなづいた」といった説明シーンを特徴付けるコンテキスト特徴量を抽出した。

### (3) 説明状況下に表示する非言語情報の機能解析・モデル化

説明評価値と4.2の多様なマルチモーダル特徴量の間をピアソン積率相関分析・回帰分析・機械学習により関連付けることで、

説明評価値(説明スキル)に関連する特徴量を明らかにした。相関分析の結果、説明時間の長さ、説明者の声のピッチ、説明者の顔向け方向を変化させる回数、などが説明スキルと有意な正相関を持つことがわかった。一方で、聞き手の発話回数、説明者同士の相互注視などは有意な負相関を持つことがわかった。また、共起パターンとして「説明者が他の説明者を見ながら発話する」、「説明者がジェスチャを伴いながら説明を行う」、「説明者が話している間、聞き手がうなづいている」といったイベントが説明スキルと有意な正相関を持つことがわかった。

最後に、抽出した特徴量セットから段階的に特徴量(説明変数)を有意でない変数を除きながら回帰分析を行った結果、10変数の説明スキル評価値の合計(目的変数)を、決定係数0.72で説明できることを示した。最後に、高レベル・低レベルに説明スコアを分類し、10分割交差検定により説明スキルレベルの予測実験を行ったところ、学習したSVM(Support Vector Machine)モデルは82%の精度を得ることがわかった。各説明評価値の予測精度を図1に記載する。この結果は、説明会話中に会話者が表出する多様な言語・非言語特徴量を用いて、説明が上手い(上手くない)シーンを高精度で予測できることを示している。

### (4) 複数の会話場における非言語情報のセンシング環境構築

着座会話だけでなく、人が移動するパーティ会場や、イベント会場のような場所での会話における非言語情報をセンシングするための環境を複数台の深度センサを用いて構築した。3枚のポスタを配置し、各ポスタの前で各発表者が聞き手に説明を行い、10名以上の聴衆が、自由にそのポスタ間を行き来するポスタセッションを行い、各参加者の姿勢・位置を追跡する実験を行った。

まだ精度は不十分であるものの、各会話参加

者の役割 (ポスタ説明者, 聞き手, 某参与者, 通行人) を 7 割程度認識できることを確認した。

(5) 認知症患者の発話中の状態推定への応用成果 (3) のマルチモーダル情報の認識手法は会話エージェント・ロボットの会話相手ユーザの認識モジュールに応用可能である。

この研究では, 対話機能を備えたヴァーチャルエージェントと認知症患者 (ユーザ) の対話データを収集し, エージェントが質問を続けるべきか, ユーザの返答を待つべきかといった状態を, 音声区間・言いよどみ・視線変化・頭部ジェスチャから予測するためのモジュールを構築し, 評価した結果, 8 割の精度で適切な状態を認識できることを示した。

#### (6) 残った課題

外部から評価された説明スキル値を説明・予測するモデルを構築することで, 説明スキルに関わる言語・非言語情報を明らかにした。今後, 提案した説明スキル評価モデルを利用して, 説得力のある, 説明シーンのリトリバルシステムに応用する予定である。この新規の目標を達成するために, ビジネスシーンにおける説明シーン, 論理的な事象の説明といった, 他の題目に関する, 説明会話データを収集し説明スキル評価を行う予定である。また音声認識を利用した, 単語特徴量の自動抽出に関しても今後の課題である。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 10 件)

- [1] 岡田 将吾, 坊農 真弓, 高梨 克也, 角 康之, 新田 克己, 非言語マルチ モーダル情報を利用したグループ対話におけるジェスチャの機能認識, 電子情報通信学会論文誌, 査読有, Vol.J98-A, No.1 2015, pp.63-75 ,  
[http://search.ieice.org/bin/pdf.php?lang=J&year=2015&fname=j98-a\\_1\\_63&abs](http://search.ieice.org/bin/pdf.php?lang=J&year=2015&fname=j98-a_1_63&abs)
- [2] 片桐 恭弘, 石崎 雅人, 伝 康晴, 高梨 克也, 榎本 美香, 岡田 将吾, 対話コミュニケーションによる相互信頼感形成の共関心モデル, 「認知科学」22 巻 1 号 , 査読有, 2015, pp. 97-109 ,  
[http://www.jcss.gr.jp/publishing/journals/journal\\_backnumber/entry-143.html](http://www.jcss.gr.jp/publishing/journals/journal_backnumber/entry-143.html)
- [3] 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コーパス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, 査読有, , 2015, Vol,56,No.2, pp. 1217-1227  
[https://ipsj.ixsq.nii.ac.jp/ej/?action=page\\_s\\_view\\_main&active\\_action=repository\\_view\\_main\\_item\\_detail&item\\_id=141595&item\\_no=1&page\\_id=13&block\\_id=8](https://ipsj.ixsq.nii.ac.jp/ej/?action=page_s_view_main&active_action=repository_view_main_item_detail&item_id=141595&item_no=1&page_id=13&block_id=8)
- [4] 田村 理乃, 岡田 将吾, 西垣 貴央, 新田 克己, 力覚つきテキストコミュニケーションシステム“HAPPicom”の提案と評価, 人工知能学会論文誌, 査読有, 第 30 巻 2 号, 2015, pp.417-428  
[https://www.jstage.jst.go.jp/article/tjsai/30/2/30\\_30\\_417/\\_article/-char/ja/](https://www.jstage.jst.go.jp/article/tjsai/30/2/30_30_417/_article/-char/ja/)
- [5] Meng Chen, Shogo Okada, Katsumi Nitta, "Evaluation of Haptic Interaction in Intercultural Online Negotiation", Journal of Advanced Computational Intelligence and Intelligent Informatics, 査読有, Vol.17 No.6, 2013, pp.779-790 ,  
<http://www.fujipress.jp/finder/xslt.php?mode=present&inputfile=JACII001700060001.xml>
- [6] Naoko Saito, Shogo Okasa, Katsumi Nitta, Yukiko I. Nakano, and Yuki Hayashi, Estimating User's Attitude in Multimodal Conversational System for Elderly People with Dementia, In proceedings of AAAI Spring Symposium, 査読有, 2015, p100-103, <https://www.aaai.org/ocs/index.php/SSS/SSS15/paper/view/10274>
- [7] Makoto Sakahara, Shogo Okada, Katsumi Nitta: Domain-Independent Unsupervised Text Segmentation for

Data Management. In proceedings of ICDM Workshops, 査読有,2014,481-487,

- [8] Fumio Nihei, Yukiko I. Nakano, Yuki Hayashi, Hung-Hsuan Huang, **Shogo Okada**, Predicting Influential Statements in Group Discussions using Speech and Head Motion Information. In Proceedings of the 16th ACM on International conference on multimodal interaction (ICMI), 2014, 査読有, pp.136-143 , DOI: 10.1145/2663204.2663248

- [9] **Shogo Okada**, Mayumi Bono, Katsuya Takanashi, Yasuyuki Sumi, and Katsumi Nitta. Context-based conversational hand gesture classification in narrative interaction. In Proceedings of the 15th ACM on International conference on multimodal interaction (ICMI), 2013, 査読有, pp.303-310, DOI: 10.1145/2522848.2522898

- [10] Jing Xu, **Shogo Okada**, Katsumi Nitta: A Semantic-Similarity-Based Method for Object Description and Clustering. In Proceedings IEEE Systems Man and Cybernetics, 査読有, 2013 pp.3669-3676, DOI: 10.1109/SMC.2013.625

[学会発表] (計 10 件)

- [1] Naoko Saito, **Shogo Okasa**, Katsumi Nitta, Yukiko I. Nakano, and Yuki Hayashi, Estimating User' s Attitude in Multimodal Conversational System for Elderly People with Dementia, AAAI Spring Symposium,2015 年 3 月 25 日, CA (USA)

- [2] Makoto Sakahara, **Shogo Okada**, Katsumi Nitta: Domain-Independent Unsupervised Text Segmentation for Data Management. ICDM Workshops, 2014 年 12 月 14 日, Shenzhen (China)

- [3] Fumio Nihei, Yukiko I. Nakano, Yuki Hayashi, Hung-Hsuan Huang, **Shogo Okada**, Predicting Influential Statements in Group Discussions using Speech and Head Motion Information. 16th ACM on International conference on multimodal interaction (ICMI), 2014 年 11 月 13 日, Istanbul (Turkey)

- [4] 齋藤 直子 **岡田 将吾** 新田 克己 林 佑樹 中野 有紀子 音声対話エージェントを利用した認知症患者の状態把握支援の試み 第 27 回人工知能学会全国大会, 2014 年 5 月 13 日, ひめぎんホール (松山)

- [5] 米 航 **岡田 将吾** 新田 克己 会話中に表出する言語・非言語情報のデータマイニングに基づく説明会話の客観的評価指標の獲得 第 27 回人工知能学会全国大会, 2014 年 5 月 12 日, ひめぎんホール (松山)

- [6] 梶 航士、**岡田 将吾**、新田 克己: 屋内散策行動分析のための歩行者自律測位手法の提案と評価, 情報処理学会 第 70 回モバイルコンピューティングとユビキタス通信研究会 2014 年 3 月 14 日, 慶應大学 日吉キャンパス (横浜)

- [7] 小川 環 **岡田 将吾** 新田 克己: ポスタ発表における支援ロボットのための役割認識モデル, 信学技報 PRMU2013-129 CNR2013-37 pp.47-50, 2014 年 2 月 14 日, 福岡大学 (福岡)

- [8] **Shogo Okada**, Mayumi Bono, Katsuya Takanashi, Yasuyuki Sumi, and Katsumi Nitta. Context-based conversational hand gesture classification in narrative interaction. 15th ACM on International conference on multimodal interaction (ICMI), 2013 年 12 月 11 日, CA (USA)

- [9] Jing Xu, **Shogo Okada**, Katsumi Nitta: A Semantic-Similarity-Based Method for Object Description and Clustering. IEEE Systems Man and

Cybernetics,2013年10月13日  
Manchester (UK).

- [10] 岡田 将吾, 坊農 真弓, 高梨 克也, 角  
康之, 新田 克己, 非言語会話構造を利用した複数人対話における状況説明ジ  
ェスチャの分析・認識, 第68回 人工知  
能学会 言語・音声理解と対話処理研究  
会 (SIG-SLUD), 2013年9月19日,  
千葉大学 (千葉)

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

名称:

発明者:

権利者:

種類:

番号:

出願年月日:

国内外の別:

[その他]

ホームページ等

<http://www.ntt.dis.titech.ac.jp/~okada/index.html>

## 6. 研究組織

### (1)研究代表者

岡田 将吾 (OKADA SHOGO)

東京工業大学・大学院総合理工学研究科・  
助教

研究者番号: 00512261