

ミーティングにおける時間経過と感情変化の関係

産業能率大学 情報マネジメント学部 川野邊研究室 4年 黒澤 涼月 指導教員:川野邊 誠

研究概要

研究背景

- 近年、自然言語処理タスクの1つとして感情分析の研究が活発に行われており、特にコールセンター対応やSNSマーケティングなどに応用されている
- 一方で、テキストのみから感情を推定する手法には、非言語的ニュアンスの欠落などにより精度に限界があることが、多くの先行研究で指摘されている
- 音声や表情などの非言語情報を組み合わせることで各モダリティの欠点を補うことができるため、より精度の高い感情分類が可能になることが示されている
- しかし、複数人が参加するミーティング（会議）のような集団討議の場を対象とした感情分析の研究は少なく、特に時間経過に伴う感情変化を、マルチモーダル感情分類器を用いて詳細に分析した研究事例は極めて稀である

研究目的

自然言語処理、音声感情推定、画像処理の各技術を用いてグループミーティング中の時間経過に伴う感情変化を明らかにすることで、対話全体における感情の推移を評価する手法を提案する

期待効果

ミーティング中の感情変化を時系列で捉えることが可能になれば、ミーティングの効率性や生産性を高めるための具体的な改善策を提案できるようになると考える

研究手法

オンラインミーティング(Zoom)を実施し、その様子を録画

感情データベース構築

①主観アノテーション

自身のミーティング中の発言内容に対して自分が抱いた感情を基本6感情から選択

②客観アノテーション

他人の発言内容に対して、自分が抱いた感情を基本6感情から選択

*基本6感情：喜び、悲しみ、驚き、怒り、恐れ、嫌悪

深層学習を用いた感情予測

『テキスト+音声+表情』の感情分類器を用いてミーティング時の参加者の感情を予測

- 感情データベースを統計的に分析
- 感情データベースと感情分類器の予測結果を比較し感情分類器の性能を評価

研究成果

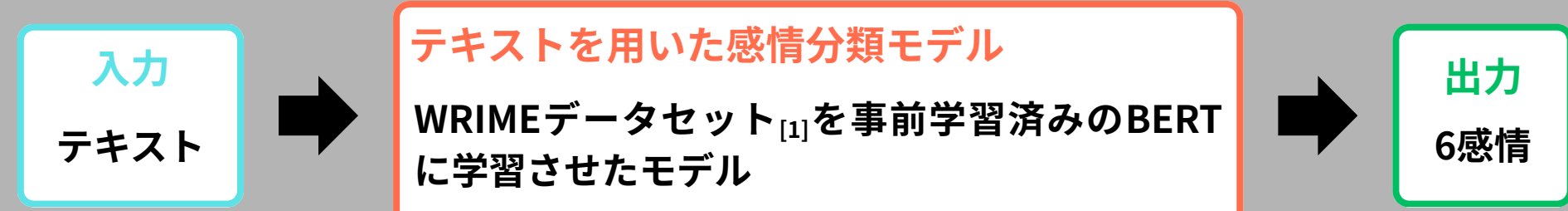
【テキストを用いた感情分類】

目的

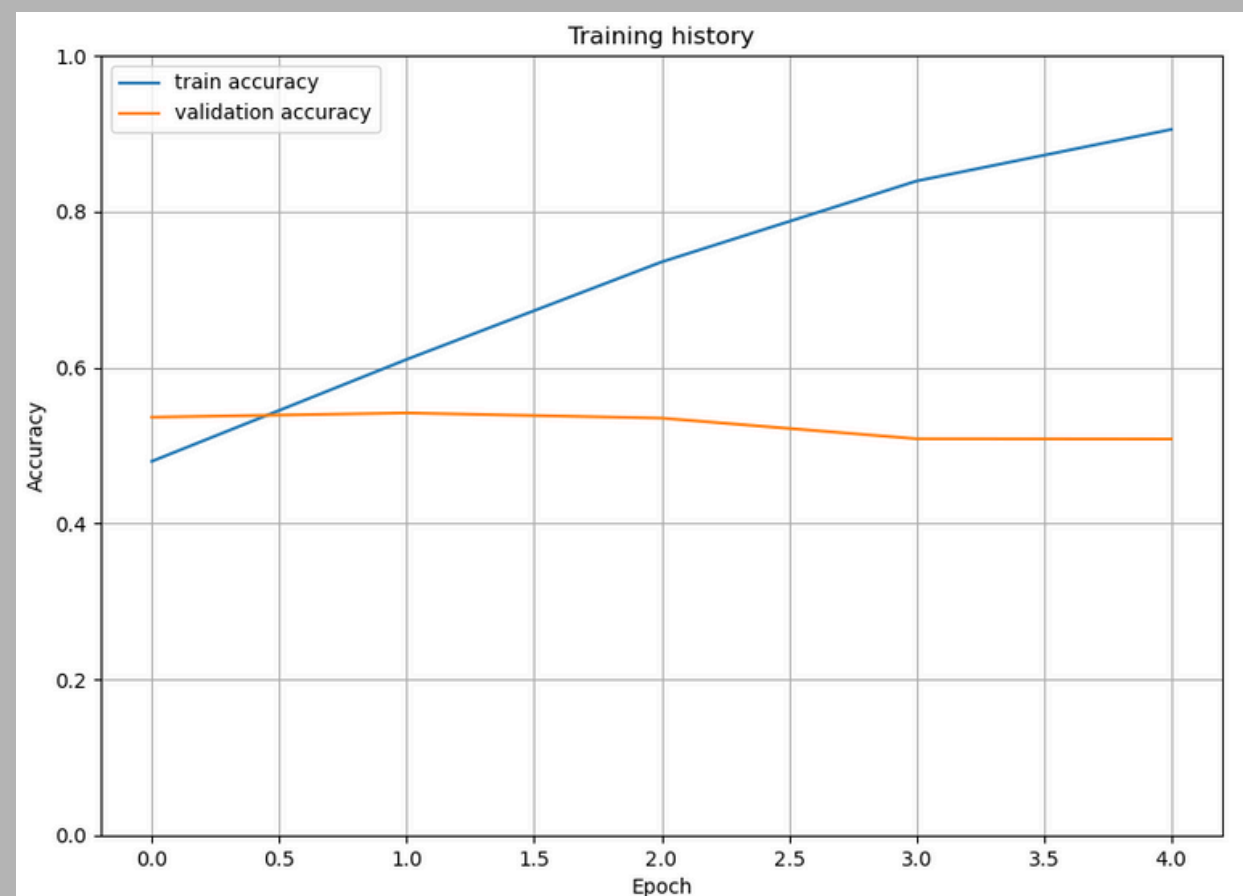
深層学習を用いてテキストから感情を推定する

手法

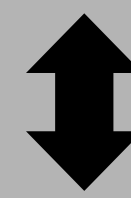
- WRIMEデータセット^[1]を8感情×4強度から6感情(強度なし)に加工した後訓練データを60%、検証データを20%、テストデータを20%に分割
- 事前学習済みのBERTモデルをファインチューニングし、テキストに対して6つの感情を予測するモデルを構築
- 学習は10エポック(早期停止は3エポック)で実施



結果



訓練データのaccuracy(精度)はエポック数が増えるにつれ上昇している



F1スコアが0.45であることからモデルがある程度感情を分類できる一方で、改善の余地があると言える

「F1スコア」
学習に使用するデータクラスに偏りがある場合に用いられるモデルの正確度を評価するための指標

【表情を用いた感情分類】

目的

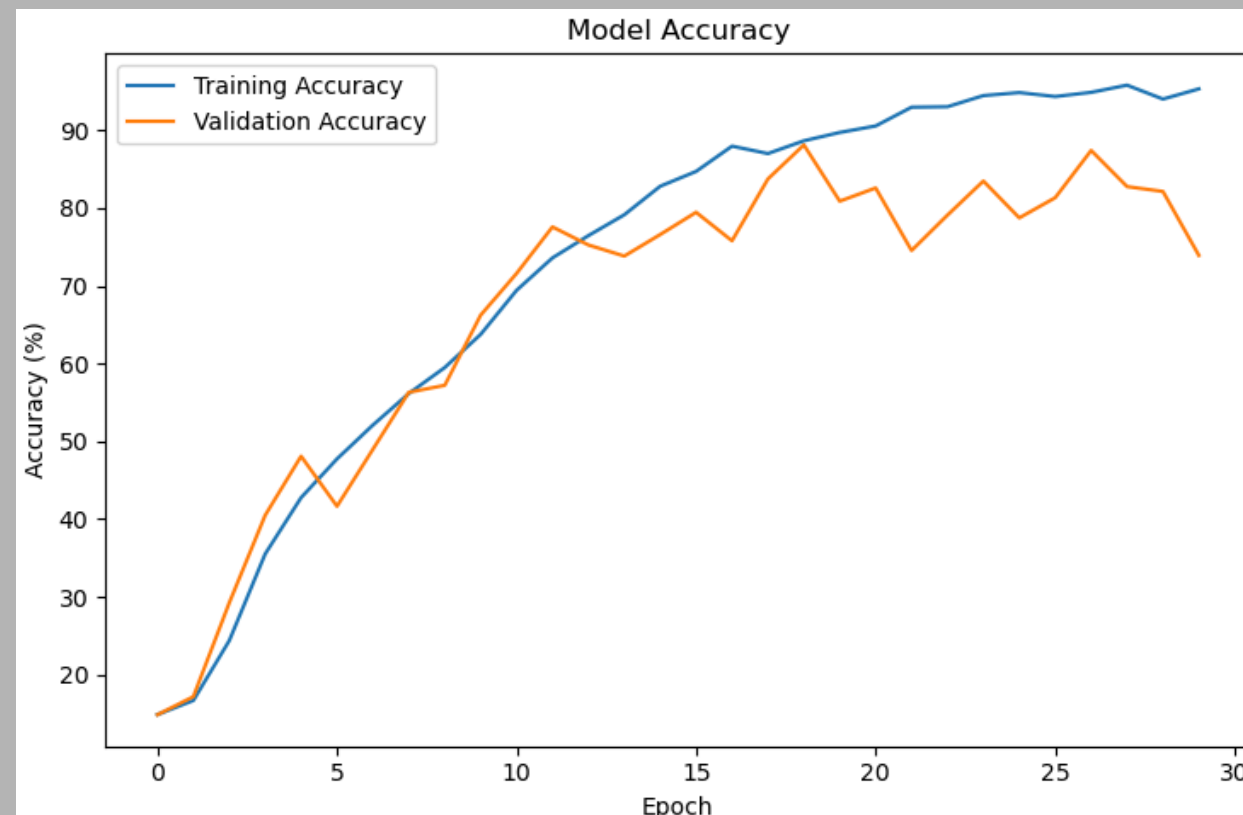
深層学習を用いて表情画像から感情を推定する

手法

- KRCデータセット^[2]に含まれる*基本6感情を分類対象とし、訓練データを80%、検証データを20%に分割
- 3段階のステージで構成され、6つの出力層を持つCNN^[3]を設計
- 30エポックの学習を実施



結果



- 訓練データと検証データのaccuracyがエポック数の増加とともに上昇している
- 各感情ラベルの平均分類精度が94%で過学習もない



表情による感情の分類性能は高い

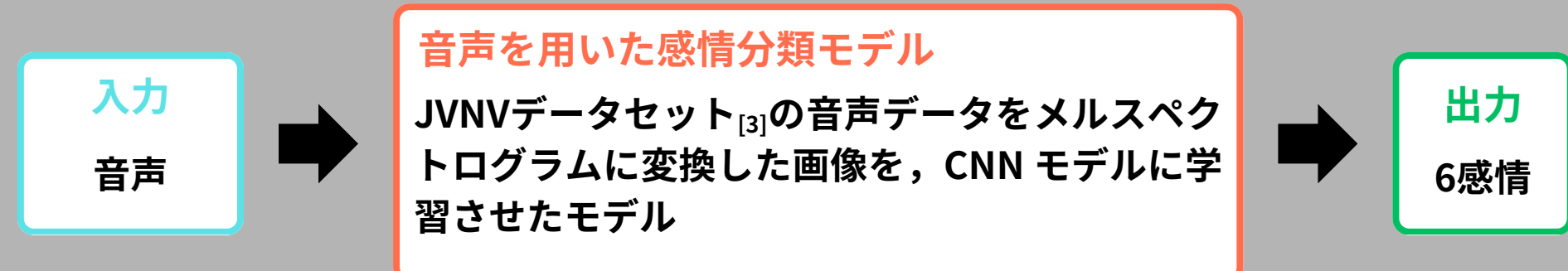
【音声を用いた感情分類】

目的

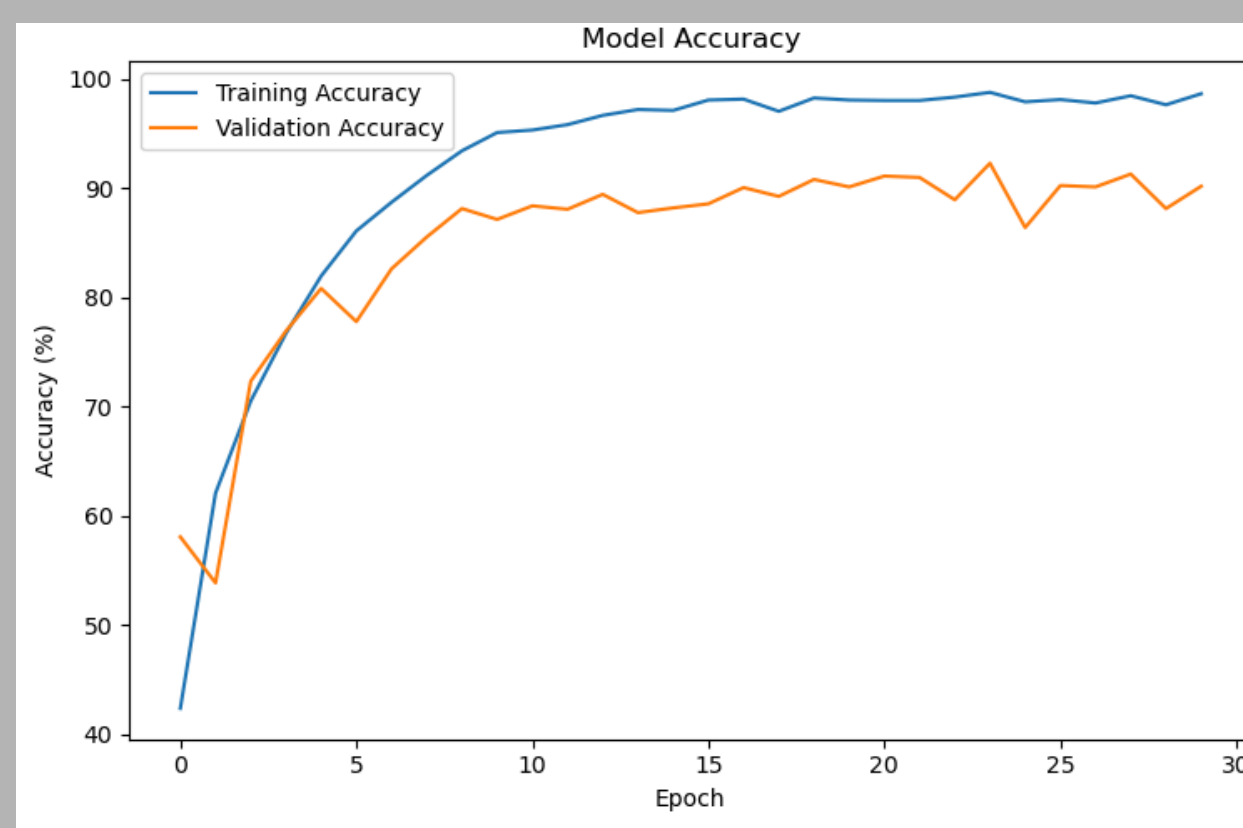
深層学習を用いて音声から感情を推定する

手法

- JVNVデータセット^[4]をデータ拡張後、6,460枚のメルスペクトログラム画像に変換
- *基本6感情を分類対象とし、訓練データを80%、検証データを20%に分割
- メルスペクトログラム画像を用いて感情を分類するCNNモデルを設計
- 30エポックの学習を実施



結果



- 訓練データと検証データのaccuracyがエポック数の増加とともに上昇している
- 各感情ラベルの平均分類精度が92%で過学習もない



音声による感情の分類性能は高い

【感情データベースの評価】

目的
ミーティングにおいて時間経過とともに参加者の感情が変化するか分析する

手法
主観データを各感情ごとに時系列で可視化する

ミーティング概要

- 実施手法：Zoomによるオンラインミーティング
- 場所：大学内
- 時間：30分
- 参加者：4人(本ゼミ所属の4年生4人で構成)
- テーマ：男女差別について

結果

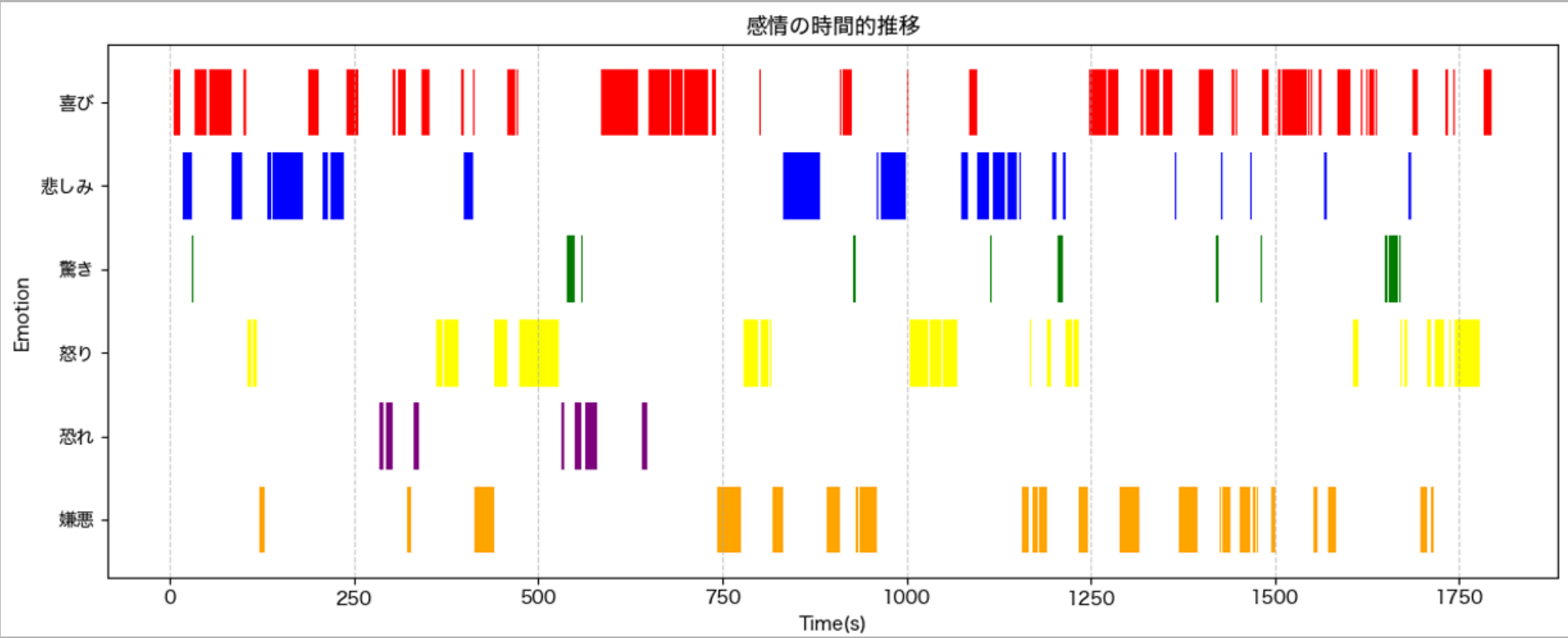


図7. 主観感情の時間的推移

- 喜びが占める割合が全体的に多い
- ネガティブな感情を抱いて発話した次の発話がネガティブであるとは限らない
- 個人ごとに感情にばらつきがある
- ミーティング終盤になるにつれて喜びの感情が増加している

【感情分類器の性能評価】

目的
実世界において(モノモーダル)感情分類器が有効か検証する

手法

テキストを用いた感情分類器の有効性検証

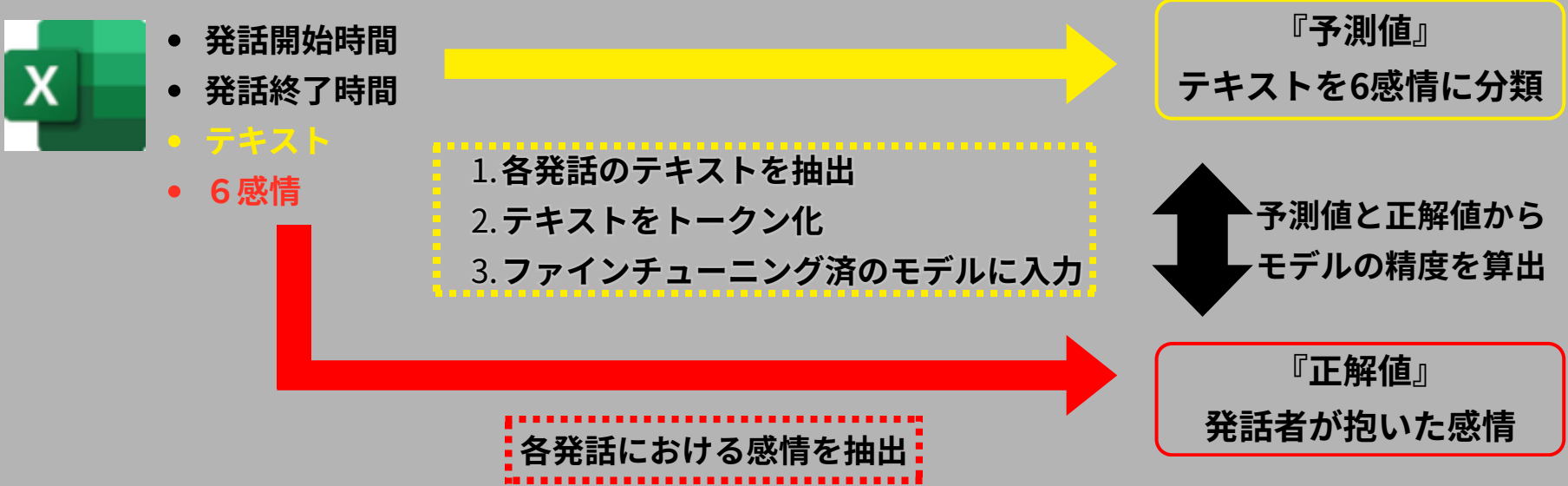


図8. テキストを用いた感情分類器の有効性検証フロー

音声を用いた感情分類器の有効性検証

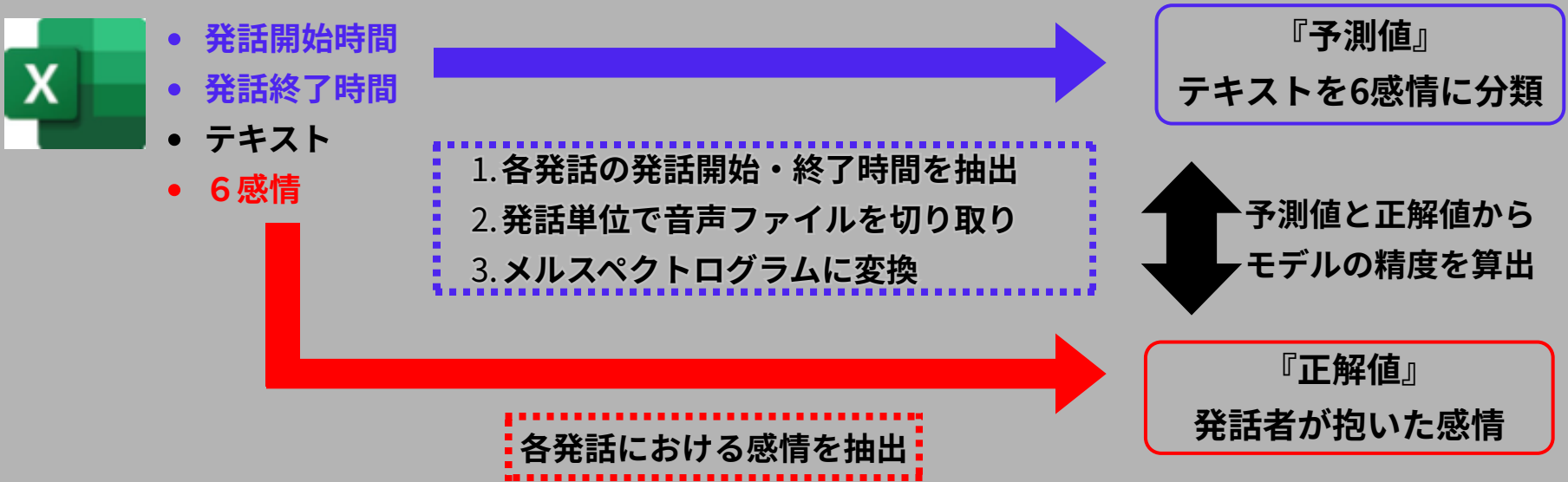


図9. 音声を用いた感情分類器の有効性検証フロー

結果

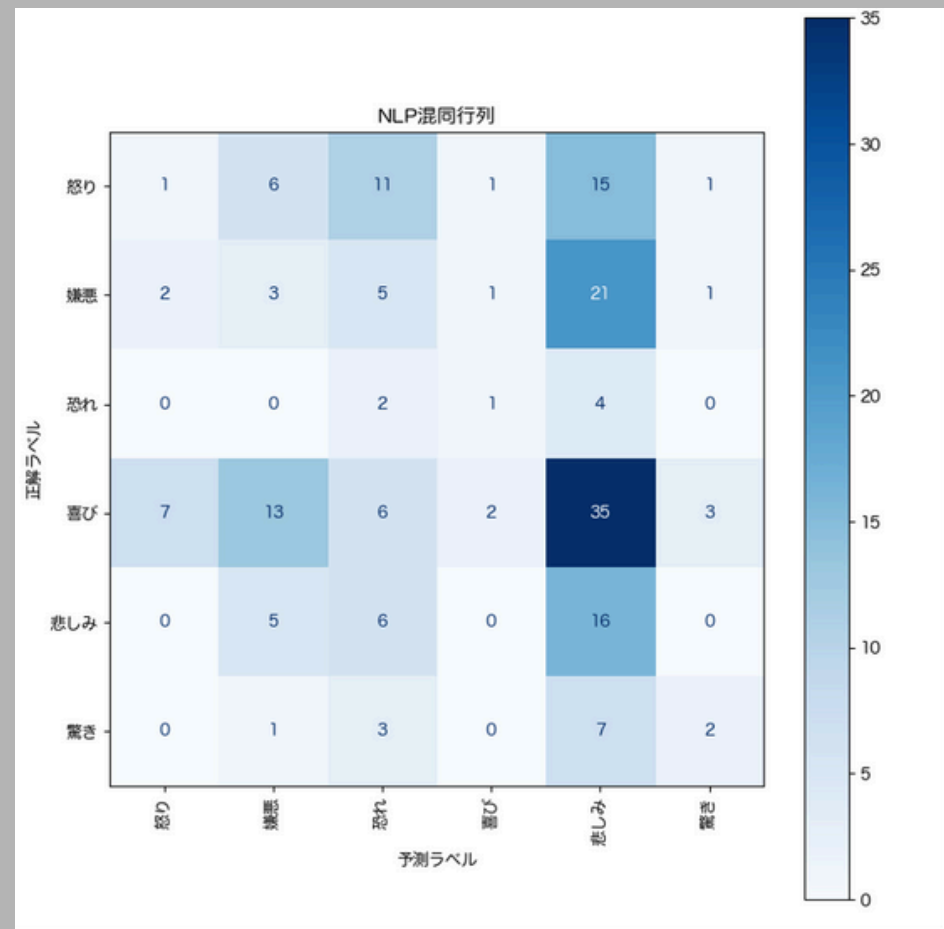


図10. NLPによる感情分類の混同行列

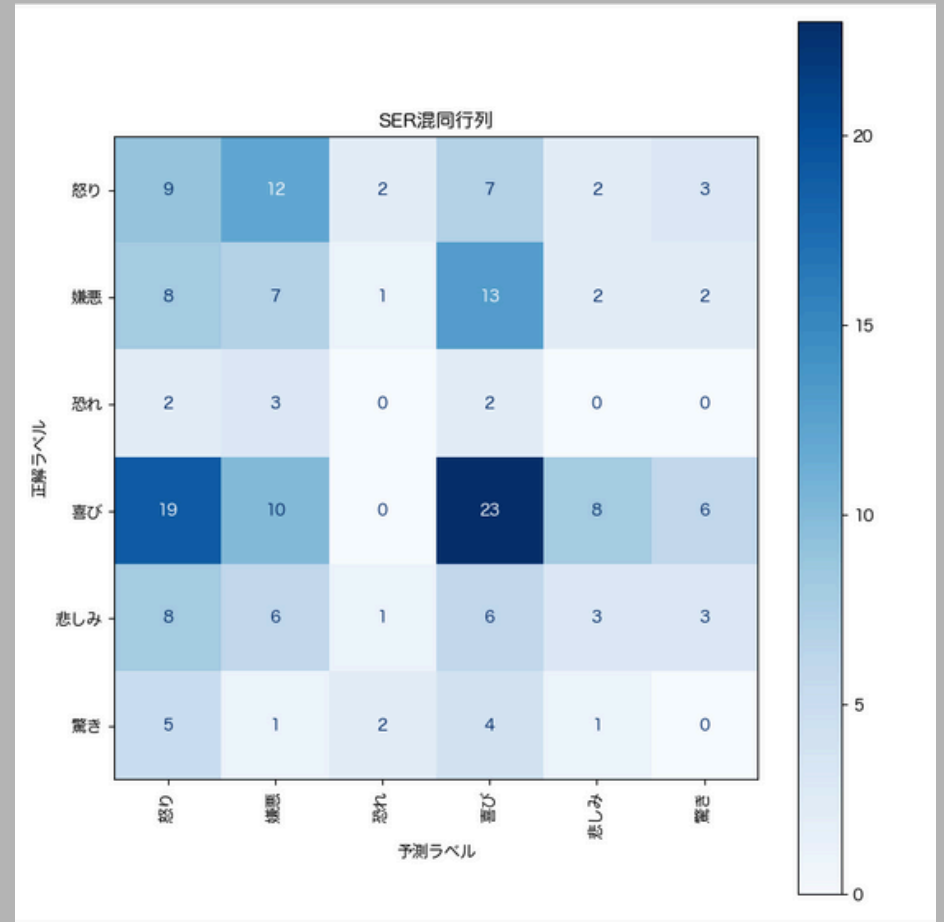


図11. SERによる感情分類の混同行列

テキストベースの感情分類器性能

- Accuracy : 18.61%
- Precision: 18.61%
- Recall: 19.70%
- F1 score: 12.59%
- 混同行列から「悲しみ」に分類される割合が高いことが判明
- 人間が「喜び」と感じているテキストに対して「悲しみ」と分類

AccuracyとF1 score がともに低いことから実世界における
応用可能性は低い

音声ベースの感情分類器性能

- Accuracy : 23.20%
- Precision: 16.03%
- Recall: 15.48%
- F1 score: 15.39%
- 混同行列から「喜び」に分類される割合が高いことが判明
- 「喜び」の分類性能はテキストによる分類性能よりも優れている

テキストベースの感情分類器より性能は高いが、Accuracy
とF1 score がともに低いことから実世界における応用可能
性は低い

【マルチモーダル感情分類器】

マルチモーダル感情分類器の処理手順

1. 入力
 - a. 各発話に対応するテキストと音声を抽出
2. 前処理
 - a. テキストをトークン化
 - b. 音声をメルスペクトログラムに変換
3. 特徴量抽出
 - a. BERT based NLP modelよりテキストの特徴量を抽出
 - b. CNN based modelより音声の特徴量を抽出
4. 結合
 - a. テキストと音声の特徴量をconcatし、全結合層を通して共通埋め込み空間を作成
5. 出力
 - a. 結合ベクトルを出力層に入力
 - b. 6感情それぞれの確率を出力

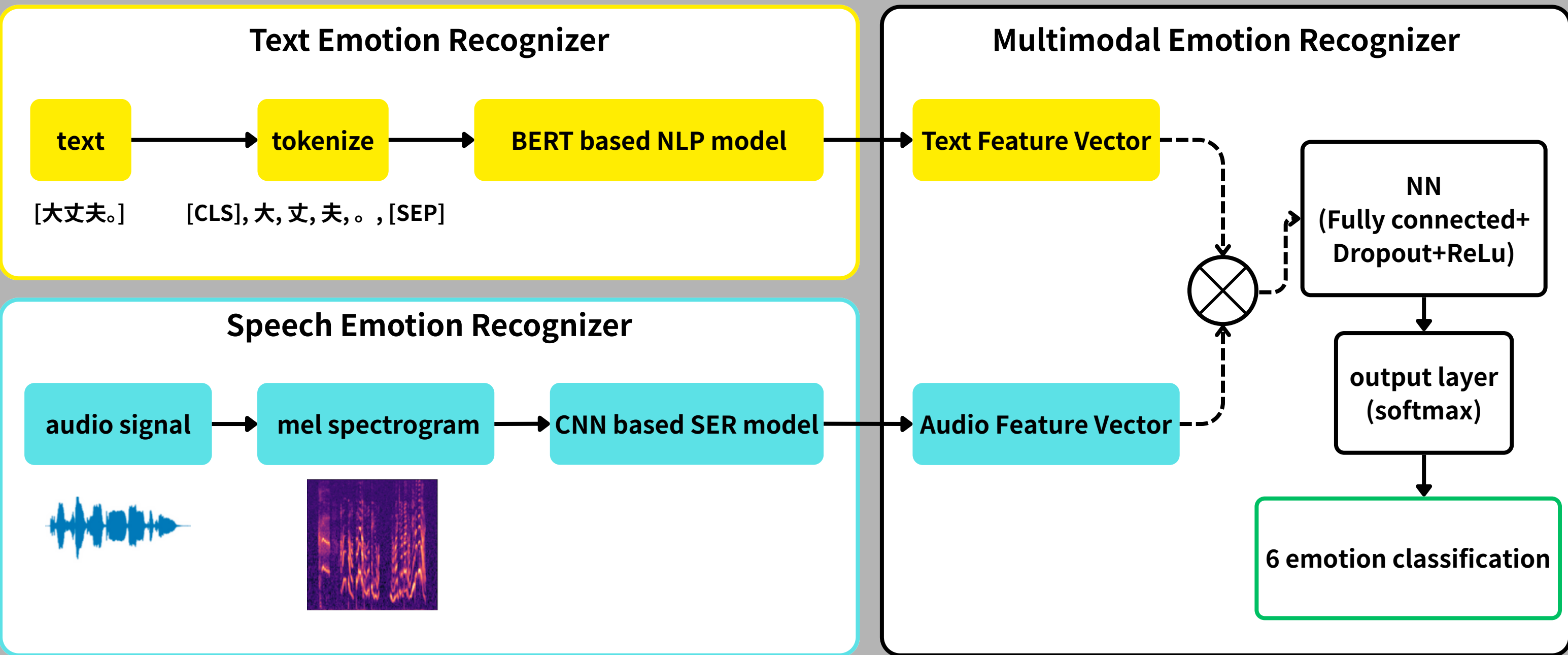


図12. マルチモーダル感情分類器のアーキテクチャ

今後の予定

- マルチモーダル感情分類器の性能評価
- 卒業論文執筆

参考文献

[1] 梶原智之, Chenhui Chu, 武村紀子, 中島悠太, 長原一. “主観感情と客観感情の強度推定のための日本語データセット”, 言語処理学会第 27 回年次大会 pp. 523-527, 2021.
[2] Ueda Yoshiyuki, Masato Nunoi, and Sakiko Yoshikawa. “Development and validation of the Kokoro Research Center (KRC) facial expression database.” Psychologia 61, no. 4 pp. 221-240, 2019.
[3] Elsheikh, Reham A., M. A. Mohamed, Ahmed Mohamed Abou-Taleb and Mohamed Maher Ata. "Improved facial emotion recognition model based on a novel deep convolutional structure." Scientific Reports 14, no. 1 (2024): 29050.
[4] Xin, Detai, et al. “JVNV: A corpus of japanese emotional speech with verbal content and nonverbal expressions”, IEEE Access 12, pp.19752-19764, 2024.