

オンラインミーティング参加者の感情推定における 機械学習モデルと客観的評価との比較

鳥羽望海¹ 藤本まなと² 諏訪 博彦^{1,3} 酒井 元気⁴ 酒造 正樹⁵ 安本 慶一¹

概要: 労働者の心理状態を表す Work Engagement を知ることで, 上司や産業医は労働者の好不調を事前に把握することができ, 結果的に健全な組織運営を実現することができる. 一方, 昨今の COVID-19 の影響で, 労働者がテレワークが取り入れられるようになり, 上司や産業医にとってテレワーク中の労働者の心理状態は把握しにくい. 我々は費用と労力, 感染リスクを減らせるメリットがあるオンラインミーティングサービスに着目し, オンラインミーティングで得られる動画や音声などのマルチモーダルデータを用いて労働者の感情を推定する手法を検討する. オンラインミーティングサービスを用いてグループディスカッションを参加者にしてもらい, そこから映像, 心拍など様々なデータをセンシングして参加者の状態を推定する. グループディスカッション中の参加者の発言録から得られた感情極性, 顔のランドマーク座標, 心拍, 他者に対する感情アノテーションの 4 種類の指標から, 他者への感情アノテーションを推定するために Light GBM (Gradient Boosting Machine), SVR (Support Vector Regression), 我々自身による客観的評価を行った結果, MAE に関して機械学習モデルである Light GBM, SVR のほうが客観的評価よりも 6 つのセッションのうち 5 つのセッションで良い結果を出力することがわかった.

Comparison between Machine Learning and Objective Evaluation in Estimating Emotions of Participants in Online Meeting

Nozomi Toba¹ Manato Fujimoto² Hirohiko Suwa^{1,3} Motoki Sakai⁴ Masaki Shuzo⁵ Keiichi Yasumoto¹

1. はじめに

Zoom や Microsoft Teams などのオンラインミーティングサービスの 1 日あたりの参加者数は, 2019 年 12 月から 2020 年 4 月で, それぞれ 1,000 万人から 3 億人, 2,000 万人から 7,500 万人にまで増加した^{*1}. これは 2020 年から現在まで, 新型コロナウイルス感染症 (COVID-19) のパンデミックにより, 世界中で外出制限の措置がなされ, 仕事におけるオンラインミーティングサービスの利用が急速に広まったことに起因する.

オンラインミーティングサービスは, 費用と労力, 感染

リスクを減らすことを期待できる一方, オンサイトでのミーティングと比較して, 利用者の表情や感情の把握が難しく, 労働管理の観点から課題が指摘されている.

オンラインミーティングにおける参加者の効率性に関連する研究として, Samrose ら [1] は, オンラインミーティングに対する参加者の意識を向上させる参加者ごとにパーソナライズされたダッシュボードの構築を提案している. Cutler ら [2] は, ミーティングの有効性とミーティングの包括性を調査するため, COVID-19 以前に大規模なテクノロジー企業で大規模な電子メール調査を実施し, オンラインミーティングの有効性の多変量モデルを導き出し, 電子メールと参加者のミーティングへの貢献度の相関性を示している. ミーティングの効果を向上させることで, 企業は時間と費用を大幅に節約でき, 組織での作業環境と従業員の定着率も改善すると仮定している. しかし, これらの既存研究では, 視線や心拍に関する解析を行っておらず, 高精度な感情推定やノンバーバルコミュニケーションの情報

¹ 奈良先端科学技術大学院大学, Nara Institute of Science and Technology

² 大阪公立大学, Osaka Metropolitan University

³ 理化学研究所, RIKEN

⁴ 日本大学, Nihon University

⁵ 東京電機大学, Tokyo Denki University

^{*1} <https://blog.zoom.us/a-message-to-our-users/>

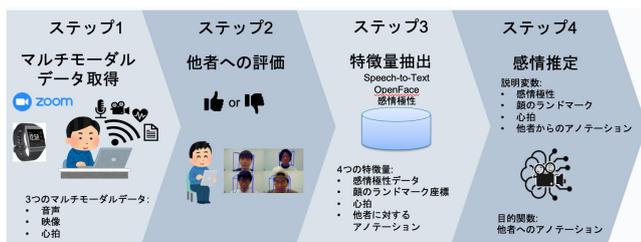


図 1 モデルの概要図

を加味できていない。さらに、アノテーションがなく、客観性が不足している。

本研究では、時系列性を考慮したマルチモーダル情報を用いた感情推定システムを提案する。本システムは、図 1 のような 4 つのステップで構成されている。具体的には、1) グループディスカッションを実行・記録するステップ、2) 他者への対するアノテーションを記録するステップ、3) 得られたデータから特徴量を抽出するステップ、4) 機械学習によりモデルを構築するステップである。モデル構築は、他者へのアノテーション付きグループディスカッションデータ、発言に関する感情極性データ、グループディスカッション中の参加者の心拍データ、顔のランドマーク座標データといったマルチモーダルなデータを用いる。複数の機械学習アルゴリズムにより感情推定モデルを構築し、その結果を比較検討する。

提案手法を評価するため、オンラインミーティングのグループディスカッションに参加者にしてもらい、そこから映像、心拍など様々なデータをセンシングして、参加者の他者への感情アノテーションを収集する実験を行う。グループディスカッション中の参加者の心拍、発言録から得られた感情極性、顔のランドマーク座標、他者からの感情アノテーションの 4 種類の指標を用いて、Light GBM, SVR による感情推定モデルを構築する。また、構築したモデルが適切に参加者の感情を推定できているかどうかを判断するため、我々自身が映像と音声を視聴して得た客観的評価と比較した。その結果、MAE に関して機械学習モデルである Light GBM, SVR のほうが客観的評価よりも 6 つのセッションのうち 5 つのセッションで良い結果を出力することがわかった。

本研究の結果は、オンラインミーティングに参加する労働者の感情の動きから Workaholic[3] や Burnout[4], [5], [6] を事前に把握し、労働者の健全な管理に貢献する可能性があることを示唆している。

2. 関連研究

本章では、オンラインミーティングにおける参加者へのフィードバック手法として MeetingCoach を紹介し、その機能と課題を述べる。その上で、課題を解決するための技術として、表現分析手法、音声データ処理手法、心拍を用

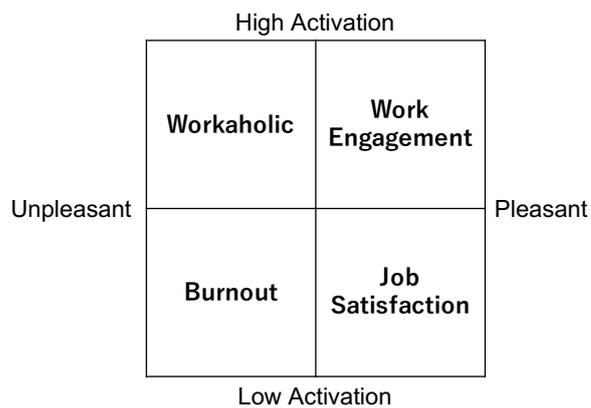


図 2 Work Engagement, Burnout, Workaholic の関係性

いた感情推定手法について述べる。

2.1 労働者の状態

労働者の状態としては Work Engagement がある [7]. Work Engagement とは、仕事に関連したポジティブな心の状態と定義されている。一方でその対概念として、Burnout[4], [5], [6] がある。Burnout とは、仕事に関連したネガティブな心の状態と定義されている。さらに、Workaholic とは、一般的に、過度に働きすぎる傾向と表現され、仕事に執着し、仕事から離れようとしなことが特徴的である [8], [9]. Workaholic は、達成したいという非常に強い欲求を持っているため、仕事をしなければならないという強迫観念に抵抗することができない傾向がある [10]. Oerlemans が構築した仕事に関連した主観的幸福感の二次元的な見方 [11] を簡略化した形式で Work Engagement, Burnout, Workaholic の関係性を図 2 に示す。Oerlemans はこれら 4 つの労働者の状態を提示し、労働者を状態に応じて適切に対処することを提案した一方で、例えばオンラインミーティングを活用した手法を用いて労働者を定量的に評価することについて触れていなかった。

2.2 参加者へのフィードバック

オンラインミーティングにおける参加者へのフィードバックをする手法として、Samrose ら [1] は、参加者ごとにパーソナライズされたダッシュボードである MeetingCoach の構築を提案している。MeetingCoach は、参加者にコンテキストおよび行動のミーティング情報を要約して提供する。また、話者交替、感情などの信号を識別する。MeetingCoach の設計は、実際のオンラインミーティングの参加者によって評価している。彼らは、4 週間にわたって 8 チームとの電話ミーティングを記録するためのツールを構築した。ダッシュボードを作成することにより、参加者の参加意欲の向上に役立つことを示している。しかしながら、(1) 心拍や視線に関する解析を行っていないこと、(2) アノテーションがなく、客観性が足りないことが課題とし

て挙げられる。

2.3 表情分析手法

表情を分析する手法として、Baltrušaitisら [12] は、OpenFace を提案している。OpenFace は、顔のランドマーク検出、頭部のポーズ推定、顔の Action Unit の認識、視線の推定が可能なオープンソースツールである。OpenFace はリアルタイム性能を備えており、特別なハードウェアを必要とせず、単純なウェブカメラから実行することが可能である。顔のランドマークの検出と追跡には Conditional Local Neural Fields (CLNF)[13] を使用している。彼らが提案した CLNF モデルでは、顔にある 68 個のランドマークをすべてまとめて検出する。また、顔の Action Unit の認識は、[14], [15] のフレームワークに基づいている。実験では 100 ピクセル以上の顔画像に対して適切な結果が得られることが示されている。

2.4 音声からの話者交代推定手法

音声処理アルゴリズムに関して、Tashev[16] は音声と雑音の異なる確率密度関数に基づいたオフライン音声活動検出器 (Voice Activity Detector: VAD) を提案している。提案された VAD アルゴリズムは、周波数領域で動作し、各オーディオフレームの各周波数に対する音声信号の存在確率と、各フレームの音声存在確率を推定し、さらにフレームごとに二値判定を行っている。これによりグループディスカッション参加者の話者交代を検出することが期待される。また、音声データからテキストデータを手軽に抽出する技術が開発されている。具体的なサービスとして、Google Speech-to-Text^{*2}などがあげられる。これらの技術を用いることにより、誰がどんな発言をしているのかを抽出することが可能となる。

2.5 心拍からの感情推定手法

Harperら [17] は、単峰性心拍時系列から感情の価値を分類する End-to-end の深層学習モデルを提案している。また、これらの価値予測に対する不確実性をモデル化するためのベイジアンフレームワークを提案している。これらの結果は、非侵襲的なデータ収集と予測の確実性が重視されるヘルスケアなどの実世界の領域における感情コンピューティングの応用の基礎となると述べている。この技術を用いることで、生体データから感情を推定することが可能となる。

2.6 本研究と関連研究の関係

本研究では、労働者の状態である Work Engagement を把握する手法で MeetingCoach で実現していない心拍や視

線、グループディスカッション参加者自身から他者へのアノテーションを付与することで、客観的な感情推定を行うことができるか検討する。具体的には、表情分析手法や音声分析、心拍からの感情推定手法を用いることで特徴量を抽出し、その特徴量を用いて Light GBM や SVR といった機械学習モデルで分析する。

3. 対象とするグループディスカッションデータ

本章では、対象とするグループディスカッションデータ (日本大学酒井元気らのグループが提供) について説明する。このデータセットには、(1) オンラインミーティングサービスを用いた参加者 3 人から 4 人が写ったグループディスカッションのギャラリー映像、(2) 発話者のみが写った映像、(3) 各参加者の音声、(4) 音声を用いて発言をテキスト化するサービス Google Speech-to-Text により生成した発言録、(5) 発言録から取得された感情極性データ [18]、(6) ウェアラブルデバイスから取得された各参加者の心拍データ、(7) 他者へのアノテーションが含まれる。それぞれの詳細は以下の通りである。

(1) ギャラリー映像：

オンラインミーティングサービスの表示におけるギャラリーモードを用いて実現している。グループディスカッション実験に参加している参加者全員が表示される映像を MP4 形式で保存している。図 3 に、2022 年 2 月 28 日 11 時 5 分から 16 時 39 分までに行われた各セッション 15 分で構成された 6 セッションのオンラインミーティングサービスを用いたグループディスカッションのギャラリー映像から抽出した 1 フレームの例を示す。

(2) 発話者のみが写った映像：

オンラインミーティングサービスの表示におけるスピーカーモードを用いて実現している。グループディスカッション実験に参加している参加者が発言した際に画面全体に表示される映像を MP4 形式で保存している。

(3) 各参加者の音声：

オンラインミーティングサービスの「参加者ごとに個別のオーディオファイルで録音」モードを用いて実現している。グループディスカッション実験に参加している参加者の音声を M4A 形式で保存している。図 4 に、上記グループディスカッションのうち参加者 A の音声の例を示す。

(4) 発言録：

Google Speech-to-Text により、録音した各参加者の音声をテキスト変換することで実現している。CSV 形式で保存している。カラムインデックスは時間、発話者、発話内容となっており時系列で保存している。

^{*2} <https://cloud.google.com/speech-to-text>



図 3 映像データ

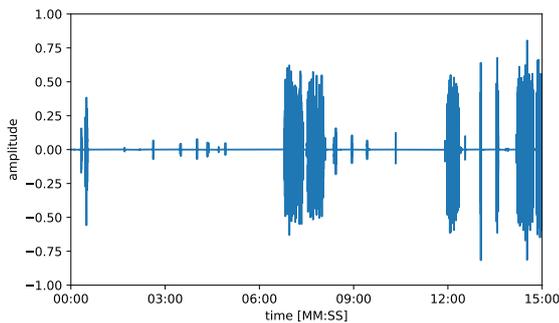


図 4 音声データ

(5) 感情極性データ：

意見や態度を含んだ感情をテキストから分析するための重要なリソースとして挙げられる単語の感情極性を用いる [18]. Google Speech-to-Text によりテキストに変換された参加者の発言内容を MeCab[19] により形態素解析を実行したのち、各単語に関して感情極性データを対応させ、それぞれの文に対して平均を取った。

(6) 心拍データ：

各参加者にウェアラブルデバイス Fitbit Ionic[®]*3を装着してもらい、そこから心拍を取得することで実現している。Fitbit SDK における Web API を用いて心拍データを抽出し、JSON 形式を変換して CSV 形式で保存している。図 5 に、上記グループディスカッションのうち参加者 A の心拍波形の例を示す。

(7) 他者へのアノテーション：

オンラインミーティングサービスによるグループディスカッション収録後、発話者のみが写った映像を用いて、各参加者がグループディスカッションに同席していた他の参加者の発言に対して、-4 点から 4 点で評価してもらい、悪ければ -4 点、良ければ 4 点の評価をつけることで CSV 形式で実現している。表 1 に、上記グループディスカッションのうち参加者 A による、他者へのアノテーションの例を示す。

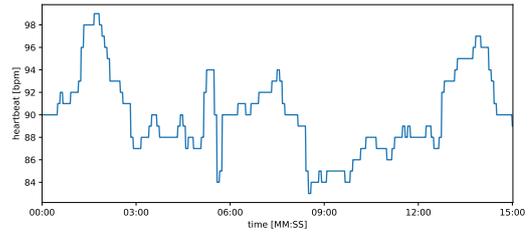


図 5 心拍データ

表 1 参加者 A の感情アノテーションを一例として示す

経過時間	感情ラベル	評価対象の参加者
00:46.0	3	B
02:00.5	2	C
03:11.3	0	D
03:30.8	-1	C
03:53.8	0	B
04:04.6	2	B
04:46.7	3	B
05:09.4	2	D
05:41.2	0	C
05:48.8	1	C
06:10.7	0	D
06:36.1	-1	D
07:01.9	0	A
07:58.8	1	C
08:00.3	0	C
08:00.9	-1	C
08:13.3	3	B
09:30.2	2	D
09:34.1	0	D
09:43.2	-1	D
09:58.9	0	B
10:09.4	2	C
10:27.9	1	D
10:40.5	0	B
11:00.6	3	B
11:46.9	2	B
12:23.2	0	D
12:55.5	-2	D
13:36.5	0	C

4. 感情推定システム

本章では、グループディスカッション参加者の感情推定システムについて提案する。提案システムでは、オンラインミーティングサービスやウェアラブルデバイス、グループディスカッション参加者が他者に対して行ったアノテーションから得られたデータから参加者の感情推定を行う。

本システムは、図 1 のような 4 つのステップで構成されている。1) グループディスカッションを実行・記録するステップ、2) 他者への対するアノテーションを記録するステップ、3) 得られたデータから特徴量を抽出するステッ

*3 <https://www.fitbit.com/>

プ、4) 機械学習によりモデルを構築するステップである。

1) では、オンラインミーティングサービスとウェアラブルデバイスを用いて、参加者4人にグループディスカッションをしてもらう。その際に音声、映像、心拍の3種類のマルチモーダルなデータを取得する。

2) では、参加者には音声と映像のデータから判断して他者の発言に対して評価をしてもらう。良いと判断した場合は4点、逆に悪いと判断した場合は-4点として9段階で判断してもらう。このデータをアノテーションデータと呼ぶことにする。

3) では、音声、映像、心拍、アノテーションの4種類のデータから Speech-to-Text や感情極性辞書、OpenFace を用いて、4種類の特徴量を取得する。4種類の特徴量とは、参加者の発言から取得した感情極性データ、映像から取得した顔のランドマーク座標、心拍、他者に対するアノテーションを指す。

4) では、得られた4種類の特徴量を説明変数として、また自身が行った他者へのアノテーションを目的関数として機械学習にかける。

5. 既存データの分析実験

本章では、提案システムの実現可能性を探るため、既存のグループディスカッションデータからの特徴量の抽出とその特徴量を用いた感情推定モデルの構築を試みる。

5.1 分析対象

分析対象は、2022年2月28日11時5分から16時39分までに行われた各セッション15分で構成された6セッションのオンラインミーティングサービスを用いたグループディスカッションデータである。参加者は、19歳から22歳の男子大学生で構成された4人で、事前にA、B、C、DのIDを割り振っている。

また、各セクションで扱ったテーマと各参加者の立場は表2のようにになっている。このテーマは、議論が活発になりそうなものを取り上げており、参加者の発言を増やすことを目的としている。ただし△で表記されているものは、中立的な立場の意見である。

5.2 特徴量の抽出

音声データからテキスト抽出をするため、Google Speech-to-Text を用いて、個別に保存されていた参加者4人の音声データをテキスト化し、発言した時間に合わせて統合を行った。図4に、参加者Aの音声データを示す。図4より、発話のタイミングを確認することができる。また、この音声データをフーリエ変換するなどすることにより、声質などを取得できるようになると考える。また、表3に、4人の参加者がグループディスカッションをした内容を実際に統合して生成したテキストデータを示す。テキストデー

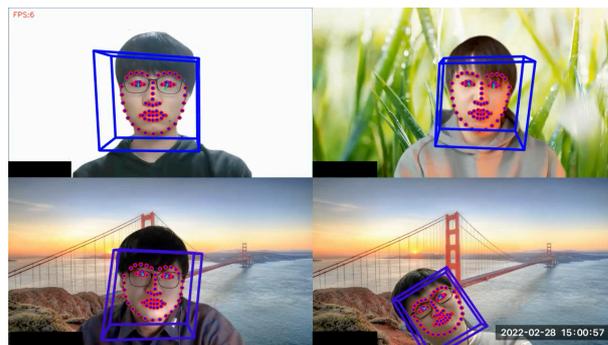


図6 OpenFaceで映像データ解析した結果

タを確認することで、与えられたテーマに沿った発言をしているかどうかを判断することや、ポジティブ・ネガティブ分析が可能となり、感情アノテーションと統合して発言の適切性を確認することが可能となる。

映像データから表情を検出するため、OpenFaceの映像内に複数人が映っている場合に表情認識を行う FaceLandmarkVidMulti を用いて、複数人の顔のランドマーク検出を行った。4人の参加者の映像(図3)を実際に OpenFace を用いて顔のランドマーク検出を行って生成した映像を図6に示す。これにより、顔の向きから視線の位置が、ランドマークから表情の検出が可能となっている。

以上の2つの特徴量に加え、心拍データと他者からの感情アノテーションを加えた4つの特徴量とする。

5.3 感情推定モデルの構築

グループディスカッション中の発言録から得られた感情極性、顔のランドマーク座標、参加者の心拍、他者からの感情アノテーションの4指標から、他者への感情アノテーションを推定するため、Light GBM[20]とSVR[21]を用いて感情推定モデルの構築を試みる。モデルの評価は、平均絶対誤差 (Mean Absolute Error: MAE) で行う。MAEは以下の式で表される。

$$M = \frac{1}{n} \sum_{t=1}^n |A_t - F_t|$$

上式において、 M がMAE、 n がフレーム数、 F_t がモデルが推定した感情ラベル、 A_t が参加者がアノテーションした感情ラベルを表す。また、個人ごとのモデルを検証するため、15分で25フレーム毎秒6セクションのデータ(全135000フレーム、縦720ピクセル、横1280ピクセル)をセクションごとに分割し、交差検証を行なっている。

6. 分析結果

表4に、参加者Cのセッション1から6の実際にアノテーションした感情ラベルとLight GBM, SVR, 客観的評価と実際の感情ラベルを比較することで得たMAEを示す。MAEは0に近ければ近いほど、推定された結果が実

表 2 各セッションで扱ったテーマ一覧

テーマ一覧	A	B	C	D
生徒は制服を着るべきか (1), 私服でも良いか (2)	△	2	1	2
宿題は有益か (1), 有益でないか (2)	1	1	2	2
人間はその人の心か (1), 見た目か (2)	2	2	1	1
自動車免許を定年制にすべきか (1), そうでないか (2)	1	1	2	△
男性専用車両は今後作るべきか (1), そうでないか (2)	△	1	2	2
タイムマシンで未来に行くか (1), 過去に行くか (2)	2	1	1	2

表 3 参加者 A の発言録データの一部

経過時間 (s)	内容
188.5	中学か中学とかは公立通ってました。
194.6	制服が来てたんですか？
198.3	で今大学生になって私服着るようになってますけど あの一それを親に決めてもらった服を着てるんですか
210.4	うーん。
212.9	でさっきの話だとあの一公立に孤立とかあの一姿勢服を 着ているような習慣がついているとあの一自分で洋服 決められないって言ってるけどでも実際はあの一 なんとなくですけど洋服は着られてますよねちゃんと選んで
231.0	なんであんまそこに関しては心配する必要がないかなと 思ってたで特にあとは個性を発揮するっていう風な話が あったと思うんですけどなんか具体的な個性の発揮の 例とかってあります

表 4 参加者 C の MAE

セッション	Light GBM	SVR	客観的評価
1	1.657266	2.146015	2.924895
2	1.246307	1.500750	1.819257
3	1.679402	1.747157	2.643581
4	1.523522	1.700253	2.653716
5	1.354815	0.917558	1.616949
6	1.727151	1.736929	1.613559

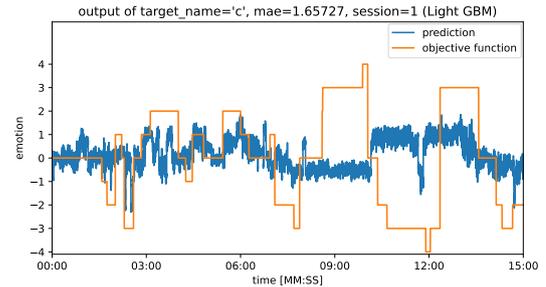


図 7 セッション 1 の参加者 C の Light GBM による分析結果

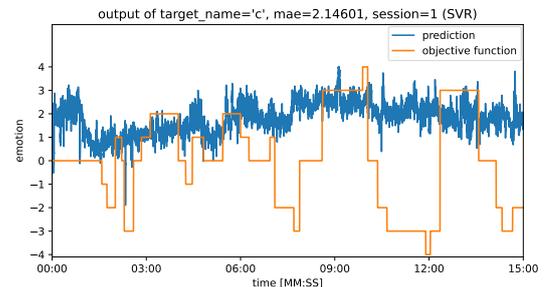


図 8 セッション 1 の参加者 C の SVR による分析結果

際の感情ラベルに近いことを表す。表 4 から見て取れるように、機械学習モデルである Light GBM, SVR により推定した結果のほうが客観的評価よりも MAE が良いことがわかる。このことから、人間よりも機械学習モデルのほうがより正確にグループディスカッションの映像や音声から参加者の感情を推定することができることが示唆される。

次に、図 7-図 15 に、参加者 C のセッション 1, 5, 6 に関して実際にアノテーションした感情ラベルと Light GBM, SVR, 客観的評価を示す。これらのセッションはそれぞれ Light GBM, SVR, 客観的評価が最も良い MAE を出力したものである (表 4)。縦軸が -4 から 4 点までの感情ラベルの値、横軸が時間、オレンジ線で描かれた矩形波が実際に参加者がアノテーションした感情ラベル、青線で描かれた波形が Light GBM, SVR, 我々自身が推定した参加者の感情ラベルを表している。オレンジ線が青線の形状に近いほど良い推定をしているということが言える。

セッション 1 (図 7-図 9) に関しては、3 分周辺の下向きを客観的評価は難しかった一方、Light GBM, SVR

で推定できている。これは、映像では参加者 C は他者の意見を聞いて頷いており、肯定していると我々自身は推定した一方、機械学習モデルは顔のランドマーク座標、心拍といった特徴量を加味した推定をしたためだと考えられる。

セッション 5 (図 10-図 12) に関しては、Light GBM, SVR, 客観的評価全てで推定するのは難しかった。これは、

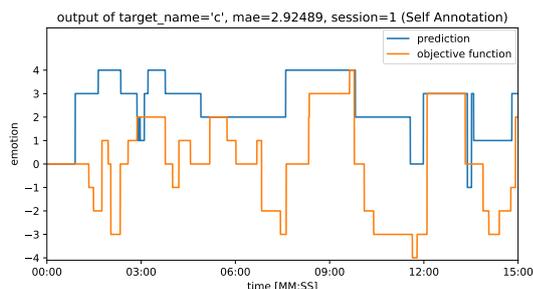


図 9 セッション 1 の参加者 C の客観的評価によるアノテーションの結果

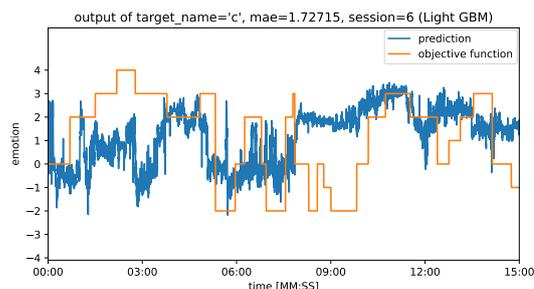


図 13 セッション 6 の参加者 C の Light GBM による分析結果

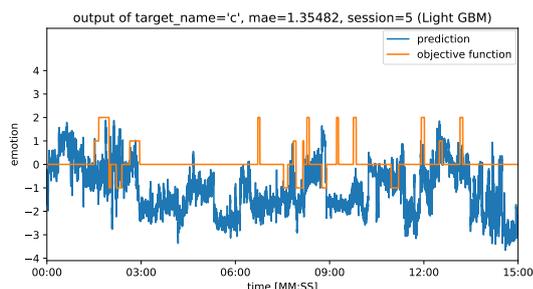


図 10 セッション 5 の参加者 C の Light GBM による分析結果

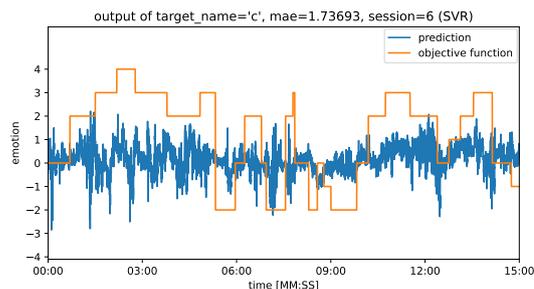


図 14 セッション 6 の参加者 C の SVR による分析結果

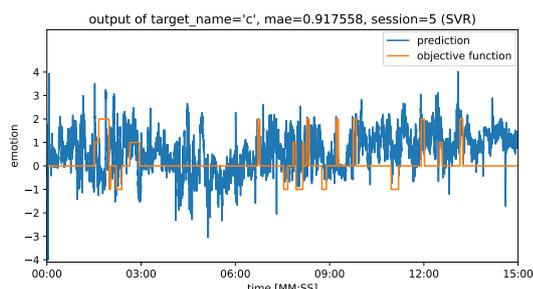


図 11 セッション 5 の参加者 C の SVR による分析結果

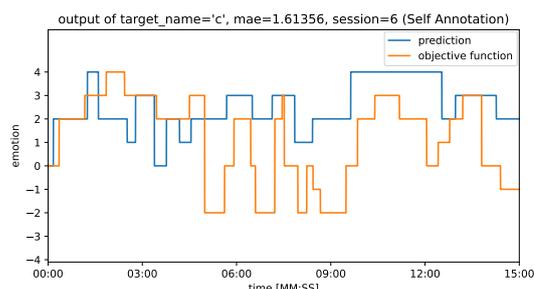


図 15 セッション 6 の参加者 C の客観的評価によるアノテーションの結果

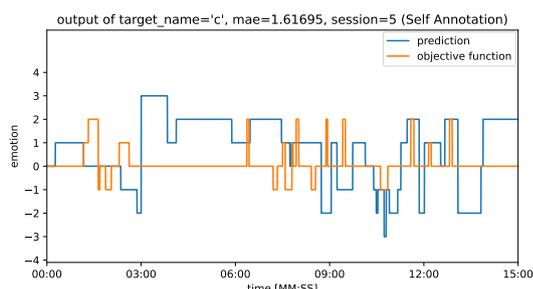


図 12 セッション 5 の参加者 C の客観的評価によるアノテーションの結果

参加者がしたアノテーションした感情ラベルの波形自体が他のセッションとは異なりパルスのような形状をしているため、他のセッションで学習してもうまく推定できなかったためだと示唆される。

セッション 6 (図 13-図 15) に関しては、後半部分の立ち上がり、立ち下がり Light GBM, SVR, 客観的評価の全てで比較的、参加者がしたアノテーションした感情ラベ

ルの波形に近い形状をしている。Light GBM に関しては、11 分周辺で正解ラベルにかなり近い値を取っている。これは、実際の映像データでも参加者 C が肯定的な発言や深い頷きをしており、これらを感情極性や顔のランドマーク座標といった特徴量をうまく学習して推定したためだと考えられる。

7. まとめ

本稿では、オンラインミーティングサービスを用いたグループディスカッションにおける感情推定システムの実現に向けた初期段階として、OpenFace による映像解析、テキスト解析のために音声からの発言録データ作成を行った。また、Light GBM, SVR, 客観的評価により感情ラベルを推定した結果、MAE において Light GBM, SVR といった機械学習モデルのほうが客観的評価よりも 6 つのセッションのうち 5 つのセッションで良い結果を得ることがわかった。一方、頷きがうまく取れていない可能性や発言録から

感情極性を組成するステップで情報が落ちている可能性が示唆される。今後は、単位秒ごとのフレーム数や顔きの回数のカウントの方式といった説明変数のデータの集約方法を見直すことや、参加者の声質といった特徴量も加味することで、音声・映像・心拍・テキストデータをマルチモーダルに用いてオンラインミーティングサービスを用いたグループディスカッションにおける感情推定システムの実現を目指す。

謝辞

本研究の一部は、Society 5.0 実現化研究拠点支援事業および科研費基盤研究 (B)(No.19H01719) , 科研費基盤研究 (B)(No.20H04177) の助成によって行った。

参考文献

- [1] Samiha Samrose, Jina Suh, Sean Rintel, Daniel McDuf, Kael Rowan, Kevin Moynihan, Robert Sim, Javier Hernandez, and Mary Czerwinski. Meetingcoach: An intelligent dashboard for supporting effective & inclusive meetings. *Conference on Human Factors in Computing Systems*, Vol. 16, No. 252, pp. 1–13, 2021.
- [2] Ross Cutler, Yasaman Hosseinkashi, Jamie Pool, Senja Filipi, Robert Aichner, Yuan Tu, and Johannes Gehrke. Meeting effectiveness and inclusiveness in remote collaboration. *Proceedings of the ACM on Human-Computer Interaction*, Vol. 5, No. 173, pp. 1–29, 2021.
- [3] Janet T. Spence and Ann S. Robbins. Workaholism: Definition, measurement, and preliminary results. *Journal of Personality Assessment*, Vol. 58, No. 1, pp. 160–178, 1992.
- [4] Herbert J. Freudenberger. Staff burn-out. *Journal of Social Issues*, Vol. 30, No. 1, pp. 159–165, 1974.
- [5] Christina Maslach and Susan E. Jackson. The measurement of experienced burnout. *Journal of Occupational Behaviour*, Vol. 2, pp. 99–113, 1981.
- [6] Christina Maslach, Wilmar B. Schaufeli, and Michael P. Leiter. Job burnout. *Annual Review of Psychology*, Vol. 52, pp. 397–422, 2001.
- [7] Hely Innanen, Asko Tolvanen, and Katariina Salmela-Aro. Burnout, work engagement and workaholism among highly educated employees: Profiles, antecedents and outcomes. *Burnout Research*, Vol. 1, No. 1, pp. 38–49, 2014.
- [8] Wayne E. Oates. On being a "workaholic". *Pastoral Psychology*, Vol. 19, pp. 16–20, 1968.
- [9] Lynley H. W. McMillan, Michael P. O'Driscoll, Nigel V. Marsh, and Elizabeth C. Brady. Understanding workaholism: Data synthesis, theoretical critique, and future design strategies. *International Journal of Stress Management*, Vol. 8, pp. 69–91, 2001.
- [10] Peter E. Mudrack and Thomas J. Naughton. The assessment of workaholism as behavioral tendencies: Scale development and preliminary empirical testing. *International Journal of Stress Management*, Vol. 8, pp. 93–111, 2001.
- [11] Wido Oerlemans. Subjective well-being in organizations. *The Oxford handbook of positive organizational scholarship*, pp. 178–189, 2011.
- [12] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. Openface: An open source facial behavior analysis toolkit. *Winter Conference on Applications of Computer Vision*, 2016.
- [13] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. Constrained local neural fields for robust facial landmark detection in the wild. *International Conference on Computer Vision Workshops*, 2013.
- [14] Tadas Baltrušaitis, Marwa Mahmoud, and Peter Robinson. Cross-dataset learning and person-specic normalisation for automatic action unit detection.
- [15] Jessica L. Tracy and David Matsumoto. The spontaneous expression of pride and shame: Evidence for biologically innate nonverbal displays. *the National Academy of Sciences*, Vol. 105, No. 33, pp. 11655–11660, 2008.
- [16] Ivan J. Tashev. Offline voice activity detector using speech supergaussianity. *Information Theory and Applications Workshop*, 2015.
- [17] Ross Harper and Joshua Southern. A bayesian deep learning framework for end-to-end prediction of emotion from heartbeat. *IEEE Transactions on Affective Computing*, 2020.
- [18] Hiroya Takamura, Takashi Inui, and Manabu Okumura. Extracting semantic orientations of words using spin model. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 133–140, 2005.
- [19] Taku Kudo. Mecab: Yet another part-of-speech and morphological analyzer. <http://mecab.sourceforge.jp>, 2006.
- [20] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, Vol. 30, pp. 3149–3157, 2017.
- [21] Mariette Awad and Rahul Khanna. Support vector regression. In *Efficient learning machines*, pp. 67–80. Springer, 2015.