

# 1 イントロ・先行研究

**背景:**

- 近年AIやデータサイエンスの発展に裏打ちされSNSや様々なデジタルフットプリントなデータが研究の資材となっている（Salganik 2018）。
- 代表的なテーマとしてText as Dataは近年の一つの主要な発展だが、映像といったVisual mediaの分析もフロンティアとなっている（Video as data）。
- そしてvideoの分析はその性質上マルチモーダルに行うことが求められる。
- 経験的には動画メディアはYoutubeやテレビに代表されるようにかなりの程度で我々の生活に普及し、娯乐的にも情報収集的にもなくてはならないものである。

**問題点:**

- 映像をどういう切り口で整理を行うべきか。どんな分析枠組みを設定すべきか。
- マルチモーダルに分析したいが技術的に可能なのか。

**先行研究:**

<映像コンテンツ分析のフレームワーク>

- 分析の基盤：Fields (1988) の質的内容分析
- Fields (1988) の8プロセス（ユニット化-書き起こし-カテゴリ開発-言語分析-音声・表情分析-シーン構成-要素の相互作用記述-説明）を出発点とする。
- 本研究の8つの観点
- ①目的・主題②出演者・キャスト③構成・演出スタイル④語り口・情報提示方法⑤視聴者とのインタラクション⑥長さ・フォーマット⑦主な配信・流通チャンネル⑧ビジネスモデル・収益構造

<video as dataマルチモーダルな分析について>

Pinto (2024) の映像計量分析（"Video as Data"）

対象：イタリアの政治討論番組 約2,000エピソード（2018～2023年）

- ・手法について
- 音声処理：①話者識別（Pyannote）②自動文字起こし（OpenAI Whisper）③音声感情分析（wav2vec2-IEMOCAP）
- 画像処理：①顔認識（MTCNN等）②感情分析（HSEmotion）
- ・メタデータ視聴率データ（Auditel™）との連携

主なimplication：

専門家の出演増加、トピックの周期性、政治的傾向と感情トーンの関係、視聴率への影響

# 2 対象データと環境

**対象データ**（Youtube APIを通じて取得）：

①「【東京都知事選2024】小池・石丸・田母神・蓮舫の4候補がネット討論会（2024年6月24日）」

②Youtubeで公開されている「Official髭男dism」の2015年から2023年までのPV/MV

**環境:**GoogleColabで実装（T4 GPUおよびv5e-1 TPU）

# 3 ケーススタディ

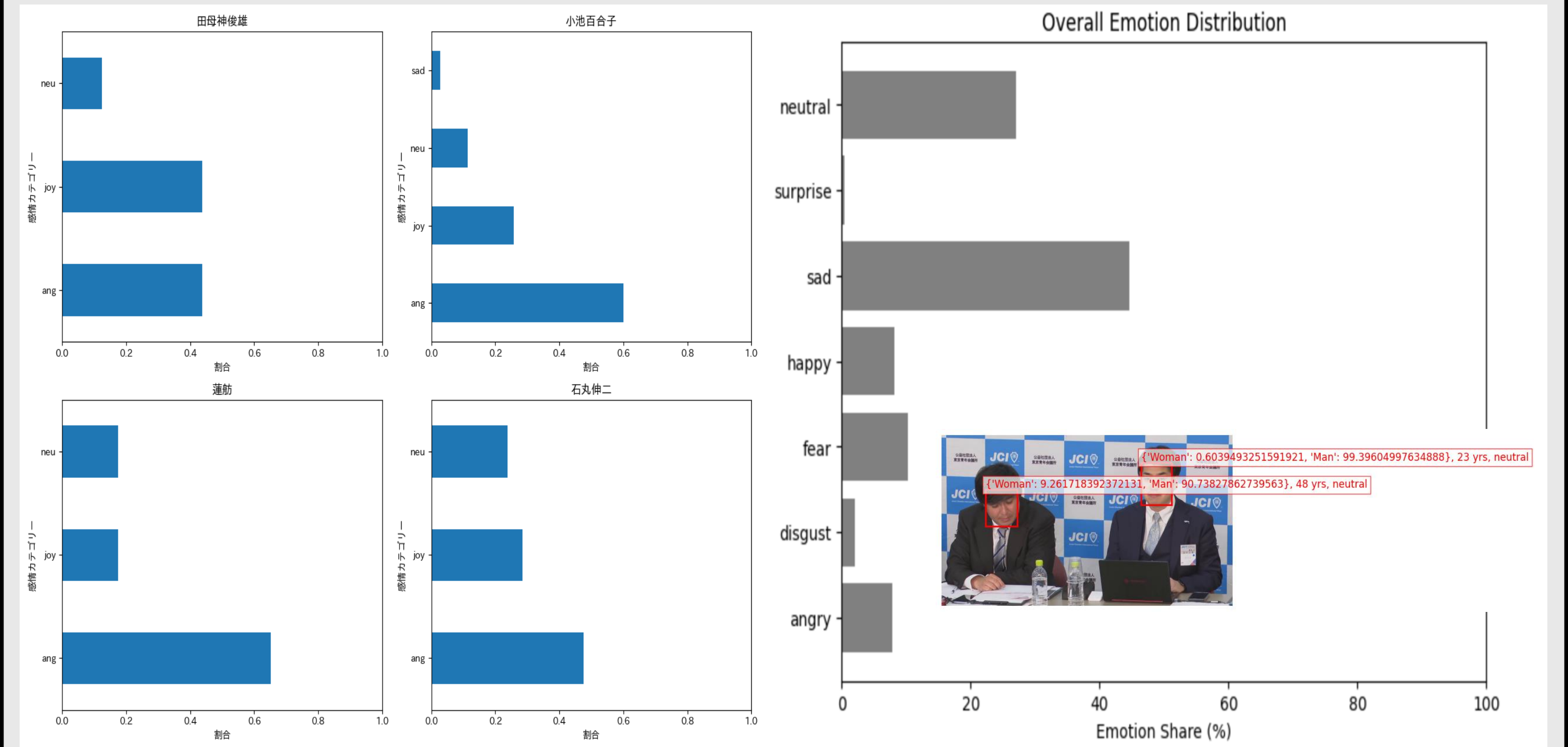
①ネット政治討論番組において出演者がどう提示されているのか:

- 音声情報について
- 発話時間を計測**(Pyannoteとwhisperを用いて話者識別/文字起こしを自動化するパイプラインを作成)

候補者名	小池百合子	石丸伸二	田母神俊雄	蓮舫
総分数（分）	23 (1394.95 秒)	14 (869.56 秒)	15 (907.9 秒)	15 (924.12 秒)

→小池氏の発言時間が8-9分近く多かった。

-**候補者ごとに発話の感情を分類**（前述したパイプラインの中でwav2vecの日本語で学習したモデルを使用）

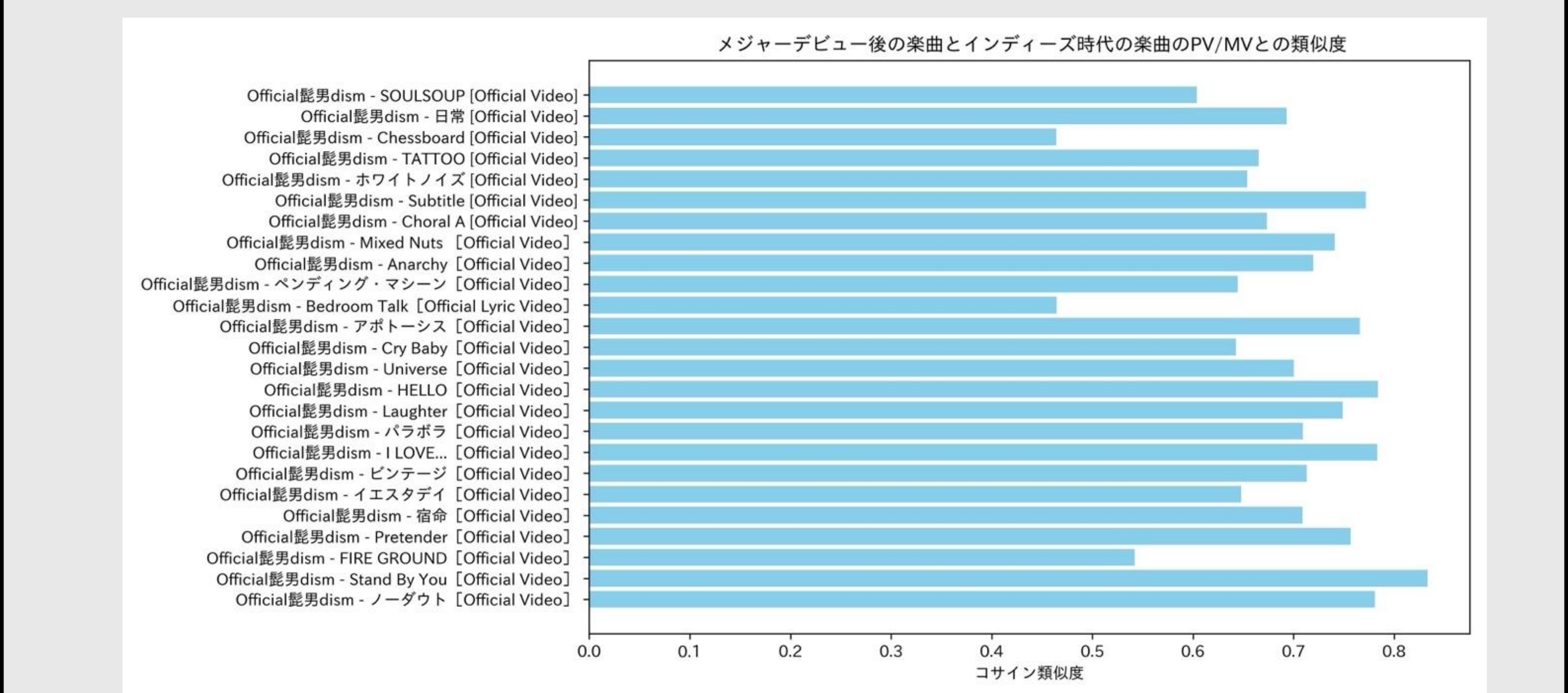


→音声感情認識(左)では、全候補者で「怒り（anger）」が支配的だったが、田母神氏は「喜び（joy）」が多く、小池氏は多様な感情を示し、石丸氏はバランスの取れた分布を示した。

- ・映像情報について
- 動画全体の感情の分布を可視化**（MoviePyライブラリでフレームを析出しMTCNNで感情分類）\*上述の右

②メジャーデビューの前後でMV/PVの表現は変わるのか:

- ・文化社会学の実証研究では、文化生産の変化に伴い、文化産品（cultural output）が変容することが指摘されている。PySceneDetectを用いてPV/MVのカットごとにフレームを抽出。Vision Transformer（google/vit-base-patch16-224-in21k）で各シーンをベクトル化し、各PV/MVのシーンベクトルの重心（centroid）を計算し、一つのベクトルに統合。インディーズ時代の楽曲とのコサイン類似度を測定。



- 課題（主に技術的課題）：
- 計算資源および保存容量の制約(感情推定の推論処理は計量資源が必要。またフレーム処理は保存容量が必要)
- 感情認識モデルの精度不足（社会科学の運用に耐えうる制度が望ましい）+マルチモーダルな分析枠組みの不在