

# 遠隔会議における円滑な議論を支援するための 興味推定モデルに関する検討

鳥山 英峻<sup>1,a)</sup> 石田 繁巳<sup>2</sup> 白石 陽<sup>2</sup>

**概要:** 遠隔会議では、他者が現在の議論へどれほどの興味を持ち合わせているかなどの意思疎通の難しさが、議論の長時間化や理解度・関与度の低下につながる要因となっている。このような遠隔会議特有の問題解決に向け、参加者が議論中にシステムを操作することで興味などを示させる研究が複数行われている。しかし、会話への注意の逸脱などが生じ、円滑な議論を停滞させる要因となるため主体的操作を必要としないものが望ましい。様々な分野において行動特徴を用いた興味推定に関する研究が複数存在することから、本研究では遠隔会議において、ユーザ操作を必要とすることなく参加者の行動特徴に基づいて意思疎通を支援するシステムを実現する。この実現に向けて本稿では、遠隔会議における参加者の議論内容に対する興味の推定に着目し、推定モデルに関する評価を行った。結果として、各参加者個人の興味推定モデルの F-measure が 5 割弱から 8 割強程度の範囲内、類似した行動特徴を持つ参加者同士のデータを用いたグループモデルは、5 割強から 7 割強程度の範囲内となり、事前に個人ごとの行動特徴と興味評価を多量に取らずとも、グループモデルを用いることで高精度に興味を推定可能である。

**キーワード:** 遠隔会議, 身体特徴, 発話特徴, 予測モデル

## 1. はじめに

近年、リモートワークやオンライン授業などの普及による遠隔会議システムの利用が急激に増加しており、今後も継続的に利用されていくことが予想される [1, 2].

遠隔会議では、相手の表情や仕草を汲み取りにくく、他の会議参加者の議論への興味度合いを把握することが難しい [3]. 他者が現在の議論へどれほどの興味を持ち合わせているかなど、意思疎通が難しいために議論の長時間化や参加者の理解度・関与度の低下につながっている [4, 5].

このような遠隔会議特有の問題解決に向けて、文献 [6–8] などの研究が複数報告されている。これらの研究は、議論中、会議参加者にシステムを操作させて興味などの意思を自ら示させるシステムを提案している。しかし、議論中のシステム操作は会話からの注意の逸脱など、円滑な議論を停滞させる要素となるため、主体的なシステム操作を必要としないものが望ましい。

一方、様々な分野において対象への興味の推定研究が複数行われている。文献 [9–11] では姿勢、頭部運動、表情などの身体動作特徴から、文献 [12, 13] では発話特徴から興味などを推定可能であることが報告されている。

本研究では、遠隔会議の議論においても興味度合いが身体動作や発話などの行動特徴に表れると考え、これらの特徴を利用し、円滑な会議のために意思疎通を支援するシステムを実現する。意思疎通を支援するため、システムは遠隔会議参加者の行動特徴に基づいて現在の議論内容への興味を推定し、他の会議参加者に共有・可視化を行う。また、会議参加者が支援を受けるにあたり、操作を不要にすることで議論の妨げにならないシステムを実現する。

著者らは、先行研究 [14] にて興味推定に有効と思われる行動特徴を考察して評価を行った。調査の結果、遠隔会議中の参加者の「顔の動き」「姿勢の変化」「感情の変化」「発話頻度」については議論内容への興味が特徴に表れやすく、推定に有効な特徴であることを確認した。しかしながら、興味を推定するためのモデル構築までは行っていない。

本稿では、意思疎通支援システムの実現に向けて、会議中の行動特徴を用いた興味推定モデルの構築および評価を行う。参加者個人のデータのみを学習させた個人モデルと、類似した行動特徴を持つ参加者同士のデータを用いて学習させたグループモデルのそれぞれの評価を行う。モデル構築のために特に重要な行動特徴、および興味推定におけるグループモデルの汎用性を明らかにする。

結果として、各参加者の個人モデルは変数重要度と推定精度の間に関連性は見られず、モデルごとで重要となる特

<sup>1</sup> 公立はこだて未来大学大学院 システム情報科学研究科

<sup>2</sup> 公立はこだて未来大学 システム情報科学部

a) g2122043@fun.ac.jp

徴が大きく異なることとなった。各個人モデルの評価結果として、推定精度は5割弱から8割強程度、各グループモデルの評価結果として、5割強から7割強程度の範囲内であった。事前に個人ごとの行動特徴と興味評価を多量に取らずとも、グループモデルに対して興味の推移時に表れる行動特徴の傾向が近い被験者であれば、高い精度で興味を推定可能である。

本稿の構成は以下の通りである。2.では関連研究を基に本研究の位置づけを述べ、3.で提案する手法の詳細を述べる。そして、4.で評価について述べ、5.でグループモデルの改良を行う。6.で意思疎通支援システムの実現のため、議論を行い、7.でまとめとする。

## 2. 関連研究

### 2.1 対象への興味推定に関する研究

行動特徴から対象への興味を推定する研究は、複数報告されている [9–12]。文献 [9] では、実店舗向け購買支援システムを提案するために、深度センサを用いた頭部分析とアイトラッカからユーザの興味を推定している。文献 [10] では、Web上の記事を読んでいるときのユーザの記事に対する興味度合いを推定したところ、注視時間の比率と興味に相関があることを報告している。文献 [11] では、公共ディスプレイに対する興味度を、利用者の位置や体の向きから推定するアルゴリズムを報告している。文献 [12] では、発話頻度に注目しており、議論内の重要単語が切り替わるまでの個人の発話総数は、議論への満足度に対して影響があると述べている。

これらの研究では、対象への興味を推定するために、様々な身体動作特徴や発話特徴を用いている。著者らの調べた範囲では、遠隔会議における議論内容に対する興味推定においては、どの行動特徴が推定に有効か、報告された研究は確認していない。

### 2.2 遠隔会議の円滑化に関する研究

遠隔会議において意思疎通の補助に着目した研究 [6–8] が報告されている。文献 [6] では、消極的参加者に発言を促すために、会議と併用可能な発言リクエスト機能を有するコミュニケーションシステムを提案している。文献 [7] では、ボタンで簡単に意思を示すことが可能なシステムを提案している。文献 [8] では、ジェスチャー表現を用いた意思疎通システムを提案している。これらの研究のような会議参加者に主体的な操作を求めるシステムは「会話の注意を逸らされる」ことや、「自ら積極的に使用しにくい」などの意見が被験者により述べられていることが多くある。

会議円滑化のためのフィードバック研究 [15–17] が報告されている。文献 [15] では、遠隔会議の情報を分析して会議の状況や行動情報など様々な情報をフィードバックするシステムを提案している。文献 [16] では、暗示的な発話に

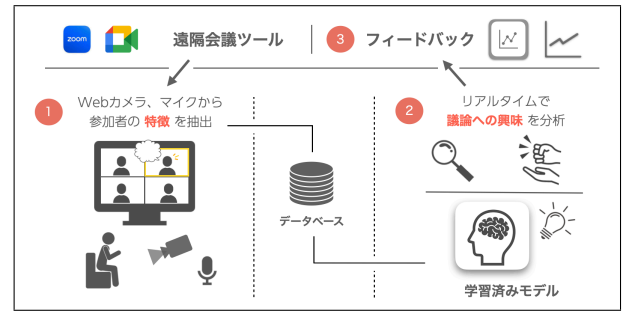


図 1: 意思疎通支援システムの概要図

対する提示方法で、被験者が主体的に発話の流れを促進するシステムを提案している。文献 [17] では、発話などの会議の関与度合いを花のメタファーに変換するフィードバック手法を提案している。これらの研究は、会議参加者に何らかの行動を促すシステムではあるが、興味度合いの推定とフィードバックは行われていない。フィードバックするためのUIに関しては検討の余地があるが、本稿では評価せず、議論への興味推定モデルの構築と評価を行う。

## 3. 提案手法

### 3.1 概要

図 1 に、意思疎通支援システムの概要を示す。意思疎通支援システムは、特徴量の抽出、興味の推定、フィードバックの3つのステップで構成される。次節以降では、各ステップについて詳細に述べる。

### 3.2 特徴量の抽出

特徴量の抽出では、会議参加者の身体の動きや声などといった行動特徴を抽出する。具体的に、先行研究 [14] で調査し、興味推定に有効であると示された参加者の「顔の動き」「姿勢の変化」「感情の変化」「発話頻度」を中心に抽出する。

議論中の興味推定を目的とした、会議参加者の行動特徴の抽出は2.2で述べた主体的な操作や、特別な外部デバイスの使用を避けるためWebカメラとマイクのみを用いて行う。特徴を抽出可能なフレームワークを表 1、具体的な抽出内容については表 2 に記述する。

顔の角度については Mediapipe Facemesh [18] を用いることで会議中の顔の角度を抽出する。432個の顔特徴から線形ベクトルを作成し、映像内の会議参加者から見て、左から右の水平方向 (rad\_x)、下から上の垂直方向 (rad\_y)、手前から奥の前後方向 (rad\_z) の角度に換算する。感情については Amazon Rekognition [19] を用いて、顔の特徴点を感情に変換し推定する。Amazon Rekognition が提供する API は顔の特徴点から ANGRY (怒り), CALM (平静), DISGUSTED (嫌悪), FEAR (恐怖), HAPPY (喜び), SAD (悲しみ), SURPRISED (驚き), CONFUSED (困惑) の8つの感情に分類する。姿勢は OpenPose [20]

表 1: 特徴抽出に使用したフレームワーク

取得する特徴	使用するフレームワーク
1. 顔の角度	Mediapipe Facemesh
2. 姿勢	OpenPose
3. 感情	Amazon Rekognition
4. 発話頻度	OpenAI Whisper

表 2: 推定モデル構築に用いる身体特徴

身体特徴	特徴名
顔の角度	水平方向 (rad_x), 垂直方向 (rad_y) 前後方向 (rad_z)
感情特徴	ANGRY (怒り), CALM (平静) DISGUSTED (嫌悪), FEAR (恐怖) HAPPY (喜び), SAD (悲しみ) SURPRISED (驚き), CONFUSED (困惑)
姿勢特徴	胸部-x, 胸部-y, 右肩-x, 右肩-y, 右肘-x, 右肘-y 左肩-x, 左肩-y, 左肘-x, 左肘-y, 右目-x, 右目-y 左目-x, 左目-y, 右耳-x, 右耳-y, 左耳-x, 左耳-y

を用いて推定する。映像から見て、胸部、首、両目、両耳、両肩、両肘の特徴を、x 軸、y 軸の座標位置で抽出する。発話頻度に関して、1 発話の開始から終了までの判断を OpenAI Whisper [21] を用いて推定する。

行われる遠隔会議において、各行動特徴は時系列データとしてタイムスタンプを持たせながら、定期的に取得する。具体的に、会議開始時から 2.5 秒間隔で顔の角度特徴、姿勢特徴、感情特徴を取得する。発話頻度に関しては、過去 2.5 秒間の間に発話しているかどうかを手動でラベリングし、特徴として活用する。

### 3.3 興味の推定

3.2 で述べた特徴量を用いて、教師あり機械学習により、会議参加者の興味推定を行う機械学習モデルを構築する。議論中の興味推定を目的とした、学習モデルは興味度合いに対して 4 段階で推定を行う。4 段階の理由としては 3.4 で後述する。興味推定モデルは、参加者個人のデータのみを学習させた個人モデルと、各参加者データを組み合わせで学習させたグループモデルを構築する。参加者個人のデータのみを学習させることで、個人でしか現れない行動特徴も学習することが可能であり、高い精度で興味を推定可能であると考え。この個人モデルとは別に、様々な会議参加者のデータを学習させたグループモデルを構築することで、事前に個人の特徴を学習することなく、興味を推定することが可能であると考え。

本稿では、特徴量の重要度を評価することが可能である Random Forest を用いる。興味の推移時に表れる行動特徴は各個人で異なると考え、どのような行動特徴に変化が表れるのかについて重要度を確認することで、分析することが可能である。また、様々な個人に対応できるよう次元数の大きい特徴量を用いるため、Random Forest を用い

て過学習を防ぐ。なお、本提案手法では使用する機械学習アルゴリズムは限定せず、行動特徴を用いて「興味がある」「やや興味がある」「やや興味がない」「興味がない」のような 4 クラスの分類を可能な教師あり学習アルゴリズムであればどのようなものでも適用可能である。

学習に用いる正解データは、会議参加者が自ら議論に対する興味度合いを判断し、手動で興味度合いを入力させたものを用いる。詳細な手順としては、4.1 で後述する。

### 3.4 フィードバック

3.3 で求めた興味推定結果を用いて、データの整形を行い、他の会議参加者にフィードバックを行う。フィードバックについては各個人の興味推定結果を集計し、平均化したものを表示することで匿名性を保つ。表示方法として、2.2 で述べたように、注意の逸脱となるため主体的操作を必要としないものが望ましい。また、フィードバックに含まれる情報量が多くなると、会議参加者の注意を逸らす要因となることが考えられる。これらの理由から、会議の時系列に沿った興味の推移を折れ線グラフとして、4 段階の簡易的なフィードバックを行う。また、本稿ではフィードバックシステムの評価は行わない。

## 4. 評価

本章では、興味推定モデル構築のために特に重要な行動特徴、および興味推定におけるグループモデルの汎用性を評価する。4.1 では行動特徴を抽出する実験を行う上での条件について述べ、4.2 では実験結果について述べる。4.3 ではモデルの評価環境を述べ、4.4 でモデルの評価を行い、4.5 で考察する。

### 4.1 実験条件

被験者は 20 代の 4 名を 1 グループとして、計 12 名の 3 グループに Web 会議ツールである Zoom を使用させ、議論を行わせた。各グループ進行役となるファシリテータ、議論内容をまとめる書記を設け、ビデオカメラは off で議論を行わせた。グループ A は「友達作りやすくて、若者の普及率が高くなるようなアプリアイデア」、グループ B、C は「若者の投票率を上げるためにはどのような政策が有効か」というテーマを設定し議論させた。議論の補助ツールとして Web ホワイトボードアプリケーションの miro を使用させた (図 2)。

議論時間に関して、予備実験時に発散・収束合わせて 15 分の議論時間を設けて議論させたが、実験設定の議論時間が短いため集中力が途切れず特徴が大きく変化しなかった。このことから、意見発散フェーズと収束フェーズに分けて 20 分、10 分の計 30 分の議論を行わせた。

行動特徴と興味の関連性を調査するため、議論を行わせたあとに会議の録画を閲覧させながら、議論中の自身がど

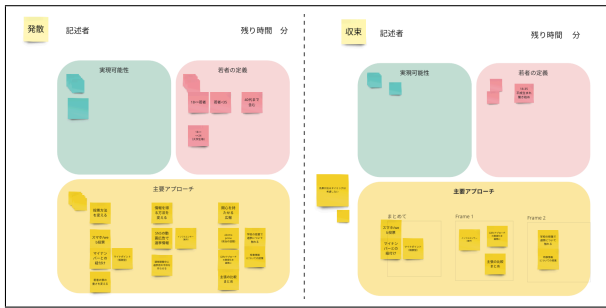


図 2: Web ホワイトボード miro 使用時の様子

表 3: 興味評価ラベル総数

興味段階	各ラベル数
興味がある (ラベル 3, 4)	597
やや興味がある (ラベル 2, 1)	603
やや興味がない (ラベル -1, -2)	126
興味がない (ラベル -3, -4)	83
計	1409

のような興味状態だったか、自らラベリングを行わせた。興味評価は -4 から +4 までの 8 段階で、会議の録画すべてに対して Google Sheets を用いて 5 秒刻みに判断させた。

## 4.2 実験結果

被験者 12 名のうち、会議中安定して行動特徴を取得できた 9 名が回答した各興味評価ラベルの総数を表 3 に示す。評価はフィードバックすることを想定し、ラベル 3, 4 の回答を「興味がある」、1, 2 を「やや興味がある」、-2, -1 を「やや興味を持ってない」、-4, -3 を「興味を持ってない」の 4 分類にした。結果として、「興味がある」と回答された回数は 597 回、「やや興味がある」と回答された回数は 603 回、「やや興味がない」と回答された回数は 126 回、「興味がない」と回答された回数は 83 回となった。

一部ピックアップした行動特徴と興味評価の相関係数を表 4 に示す。相関係数はピアソンの積率相関係数を計算した。また、被験者の姿勢などの要因で取得できなかった特徴は「-」で記載している。各被験者ごとのデータ数として、議論中に取得した顔の角度特徴、姿勢特徴が 1 次元あたり 4,345 件、感情特徴が 1 次元あたり 867 件となり、正解データとなる興味評価ラベルは 360 件となった。

## 4.3 評価環境

各行動特徴に関して、興味評価ラベルに紐付けるよう、5 秒ごとにダウンサンプリングを行った。ダウンサンプリングは 5 秒区間に取得した特徴の最大値に合わせるように行った。評価指標として、5 分割交差検証を用いて、推定精度、および F-measure を重みつき平均で算出した。また、各特徴に関する次元数の削減、ツリーの深度、ノード数、決定木数は F-measure を最大化するようにモデルごとにグリッドサーチにて決定した。

また、本稿の議論形態として前半議論、後半議論ともに各グループに 2 人ずつファシリテータと書記の役割が存在するため、これを特徴として学習させた。

## 4.4 評価結果

構築した各被験者ごとの興味推定モデルの精度、および変数重要度が高い特徴の 3 つを表 5 に示す。結果として、推定精度、F-measure とともに 5 割弱から 8 割強程度の範囲内となった。被験者 F は、最も推定精度が高く 0.855、F-measure が 0.728、次に推定精度が高い被験者 A は推定精度が 0.845、F-measure が 0.754 となった。一方、最も推定精度が低い被験者 K は 0.517、F-measure は 0.423 であり、次に推定精度が低い被験者 M は 0.641、F-measure は 0.519 となった。また、被験者全員のデータを学習させたグループモデルの推定精度は 0.688、F-measure は 0.657 となった。

各個人モデルにおいて重要度が高い行動特徴として、顔の動きに関しては、水平方向がもっとも重要度が高い結果となった。顔の奥行き、垂直方向に関しては被験者ごとにばらつきが見られた。次に感情特徴に関しては、喜び、怒りの 2 つが重要度が平均的に高く、平静や恐怖、驚きに関しては被験者ごとにばらつきが見られた。姿勢特徴に関しては目と肩の y 座標が特に高い重要性を持つ結果となった。発話頻度、議論内の役割に関する特徴は、各被験者とも身体特徴と比較して重要度が低い結果となった。一方、被験者全員のデータを含めたグループモデルに関しては、感情特徴以外は重要度が低い結果となった。

## 4.5 考察

表 5 に示したように被験者ごとに推定精度が大きく変化する結果となった。この一要因として、各被験者の興味評価の付け方が大きく異なっていたことが挙げられる。各被験者が実験後に評価した興味ラベルの平均値と分散度合いを表 6 に示す。平均値の違いに関する推定精度の変化は見られない一方で、興味評価の分散度合いが大きいほど推定精度が低くなる傾向が見られる。各被験者の興味評価の分散値を比較すると、推定精度が最も高かった被験者 F が 0.85、最も低かった被験者 K は 3.77 と大きな差が生まれている。また、次に推定精度が高い被験者 H と、次に推定精度が低い被験者 M の分散値は、それぞれ 0.94, 3.38 と大きな差が生まれている。

図 3 に被験者 F と被験者 K の興味推定モデルの混同行列を示す。各被験者を比較すると、被験者 F は、行った会議に対して「興味がない (ラベル -4, -3)」をつけていないことから分類先が 1 クラス少ない。また、「興味がある (ラベル 3, 4)」の評価数が極端に多いことがわかる。一方で被験者 K は、「興味がない (ラベル -4, -3)」は少ないものの、ある程度均等に評価をつけていることがわかる。

表 4: 行動特徴と興味評価の相関関係

特徴名	被験者 A	被験者 F	被験者 G	被験者 H	被験者 I	被験者 J	被験者 K	被験者 L	被験者 M
水平方向 (rad_x)	0.240	-0.020	-0.188	-0.219	-0.144	-0.014	0.000	0.184	0.111
前後方向 (rad_z)	0.318	-0.004	0.221	0.286	0.022	0.068	0.109	0.089	0.030
ANGRY (怒り)	-0.126	0.141	0.308	0.005	-0.095	0.179	-0.016	-0.052	-0.154
CALM (平静)	-0.429	-0.053	-0.198	-0.124	0.174	0.054	-0.179	-0.057	-0.199
CONFUSED (困惑)	0.269	-0.050	0.049	0.117	-0.028	-0.160	0.032	0.090	-0.107
HAPPY (喜び)	0.348	0.107	0.251	0.191	0.009	0.094	0.203	-0.123	0.251
SURPRISED (驚き)	0.158	0.123	0.136	0.063	-0.112	-0.044	-0.004	-0.149	0.013
右肩_x	-0.013	-0.022	-0.029	-0.169	0.082	-0.014	-0.014	0.033	0.086
右肘_x	-	-0.016	-0.121	0.138	-	0.046	-0.026	0.079	-0.127
右目_x	0.225	0.038	-0.168	-0.194	-0.104	-0.010	-0.002	0.159	0.102
左目_x	0.224	0.064	-0.132	-0.190	-0.106	-	0.019	0.212	0.154
右耳_x	0.168	0.031	-0.014	-0.235	-0.075	0.011	0.014	0.072	0.050

表 5: 各被験者ごとの興味推定モデルの分類精度 (4 分類)

条件	被験者 A	被験者 F	被験者 G	被験者 H	被験者 I	被験者 J	被験者 K	被験者 L	被験者 M	被験者全員
推定精度	0.845	0.855	0.746	0.835	0.687	0.725	0.517	0.647	0.641	0.688
F-measure	0.754	0.728	0.672	0.800	0.606	0.584	0.423	0.556	0.519	0.657
重要度が 高い特徴	喜び	驚き	胸部 y	喜び	恐怖	怒り	顔の水平	左肩 y	平静	困惑
	平静	左肩 x	右肩 y	左目 y	顔の水平	喜び	恐怖	胸部 y	喜び	平静
	顔の水平	左肩 y	怒り	顔の垂直	怒り	平静	顔の前後	左目 y	顔の前後	怒り

このことから、各被験者ごとに会議に感じる興味度合いに、個人差があることがわかる。

また、各被験者の変数重要度と推定精度に関しては関連性は見られず、個人ごとに重要となる特徴が大きく異なることとなった。最も推定精度が高かった被験者 F と、次に高い被験者 H の変数重要度の上位 3 件を比較すると共通となる特徴は見られていない。実際、精度が低かった被験者 K や被験者 M は被験者 A と共通した重要性の強い特徴が見られるが、精度には大きく差がある。このことから各被験者ごとに推定精度に関わらず、興味に対して表れる行動特徴が大きく異なることがわかる。

全会議参加者の行動特徴を学習させたグループモデルはほとんどの個人モデルと比較して推定精度が低い結果となった。また、変数重要度に関しては上位 6 件が感情特徴という結果となった。この結果は、各個人の興味の推移により表れる個人特有の動きや癖がうまくモデルに反映されておらず、推定に汎用的である感情特徴のみが重要度が高くなったと考えられる。このことから、グループモデルにおいても推定精度を高めるためには、ある程度行動特徴が似た被験者でグループモデルを構築する必要がある。

## 5. 改善手法

本章では、4.5 で述べた考察を踏まえ、興味の推移時に表れる行動特徴が似た被験者同士でグループモデルを構築し、推定精度を評価する。5.1 でグループモデル改善のための考察を行い、5.2 でモデルを構築、5.3 で評価結果、5.4 で考察を述べる。

### 5.1 モデル改善考察

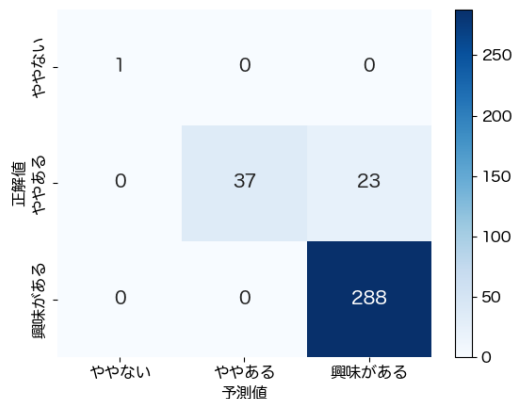
本節では、表 4 と表 5 に示した結果から、グループモデルの構築に必要な要素を考察する。

顔の動きは変数重要度が高い被験者が一定数存在するため、予測モデル構築に有効と考えられる。顔の水平方向の動きは、図 4 に示すようにカメラに対して正面を向いていることを指す 0.6 あたりを中心に、左右に大きな分散が見られた。これは、被験者ごとにモニターの位置が違ふことなどから生じたものも含まれると考えられる。また、OpenPose を用いた姿勢分析の結果からも同じことが述べられる。図 5 に示すように、全被験者において肩や胸などの x 軸に関する特徴は、分散されていた顔の水平方向の特徴と正の相関にあることがわかる。このことから、全被験者のデータを学習させたグループモデルは、被験者ごとの議論環境によって表れる顔と体が平均的に左右どちらかを向いているか、どちらの向きで興味が出やすいかなどの特徴を打ち消していた可能性がある。よって、顔の動きの水平方向と姿勢特徴を用いる場合は、興味評価の相関の正負が同じ被験者同士を集めてモデルを構築する必要がある。

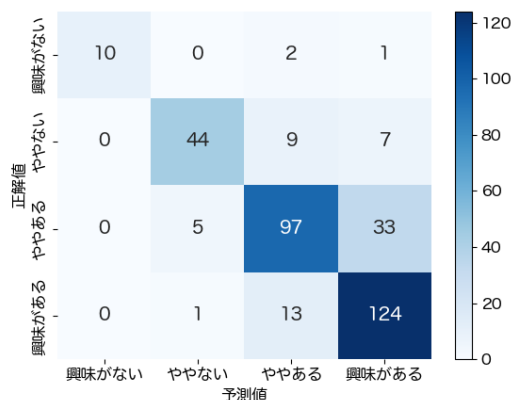
感情特徴は、変数重要度が高い被験者が一定数存在するため、予測モデル構築に有効と考えられる。しかしながら、表 4 に示したとおり、個人ごとに興味と相関のある感情にばらつきが見られた。具体的に、被験者 J は怒りの感情が高くなると興味が強くなる結果となっているが、被験者 A はその逆の結果となっている。このことから感情特徴を用いる場合も、各感情の相関係数が近い被験者同士を集めてモデルを構築する必要がある。

表 6: 各被験者における興味ラベルの分散度合いと平均値

	被験者 A	被験者 F	被験者 G	被験者 H	被験者 I	被験者 J	被験者 K	被験者 L	被験者 M
平均	1.74	3.45	1.52	1.07	2.42	3.09	1.57	1.21	0.84
順位	4	1	6	8	3	2	5	7	9
分散	0.89	0.85	1.76	0.94	2.08	1.12	3.77	3.56	3.38
順位	8	9	5	7	4	6	1	2	3



(a) 被験者 F



(b) 被験者 K

図 3: 興味推定モデルの混同行列

4.5 で考察した際、他のラベルと比較して「興味がある (ラベル 3, 4)」を多くつける被験者がいるように、被験者ごとに会議に感じる興味度合いに、個人差があることがわかる。これを踏まえ、各被験者の会議に対する興味の抱き方が近い被験者同士のグループを作成する。

## 5.2 モデル構築

推定モデルは、類似した行動特徴を持つ参加者同士でグループモデルを構築する。類似した行動特徴に関して、5.1 で述べた特徴を考慮して、構築する。具体的には、顔の角度・姿勢と興味評価の相関が正の被験者を集めたモデル、および負の被験者を集めたモデル、興味を抱いた際に表れる感情の変化が近い被験者同士を集めたグループモデル

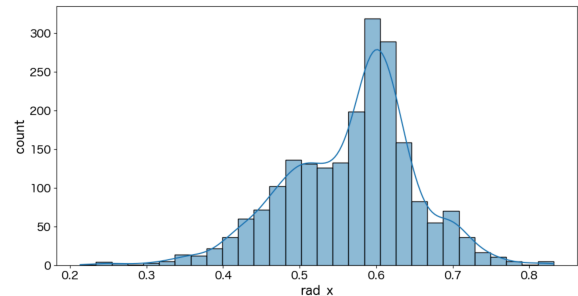


図 4: 顔の水平方向 (rad\_x) の分散度合いグラフ

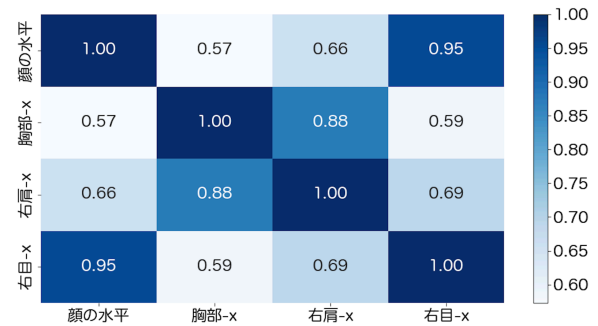


図 5: 顔の水平方向と姿勢特徴の相関係数

である。このほかに各被験者の会議に対する興味の抱き方が近い被験者同士のグループを作成する。具体的には 8 段階ある評価ラベルの分散値を参考にして、全被験者中、分散値が比較的高いグループと低いグループに 2 分割する。

本稿では、顔の角度・姿勢に関して正の相関がある 3 名、負の相関がある 6 名、感情に関して平静、怒り、喜びの各感情をそれぞれ代表値として持つ 3 名、分散値が高い 5 名と低い 4 名のデータを用いてモデルを構築した。また、モデルの評価環境に関しては、4.3 と同様である。

## 5.3 モデル評価

構築した各条件におけるグループごとの興味推定モデルの精度を表 7 に示す。表の括弧内の数字は、各グループに含まれる被験者の人数を表す。結果として、興味の推定精度は 5 割強から 7 割強程度の範囲内となった。中でも精度がもっとも高かったのは興味評価分散値が低いグループモデルであり、推定精度が 0.784、F-measure が 0.783 となった。次に精度が高かったのは怒りと興味評価の相関係数が高い被験者を集めたグループモデルであり、推定精度が 0.755、F-measure が 0.734 となった。推定精度、F-measure が共に最も低かったのは興味評価の分散値

表 7: 各グループごとの興味推定モデルの分類精度 (4 分類)

条件 (被験者数)	被験者全員	顔の角度・姿勢		最も相関係数が高い感情ラベル			興味評価分散値	
	(9)	正の相関 (3)	負の相関 (6)	平静 (3)	怒り (3)	喜び (3)	高い (5)	低い (4)
推定精度	0.688	0.673	0.702	0.754	0.755	0.648	0.589	0.784
F-measure	0.657	0.611	0.692	0.716	0.734	0.613	0.596	0.783

が高い被験者を集めたグループモデルであり、推定精度が 0.589, F-measure が 0.596 となった。

#### 5.4 考察

表 7 に示したように、興味評価との相関係数を考慮して構築したグループモデルの精度は、被験者全員のデータを含めたモデルの精度と比較してある程度高い結果となった。具体的には、顔の角度・姿勢特徴と興味評価ラベルに負の相関があるグループ、平静、怒りの感情と興味評価ラベルの相関が高いグループである。これは、単に学習データ件数を増やすだけでなく、興味の推移時に表れる行動特徴の種類が近い被験者同士のデータを集めたグループモデルを構築することは推定に有効であることを示している。そのため、事前に個人ごとの行動特徴と興味評価を多量に取らずとも、グループモデルに対して興味の推移時に表れる行動特徴の傾向が近い被験者であれば、高い精度で興味を推定可能であると考えられる。

一方で、興味評価の分散値の高低で推定精度が変化してしまう問題は残る。分散値が高い被験者を集めたグループモデルは、分散値が低い被験者を集めたグループモデルと比較して推定精度に大きく差が出てしまう結果となった。これは、4.5 でも述べたように、興味評価の分散値が高い被験者はある程度均等に興味評価をつけている傾向があるために生じたものと考えられる。

### 6. 議論

意思疎通支援システムを実現するため、今後検討すべき点が 2 つある。

1 つ目に、興味推定を行うためのモデル構築に向けた、興味評価に対する分散値の扱いである。本稿の実験によって、会議への興味の抱き方は個人によって大きく異なることがわかった。この違いは個人モデルの構築において、一種の特徴として反映することができる。しかし、グループモデルとして扱う際は個人の興味評価の基準値を一定にする必要がある。

この問題の解決法として、各被験者が評価した全興味ラベルの中央値や平均値を活用する方法がある。例として、被験者 F の全興味ラベルの平均値は 3.45 となっているが、この値を興味評価の基準値として、各評価ラベルより上か下かを新たな興味ラベルとすることで興味評価分散値の振り幅を一定にすることが可能となる。各被験者における興味評価の分散値を一定にすることで、推定精度の極端な差

を抑えることができるためグループモデル構築に最適であると考えられる。一方で、この方法は被験者の会議への興味傾向という一種の特徴を反映することができないデメリットが存在しているため、これを考慮する必要がある。

2 つ目に、興味度合いをフィードバックするための最適な UI の検討である。参加者映像に画像処理を用いた研究 [22, 23] や発話支援に着目した研究 [24, 25] が報告されている。これらの研究は主体的なシステム操作が不必要なように、参加者映像に画像処理を用いたり、発話タイミングを自動支援するシステムを提案しているが、支援されるタイミングに不満を抱く被験者も多い。このことから、本システムが適切なタイミングで参加者を支援することは重要な要素であることが考えられる。

また、議論の円滑化にはどのぐらいの興味推定精度が必要かという点に関しては、フィードバック方法で大きく異なる。現在想定しているフィードバックは、参加者の興味度合いをグラフの上下推移で表示することであり、会議参加者は過去の興味も含めて時系列的に確認することができる。そのため、興味推定精度が 7, 8 割程度で誤推定が生じたとしても常に連続的な推定が行われるため、他者の興味を判別するのに問題はないと考える。このような、興味推定精度と適切なフィードバックの関係性を今後検討し、評価する必要がある。

### 7. おわりに

本稿では、遠隔会議における意思疎通を支援するため、会議参加者の行動特徴から興味を推定する手法を提案した。先行研究 [14] から、興味推定に有効である行動特徴を抽出し、実際の議論から得られたデータを用いて、参加者個人のデータのみを学習させた個人モデルと、類似した行動特徴を持つ参加者同士のデータを用いて学習させたグループモデルを構築した。

結果として、各モデルとも変数重要度と推定精度の間に関連性は見られず、モデルごとで重要となる特徴が大きく異なることとなった。個人モデルの評価結果として、興味の推定精度、F-measure ともに 5 割弱から 8 割強程度の範囲内となった。中でも被験者 F は、最も推定精度が高く 0.855, F-measure が 0.728 という結果となった。各グループモデルの評価結果として、5 割強から 7 割強程度の範囲内となった。中でも精度が高かったのは怒りと興味評価の相関係数が高い被験者を集めたグループモデルであり、推定精度が 0.755, F-measure が 0.734 となった。事

前に個人ごとの行動特徴と興味評価を多量に取らずとも、グループモデルに対して興味の推移時に表れる行動特徴の傾向が近い被験者であれば、高い精度で興味を推定可能であると考えられる。

今後の課題として、2つのことが挙げられる。興味推定を行うためのモデル構築に向けた興味評価に対する分散値の扱い、興味度合いの最適なフィードバックについての検討である。本稿の実験によって、会議への興味の抱き方は個人によって大きく異なり、これを一種の特徴として反映できることを示したが、グループモデルとして扱う際は個人ごとの興味評価の基準値を一定にする必要がある。同じく、興味推定精度と適切なフィードバックの関係性を今後検討し、様々な会議目的を持った議論環境で、評価する必要がある。

### 参考文献

- [1] 総務省：新型コロナウイルス感染症が社会にもたらす影響:5Gが促すデジタル変革と新たな日常の構築, 令和2年版 情報通信白書 第1部, pp. 138–166 (2020).
- [2] 総務省：令和3年通信利用動向調査の結果, 情報通信統計データベース, pp. 1–27 (2021).
- [3] 宮内佑実, 遠藤正之: オンライン会議とオフライン会議の意思疎通の比較, 経営情報学会全国研究発表大会要旨集, pp. 144–147 (2020).
- [4] Yang, L. and et al.: The Effects of Remote Work on Collaboration among Information Workers, *Nature Human Behaviour* 6, No. 1, pp. 43–54 (2021).
- [5] Cao, H., Lee, C., Iqbal, S., Czerwinski, M., Wong, P., Rintel, S., Hecht, B., Teevan, J. and Yang, L.: Large Scale Analysis of Multitasking Behavior During Remote Meetings, *Proc. the CHI Conference on Human Factors in Computing Systems*, pp. 1–13 (2021).
- [6] 塩津翠彩, 高島健太郎, 西本一志: 消極的参加者に発言を促す手段を備えたチャット併用会議用コミュニケーションメディア, 情報処理学会研究報告, Vol. 2018-GN-104, pp. 1–8 (2018).
- [7] 阿部花南, 築館多藍, 桑宮 陽, 小林 稔: 会議円滑化支援を目的とした気持ち可視化ボタンの提案, 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), pp. 774–783 (2021).
- [8] 井出将弘, 大島昇時, 森 真吾, 市野順子, 田野俊一: VR会議におけるアバターを介したジェスチャー表現の影響, 情報処理学会研究報告, Vol. 2020-GN-110, No. 1, pp. 1–8 (2020).
- [9] 本松大夢, 徳永大空, 杉原賢次, 尾崎保乃花, 安田真理, 中村優吾, 荒川 豊: ユーザの商品に対する興味推定に基づく補足情報提示の受容性検証, 情報処理学会研究報告 行動変容と社会システム, Vol. 9, pp. 1–9 (2023).
- [10] 加藤勇太, 岩本健嗣, 松本三千人: タッチ操作ログを用いた Web コンテンツ閲覧時における興味度合い推定の研究, 情報処理学会論文誌, Vol. 59, No. 2, pp. 508–518 (2018).
- [11] Schiavo, G., Mencarini, E., K., V., B. and Zancanaro, M.: Sensing and reacting to users' interest: an adaptive public display, *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pp. 1545–1550 (2013).
- [12] 西條涼平, 徳永陽子, 山口大地, リドウィナアンダリニ, 松尾翔平, 戸嶋巖樹, 倉橋孝雄, 小澤史朗: 感性コミュニケーション技術の実現に向けた熟練度と対話満足度に関する取り組み, *NTT 技術ジャーナル*, Vol. 34, No. 1, pp. 32–35 (2022).
- [13] 富増紗也華, 荒木雅弘: 雑談対話におけるマルチモーダル情報からの興味の有無の判定, 人工知能学会全国大会論文集 JSAI, pp. 1–4 (2016).
- [14] 鳥山英峻, 石田繁巳, 白石 陽: 遠隔会議における円滑な議論を支援するための興味推定モデルに関する検討, 情報処理学会研究報告, Vol. 2023-MBL-107, No. 45, pp. 1–8 (2023).
- [15] Samiharoase, S., McDuff, D., Sim, R., Suh, J., Rowan, K., Hernandez, J., Rintel, S., Moynihan, K. and Czerwinski, M.: MeetingCoach: An Intelligent Dashboard for Supporting Effective & Inclusive Meetings, *CHI'21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–13 (2021).
- [16] Schiavo, G., Cappelletti, A., Mencarini, E., Stock, O. and Zancanaro, M.: Overt or subtle? Supporting group conversations with automatically targeted directives, *IUI '14: Proceedings of the 19th international conference on Intelligent User Interfaces*, pp. 225–234 (2014).
- [17] 山田楓也, 若園裕太, 鉢呂悠真, 石田繁巳, 白石 陽: フラワーガーデンフィードバックによる遠隔会議支援手法の提案, 情報処理学会 マルチメディア通信と分散処理ワークショップ論文集, Vol. 30, pp. 203–205 (2022).
- [18] Kartynnik, Y., Ablavatski, A., Grishchenko, I. and Grundmann, M.: Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs, *The Workshop on Computer Vision for Augmented and Virtual Reality in conjunction with Conference on Computer Vision and Pattern Recognition*, pp. 953–959 (2019).
- [19] Amazon Web Services, I.: Amazon Rekognition, Amazon.com. (online), available from (<https://aws.amazon.com/jp/rekognition/>) (accessed 2023/10/23).
- [20] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E. and Sheikh, Y.: OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, *arXiv* <https://doi.org/10.48550/arXiv.1812.08008>, pp. 1–14 (2019).
- [21] Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C. and Sutskever, I.: Robust Speech Recognition via Large-Scale Weak Supervision, *arXiv* <https://arxiv.org/abs/2212.04356>, pp. 1–28 (2022).
- [22] Ziting, G., 金井秀明: 遠隔会議における相手の顔を変換することによる緊張緩和手法に関する研究, 情報処理学会研究報告, Vol. 2021-GN-113, No. 3, pp. 1–8 (2021).
- [23] 鈴木啓太, 横山正典, 吉田成朗, 望月崇由, 布引純史, 鳴海拓志, 谷川智洋, 廣瀬通孝: 同調的な表情変形技術を用いた遠隔コミュニケーションの拡張, 情報処理学会論文誌, Vol. 59, No. 1, pp. 52–60 (2017).
- [24] Samiha, S.: Automated Collaboration Coach for Video-conferencing based Group Discussions, *Proc. the ACM Int. Joint Conference and Int. Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pp. 510–515 (2018).
- [25] 鍋谷航平, 村岡泰成, 石川誠彬, 江本啓訓: 消極的発言者の発言率向上を目的とした音声による個別指示議論支援システムの開発, 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), pp. 776–773 (2021).



## 正誤表

下記の箇所に誤りがございました。お詫びして訂正いたします。

訂正箇所	誤	正
1 ページ 先頭 タイトル	遠隔会議における円滑な議論を支援 するための興味推定モデルに関する 検討	遠隔会議参加者の行動特徴を用いた 興味推定モデルの提案