

2021 年度

早稲田大学大学院 基幹理工学研究科 情報理工・情報通信専攻 修士論文

感情単語と歌詞の類似度計算を用いた
日本語歌詞楽曲のムード分類に関する研究
Study on Mood Classification of Japanese Lyric Songs
by Similarity Calculation of Emotional-words and Lyrics

XU, Chi

(5120F037-5)

提出日：2022.01.24

指導教員：亀山渉教授

研究指導名：マルチメディア情報流通システム研究

目次

| | |
|------------------------------------|-----------|
| 第1章 序論 | 1 |
| 1.1 研究背景 | 1 |
| 1.2 研究目的 | 1 |
| 1.3 本論文の構成 | 1 |
| 第2章 関連研究と予備実験 | 3 |
| 2.1 英語歌詞のムード分類と予備実験..... | 3 |
| 2.2 日本語歌詞のムード分類 | 6 |
| 2.3 感情単語リストを使用した分類方法 | 7 |
| 2.4 感情単語ベクトルと歌詞ベクトルの関連性と予備実験 | 7 |
| 第3章 実験 | 10 |
| 3.1 実験概要 | 10 |
| 3.2 学習データ | 10 |
| 3.3 アンケート | 10 |
| 3.4 検証方法 | 12 |
| 3.5 実験手順 | 12 |
| 第4章 解析手法 | 13 |
| 4.1 前処理 | 13 |
| 4.2 感情単語リスト | 13 |
| 4.2.1 作成方法 | 13 |
| 4.2.2 バランス調整 | 13 |
| 4.3 ベクトル化..... | 14 |
| 4.4 特徴ベクトルの作成 | 14 |
| 4.4.1 セグメント化なしの場合 | 14 |
| 4.4.2 セグメント化ありの場合 | 15 |
| 4.5 クラスタリング | 15 |
| 4.6 分類手法 | 16 |
| 第5章 実験結果及び考察 | 17 |
| 5.1 セグメント化有無の比較 | 17 |
| 5.1.1 セグメント化なしの実験結果 | 17 |
| 5.1.2 セグメント化なしの実験結果 | 18 |
| 5.1.3 考察 | 18 |
| 5.2 感情単語のバランスを調整した解析 | 19 |

| | |
|------------------------|-----------|
| 5.2.1 実験結果 | 19 |
| 5.2.2 特徴量の重要度 | 20 |
| 5.2.3 考察 | 22 |
| 5.3 ノイズ除去を追加した解析 | 23 |
| 5.2.1 学習データでのノイズ | 23 |
| 5.2.1 ノイズ除去の手法 | 23 |
| 5.2.1 実験結果 | 24 |
| 5.2.1 考察 | 25 |
| 第6章 結論 | 26 |
| 6.1 まとめ | 26 |
| 6.2 今後の課題 | 26 |
| 謝辞 | 27 |
| 参考文献 | 28 |
| 図一覧 | 29 |
| 表一覧 | 30 |
| 付録 | 31 |
| 研究業績 | 33 |

第 1 章 序論

1.1 研究背景

近年、機械学習の発展に伴い、より高品質なコンテンツをユーザに提供するためのコンテンツ推薦システムが重要視されている。音楽サイトや音楽アプリケーションのユーザには、気分によって特定のムードの楽曲を聴きたいニーズがあると想定される。楽曲に対するムードがメタデータとして利用可能であれば、聴きたいムードの楽曲検索や音楽推薦システムへの応用が可能となる。しかしながら、音楽配信サービスに提供される楽曲は必ずしもムードタグが付けられていない。また、ムードデータがない大量の既存楽曲に対し、手作業でムード分類する方法は多額のコストを必要とするため、手作業での分類は現実的ではない。

一方、音楽に対する印象は主観的であり、人によって異なった感情が生起する可能性がある。特に、リズムやテンポなどの情報により論理的な理由でムードを判断することが困難である。その点に対し、歌詞はリズムよりも直接的に感情を表現すると考えられるため、ムード傾向の推定には重要だと推測している。Xiao Hu の研究[1]では、Bag-of-words モデルに基づいて英語の歌詞を分析し、音の補足としてムード分類の精度を向上させることを目指している。その結果、歌詞のみによる分類結果は音のみの結果より安定していることが示された。そこで、歌詞情報を利用したムード分類手法を改善できれば、分類精度の向上も可能となる。

歌詞情報を用いた従来の研究では、感情単語の出現頻度を着目点としている。感情カテゴリ別に出現頻度を計算すれば、感情を表す特徴ベクトルを作成することも可能であろう。しかし、そのような方法では、感情辞書や感情単語リストに収録していない単語に対して解析を行っていない。つまり、近い語彙を持つ単語が歌詞の中に出現しても、同じ形式で表現されていなければアルゴリズムに無視されることが問題点である。そこで、収録されている感情単語と歌詞に出現している単語の類似度を測ることが必要と考えられる。

1.2 研究目的

英語歌詞楽曲に対しては、既存のムードデータセットを利用することが可能である。しかし、日本語楽曲の場合は、ムードラベルが付けられている大規模なデータセットがなく、教師あり学習を実行することが困難である。さらに、本論文の 1.1 節に記述したように、感情単語の出現頻度に基づいた従来の研究は問題点があると考えられているが、歌詞と感情単語の類似度計算によりその問題点を解決できる可能性がある。故に本研究は日本語歌詞楽曲に対して、歌詞テキストと感情単語の類似度に基づき、教師なしで楽曲のムード分類を行うことを目的とする。

1.3 本論文の構成

本論文は、以下の 6 章で構成されている。

第 1 章「序論」では、本研究の研究背景、目的及び本論文の構成について記述している。

第 2 章「関連研究」では、英語歌詞楽曲に対する先行研究と日本語歌詞楽曲に対する先行研究を紹介し、感情単語リストに基づく方法について説明している。また、本研究の予備実験とした感情単語ベクトルと歌詞ベクトルの関連性検証についても説明している。

第3章「実験」では、本実験の概要、使用している学習データ、アンケートの実施方法、実験結果の検証方法と実験の手順について記述している。

第4章「解析手法」では、データの前処理と解析手法について説明している。

第5章「実験結果及び考察」では、各条件に基づいた実験結果と考察について記述している。

第6章「結論」では、本研究のまとめと今後の課題について記述している。

第 2 章 関連研究と予備実験

2.1 英語歌詞のムード分類と予備実験

英語歌詞のムード分類に関する研究は、教師あり学習を中心に行っている。歌詞テキストのベクトル化については、BoW モデルなど単語の出現頻度に基づく方法が伝統的である。しかし Whissell の研究[2]では、このような方法で抽出した各歌詞の特徴ベクトルは平均値との差が非常に微小で、感情別に区別することが困難であると指摘されている。Dan Yang ら[3]はその問題を改善するために、出現頻度に基づくベクトルを直接に使用せず、General Inquirer モデル[4]を通じて次元数が大幅に削減された特徴ベクトルに変換した。彼らは 1032 曲分のムードラベル付きの歌詞データを解析し、歌詞の感情抽出によって分類実験を行った。彼らの学習データから解析されたユニークワード数は 5000 語以上あるが、General Inquirer モデルの使用により 182 次元の感情特徴ベクトルへの変換が可能となった。

一方、BERT 系事前学習モデルの提出により、テキストをベクトル化する新たな手法がトレンドになっている。膨大なコーパスを利用して事前学習を行い、各種 downstream tasks のために適切な埋め込みを用意しておくことが BERT 系モデルの特徴と考えられている。そこで本研究の予備実験として、BERT 系モデルを用いた英語歌詞のムード分類を試行した。

予備実験用の歌詞データは、Erion Cano が作成した「MoodyLyrics4Q」データセット[5]を使用している。「MoodyLyrics4Q」データセットに収録されている歌詞は合計 2000 曲分であり、各歌詞が“Happy”, “Angry”, “Sad”, “Relaxed” の 4 種類の感情カテゴリのいずれかに所属している。それらのデータのうちに、400 曲分のデータをテストデータとして扱う。また、実験用の BERT 系モデルは計算効率を重視した ALBERT モデル(Google research base v3.0)[6]を使用している。Finetuning 段階のパラメーターについては、以下の値に設定している。

表 2.1.1 ALBERT のパラメーター(Finetuning)

| パラメーター | 設定値 |
|-----------------------|---------|
| Max length | 320 |
| Batch size | 16 |
| Optimizer | Adam |
| Learning rate | 1e-5 |
| Activation | Softmax |
| Training : Validation | 9 : 1 |

予備実験では、各感情カテゴリに対して二値分類の学習を行った。各カテゴリの学習曲線は、図 2.1.1, 2.1.2, 2.1.3, 2.1.4 に示している。それらの学習曲線から、学習データでの精度上昇に関わらず、検証データとテストデータでの精度が著しく上昇しないことが判った。その理由として、ALBERT モデルの過学習の傾向が強く、データのパターンが明確でないムード分類には不向きである可能性が考えられる。

Accuracy of Binary Classification 'Happy'

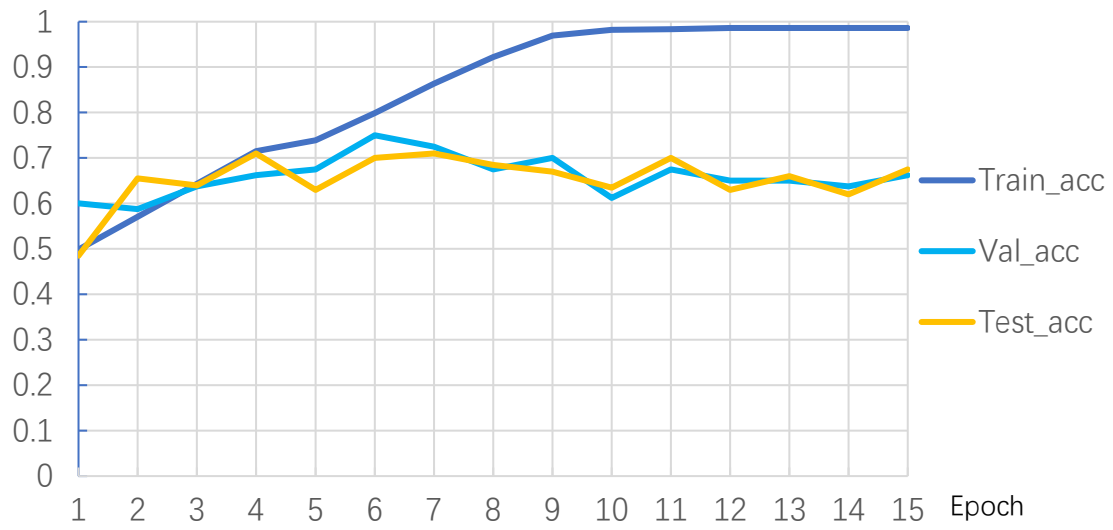


図 2.1.1 二値分類の学習曲線 (Happy)

Accuracy of Binary Classification 'Angry'

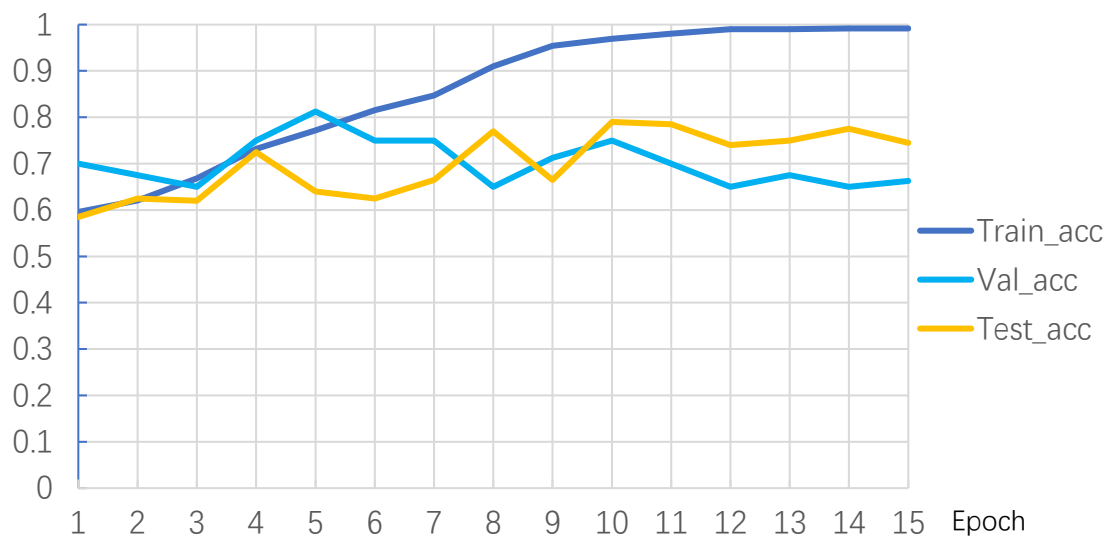


図 2.1.2 二値分類の学習曲線 (Angry)

Accuracy of Binary Classification 'Sad'

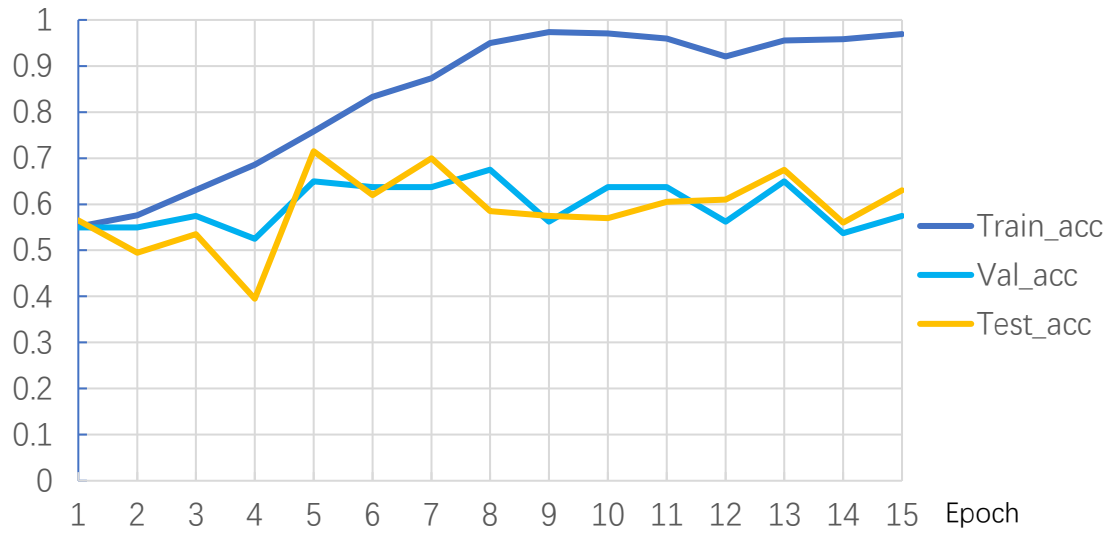


図 2.1.3 二値分類の学習曲線 (Sad)

Accuracy of Binary Classification 'Relaxed'

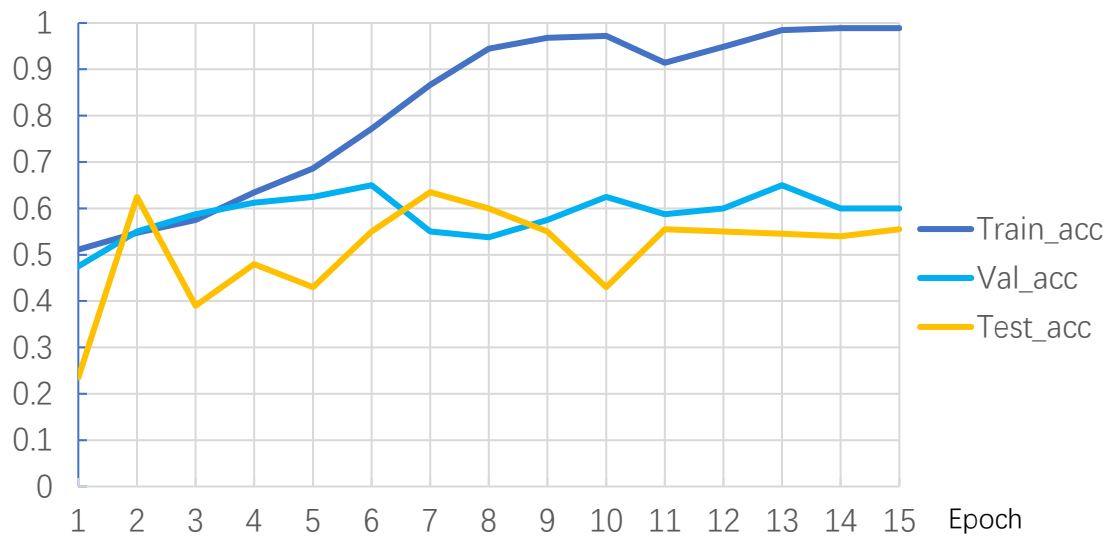


図 2.1.4 二値分類の学習曲線 (Relaxed)

Final Accuracy on Validation data & Test data

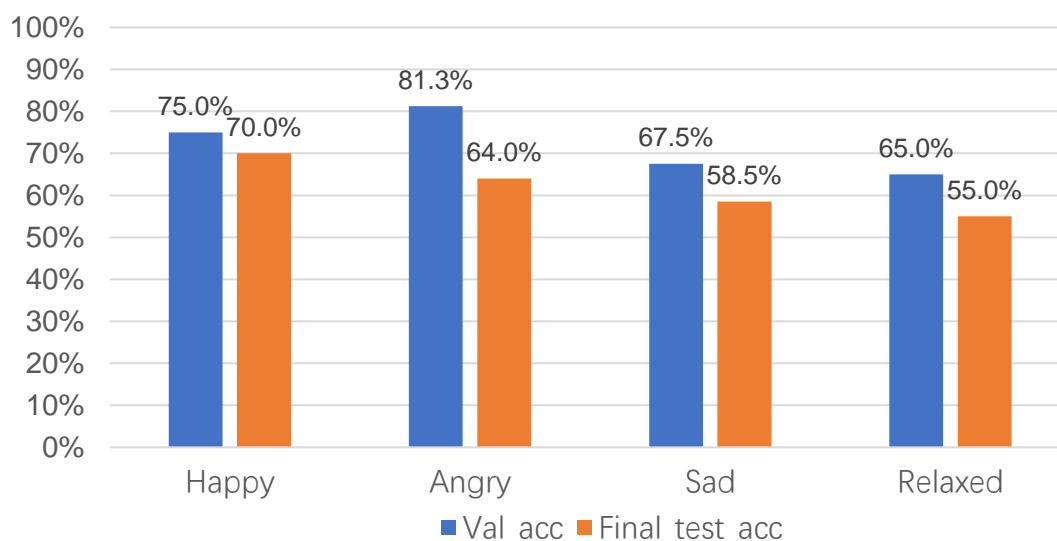


図 2.1.5 ALBERT による分類精度

検証データで精度が一番高かった分類モデルを選出し、テストデータを分類した結果は図 2.1.5 に示している。“Happy” のカテゴリが 70.0% の正解率に達しているが、それに対し“Sad” と “Relaxed” での精度が不足になっている。以上の予備実験により、事前学習モデルである ALBERT モデルを直接に適用する分類方法では、実現したムード分類精度を向上させる必要があると判った。

2.2 日本語歌詞のムード分類

歌詞情報を使用した日本語歌詞楽曲のムード分類は、英語歌詞楽曲と同様に単語の出現頻度を用いた手法から発展している。杉浦ら [7] は、独自の音楽カテゴリに沿って 5922 語の辞書を作成した。彼らは分類対象の歌詞に含まれる単語を抽出し、それを辞書データベースとの照合することにより、カテゴリ別の出現頻度数を得た。そして各単語に対し、カテゴリ別の出現頻度数と分類対象の歌詞に出現する頻度を利用してスコアの算出を行った。表 2.2.1 には、杉浦らの実験結果を示している。彼らは“海で聴きたい” など特定のカテゴリで高い再現率と適合率を得られているが、カテゴリの構築自体がムードや雰囲気よりトッピングの方に類似している。

表 2.2.1 杉浦らの実験結果(文献[7]より引用)

| カテゴリ | 再現率 | 適合率 |
|--------------|-----|------|
| 海で聴きたい曲 | 0.8 | 0.8 |
| スキー(雪)で聴きたい曲 | 0.9 | 0.34 |
| 入学で聴きたい曲 | 0.7 | 0.7 |
| 卒業で聴きたい曲 | 0.2 | 0.2 |
| クリスマスで聴きたい曲 | 0.5 | 0.35 |

2.3 感情単語リストを使用した分類方法

古屋ら[8]は、感情単語リストを利用する分類手法を提案している。彼らの研究では、歌詞に存在する感情単語の出現頻度を中心に作成した特徴ベクトルのクラスタリングが行われている。評価方法については、10人の被験者に合計50曲のクラスタリング対象楽曲を聴取させ、10の感情カテゴリから選択した楽曲聴取時の印象を正解集合としている。ただし、評価実験では半数の感情カテゴリがほとんど選択されていないため、実際分類に使用したカテゴリが5種類となっている。また、彼らも本論文の1.1節に述べた感情用語のみを解析する問題点を察知し、その対策として類似語辞書及び単語間類似度の算出により感情単語リストの拡張を行った。その結果、拡張した感情単語リストを使用する場合の分類精度は拡張しない場合より向上した。

表 2.3.1 古屋らの実験結果（文献[8]より引用）

| 対象楽曲の言語 | 再現率 | 適合率 | F 値 |
|---------|-------|-------|-------|
| 日本語 | 46.3% | 41.3% | 43.7% |
| 英語 | 26.7% | 21.6% | 23.9% |

2.4 感情単語ベクトルと歌詞ベクトルの関連性と予備実験

既存研究では、感情単語の類似語による感情リストの拡張を行っているが、依然として特徴ベクトルの計算が単語の出現頻度に基づいている。本論文では、感情単語と歌詞テキストの類似度に基づいた特徴ベクトルの作成方法を提案しているが、ベクトル化モデルによる感情単語と文書の類似度計算に関する先行研究が少ない。そこで本節では、本研究の予備実験とした感情単語ベクトルと歌詞ベクトルの関連性検証について記述する。

まず、実験で使用されているベクトル化モデルについて説明する。類似度の計算に使用するため、感情単語ベクトルと歌詞テキストのベクトルは同一のベクトル化モデルによって出力される必要がある。そこで本実験は、類似度計算に特化した学習済みの日本語 SentenceBERT [9]を使用する。SentenceBERT では、512 語までの入力をベクトルに変換することが可能であり、それによって出力されるベクトルの次元数は固定的な 768 次元である。

次に、日本語 SentenceBERT により出力された感情単語ベクトルの分布の合理性について検証する。Russell の円環モデルを参考に、16 語の感情単語を SentenceBERT でベクトル化する。さらに、UMAP で 3 次元まで削減する結果、ベクトル空間の X 軸マイナス方向からの射影が図 2.4.1 に示されている。それを Russell の円環モデルにある二次元感情分布と照合することにより、生成したベクトルが円環モデルに類似した感情分布を持っていると考えられる。

最後に、出力した感情単語ベクトルと sentence レベルの文書ベクトルのコサイン類似度を計算し、最も類似している感情を探す。ここでは、例文として使用されている sentence 単位の文書が sentence 単位で分割された歌詞と同じような性質を持つと仮定する。表 2.4.1 には、解析された類似度順位を上位の 8 位まで表示している。

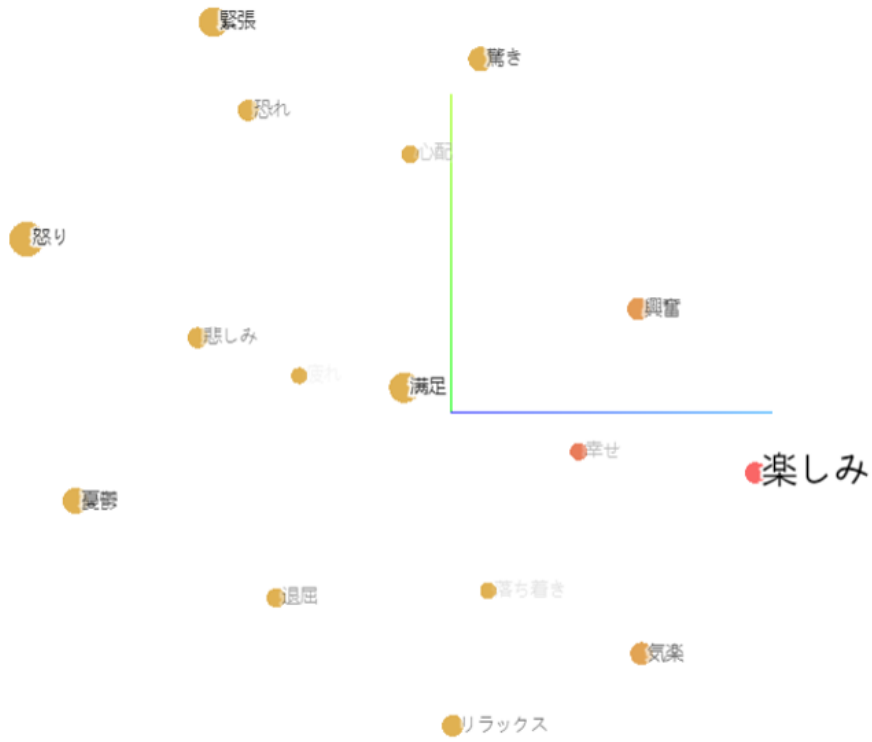


図 2.4.1 感情単語のベクトル空間

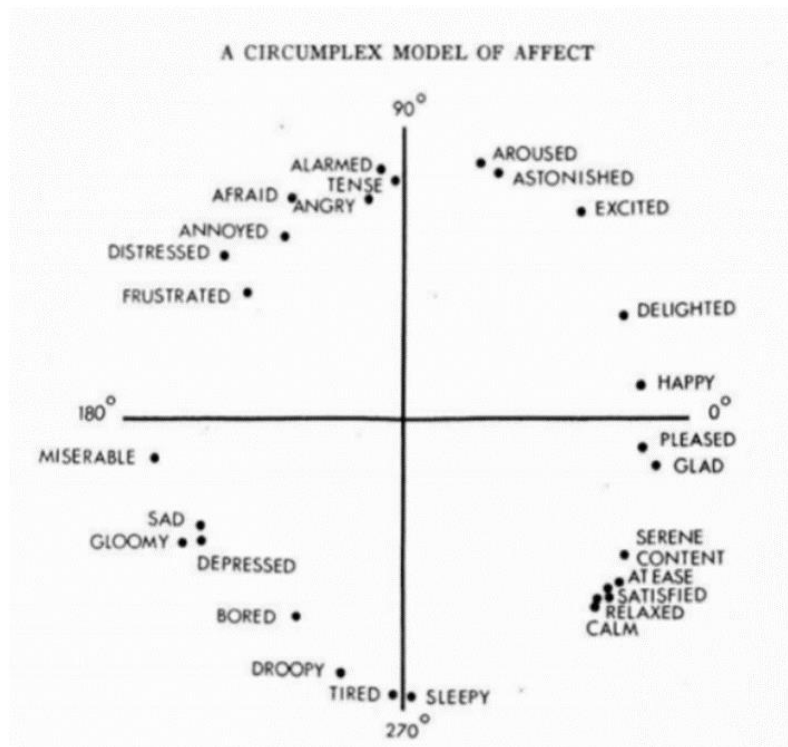


図 2.4.2 Russell の円環モデル[10]

表 2.4.1 例文における類似度順位

| 類似度順位 | 例文 A : ぬいぐるみを失くして一晩中泣いていた | 例文 B : 誕生日プレゼントをたくさんもらって嬉しい |
|-------|---------------------------|-----------------------------|
| 1 | 悲しみ | 幸せ |
| 2 | 憂鬱 | 興奮 |
| 3 | 退屈 | 楽しみ |
| 4 | 疲れ | 気楽 |
| 5 | 怒り | 驚き |
| 6 | 緊張 | 心配 |
| 7 | 恐れ | 疲れ |
| 8 | 心配 | リラックス |

類似度順位の解析から、ネガティブな感情を持っていると想定された例文 A には“悲しみ”，“憂鬱”などの感情単語が最も類似していると判断され，ポジティブな感情を持っていると想定された例文 B には“幸せ”，“興奮”などの感情単語が最も類似しているという矛盾しない結果が得られた．このような検証により，感情単語ベクトルと歌詞ベクトルのコサイン類似度を感情単語と歌詞テキストの関連度とした解析が可能であると判明した．故に本研究は，SentenceBERT モデルによる関連度解析を基本的な手法として用いることにした．

第3章 実験

3.1 実験概要

実験では、クローラで取得した歌詞データに対して前処理を行う。日本語と日本語以外の文字の割合を計算し、解析対象外のデータを除外する。最終的に、5115本の歌詞データを実験に使用した。感情単語リストの作成には、10種類の感情から計105語を選出した。その後、日本語 SentenceBERT で得られた歌詞ベクトルと各感情単語ベクトルのコサイン類似度から、105次元の特徴ベクトルを作成する。算出された特徴ベクトルに対し、k-means 法でクラスタリングを行う。クラスタリングによって取得したクラスタ番号が正解ラベルであると仮定し、Random Forest によって歌詞の分類を学習させる。検証には、オンラインで実施した主観評価アンケートの回答から、60曲分の正解集合を作成した。

3.2 学習データ

本実験の学習データは、NetEase Cloud Music の Web Service[11]から取得した。取得したメタデータには、楽曲のタイトル、URL、タグ、歌詞情報などが含まれている。前処理を通じて、9264本のメタデータから、5115本の歌詞データをクラスタリング実験の学習データとして抽出した。前処理の詳細については、本論文の4.1節で説明する。学習データには、一行の歌詞を一つの sentence として解析している。

3.3 アンケート

本実験の検証には、主観評価アンケートの結果を使用する。アンケートはオンラインで実施され、回答者の総人数は148名である。回答者のうち男性は82人、女性は66人である。アンケートでは、学習用の5115曲とは異なった合計60曲の歌詞に関する質問を設けている。著作権保護のため、アンケートは回答者にリンク先[12]に掲載された歌詞を読ませる形式となっている。回答者は歌詞を閲覧した後、歌詞から受ける感情について評価する。

アンケートに使用する感情は Russell の円環モデルを参考に、感情価 (Valence: 悲しい-楽しい) と覚醒度 (Arousal: 落ち着く-興奮/緊張) を9段階で評価するものとした。60曲に対し、回答者全体の平均点を正解ラベルとして扱う。

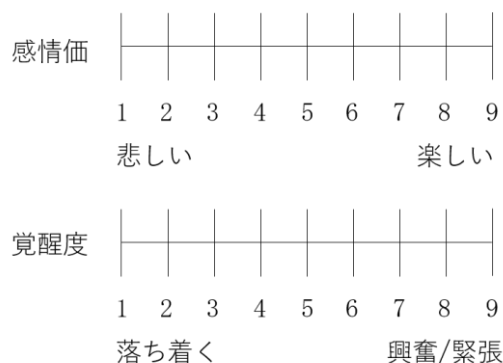


図 3.3.1 アンケートに使用する感情指標

リンク先の歌詞を読んで、その歌詞の感情について評価してください。(猫 : 1/20)
<https://utaten.com/lyric/ya17081607/>

その歌詞が表している楽しさの程度は？ *

1 2 3 4 5 6 7 8 9

悲しい ○ ○ ○ ○ ○ ○ ○ ○ ○ 楽しい

その歌詞が表している緊張または興奮の程度は？ *

1 2 3 4 5 6 7 8 9

落ち着く ○ ○ ○ ○ ○ ○ ○ ○ ○ 興奮または緊張

図 3.3.2 アンケート項目の一例

アンケートの結果は表 3.3.1 に示している. ここでは, 感情価を **V** として表記し, 覚醒度を **A** として表記する. また, 評価の平均点数が 5 未満の場合は後ろに “-” を付け, 5 以上の場合は “+” を付ける.

表 3.3.1 アンケート結果

| 感情カテゴリ | 楽曲数 |
|------------------|-----|
| V+A+ (楽しいかつ興奮) | 19 |
| V-A+ (悲しいかつ緊張) | 11 |
| V-A- (悲しいかつ落ち着く) | 23 |
| V+A- (楽しいかつ落ち着く) | 7 |

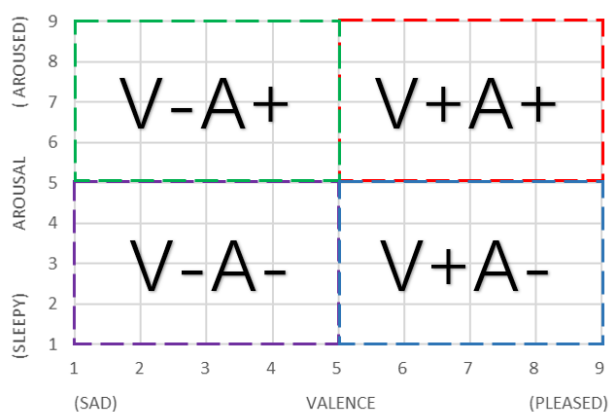


図 3.3.3 アンケート評価の感情カテゴリ

3.4 検証方法

検証では、まず、主観評価を行った 60 曲に対して、それぞれの特徴ベクトルを算出する。また、特徴ベクトルの計算方法については、本論文の 4.4 節で説明する。次に、算出された特徴ベクトルを **Random Forest** のモデルに入力し、クラスタ番号を取得する。その後、感情価と覚醒度を 2 次元とする空間に、正解としたそれぞれの値に対して、出力されたクラスタ番号を図示し、データ分布を調べることで検証を行う。

3.5 実験手順

前処理を行った 5115 本の歌詞データに対し、学習済みの日本語 **SentenceBERT** モデルでベクトル化する。次に、別途作成した感情単語リストを同一のモデルでベクトル化を行い、歌詞ベクトルと各感情単語の類似度を計算することで特徴ベクトルを作成する。

作成した特徴ベクトルに対し、**k-means** 法でクラスタリング実験を行う。その後、クラスタリングで取得したクラスタ番号が正解ラベルであると仮定し、**Random Forest** によってクラスタリング結果を学習する。

最後に、3.4 節で説明した検証方法で、実験結果の検証を行う。ただし、検証用特徴ベクトルの算出には、学習用特徴ベクトルの算出と同様な手法を使用している。

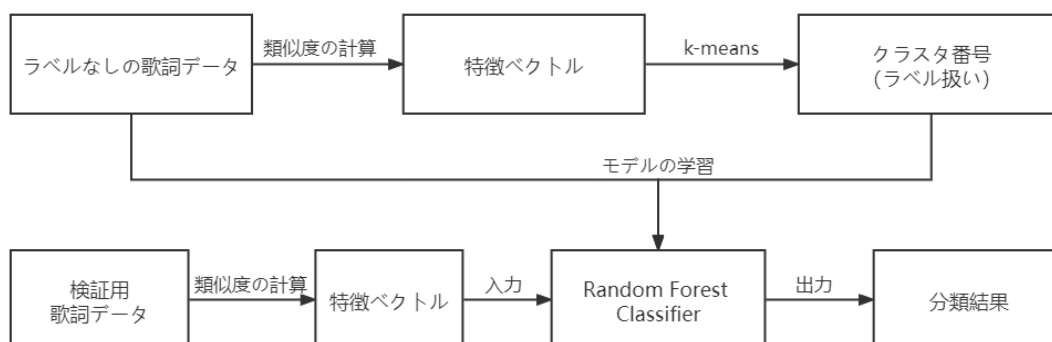


図 3.5.1 実験手順

第4章 解析手法

4.1 前処理

実験に適合する歌詞データを学習データとして抽出するために、メタデータに対して前処理を行う。

まずは、楽曲のタイトルを参考に、重複したデータを取り除く。次に、メタデータにある歌詞の有無に関する属性の値を読み取り、`instrumental` の楽曲を除外する。残った歌詞ありのメタデータに対し、訳語のデータが存在する場合はその部分を削除し、元の言語で書かれている歌詞のみを解析対象とする。その後、歌詞以外の情報を取り除き、歌詞情報のみを抽出する。

抽出された歌詞情報は、`LRC` フォーマットで登録されている。`LRC` フォーマットの歌詞には、楽曲での出現時間が該当の歌詞と共に記述されている。解析に影響しないため、そのような時間情報のパターンを正規表現で検知し、出現時間の削除を行った。句読点などに対しては、既存の句読点や記号をスペースに置換し、行ごとに区切りとして句点を付ける。さらに、このような処理を行ったテキストに対し、日本語文字と日本語以外の文字の割合を計算する。本実験では、英語などの他言語が全体の **33.3%**以上占めている歌詞を解析対象外とした。

このような前処理を行うことで、**9264**本のメタデータから**5115**本の歌詞データを抽出し、クラスタリング実験に使用した。

4.2 感情単語リスト

歌詞データと各感情の類似度を計算するため、各感情カテゴリに所属する感情単語を含める感情単語リストを作成する。

4.2.1 作成方法

感情単語リストの作成には、中村明氏の感情表現辞典[13]を使用する。感情に関する類似度計算に使用するため、本実験では可能な限りに感情面以外に同一な性質を持つ単語を選出している。選択基準については、作成用の感情単語を1単語で表現されている和語の形容詞と動詞に限定する。ただし、複合動詞、擬声語、擬態語等は使用しない。

このような条件に基づき、“喜”、“怒”、“哀”、“怖”、“恥”、“好”、“厭”、“昂”、“安”、“驚”の10種類の感情から計**105**語を選出した。

4.2.2 バランス調整

前項で述べた作成方法により作成した**105**語の感情単語リストでは、カテゴリ間に単語数の差が生じている。最大のカテゴリ“厭”には**26**の感情単語が含まれているが、それに対して最小のカテゴリ“驚”には**3**つの感情単語しか含まれていない。そこで本実験では、感情単語数のバランスを調整し、新たに合計**68**語の感情単語リストを作成した。調整前後の各感情カテゴリの単語数は、表**4.2.1**に示している。また、バランス調整の影響については、本論文の**5.2**節で結果と考察を記述する。

表 4.2.1 感情単語リストのバランス調整

| 感情カテゴリ | 喜 | 怒 | 哀 | 怖 | 恥 | 好 | 厭 | 昂 | 安 | 驚 | 合計 |
|--------------|----|---|----|---|---|----|----|---|---|---|-----|
| バランス調整なしの単語数 | 14 | 7 | 15 | 5 | 6 | 17 | 26 | 7 | 5 | 3 | 105 |
| バランス調整ありの単語数 | 10 | 5 | 10 | 5 | 5 | 10 | 10 | 5 | 5 | 3 | 68 |

4.3 ベクトル化

感情単語に対しては、直接に単語の状態ですentenceBERTモデルに入力し、感情単語ベクトルを得ることが可能である。しかし、歌詞には場合によって文字数が過多になり、モデル入力上限の512語を超えることがある。そこで本実験では、歌詞に対して行ごとにsentenceレベルのベクトルを生成する。歌詞と感情単語の類似度を計算する際には、該当する歌詞の全ての行に対して感情単語との類似度を算出し、それに基づいて特徴ベクトルの作成を行う。

4.4 特徴ベクトルの作成

4.4.1 セグメント化なしの場合

特徴ベクトルの作成には、まず行ごとに歌詞の感情特徴量を算出する。感情特徴量は、歌詞ベクトルと感情単語ベクトルのコサイン類似度のスコア S_c から構築される。スコア S_c は以下の算式で算出されている。

$$S_c(u, v) = \frac{1}{2} \left(1 - \frac{u \cdot v}{\|u\| \|v\|} \right) \quad (4.1)$$

ここでは、 u 、 v はそれぞれ sentence ベクトルと感情単語のベクトルを表している。感情単語リストに含まれる単語数に応じて、感情特徴量の次元数が異なることがある。本実験では、バランス調整なしの場合の感情特徴量は105次元であり、バランス調整ありの場合の感情特徴量は68次元である。

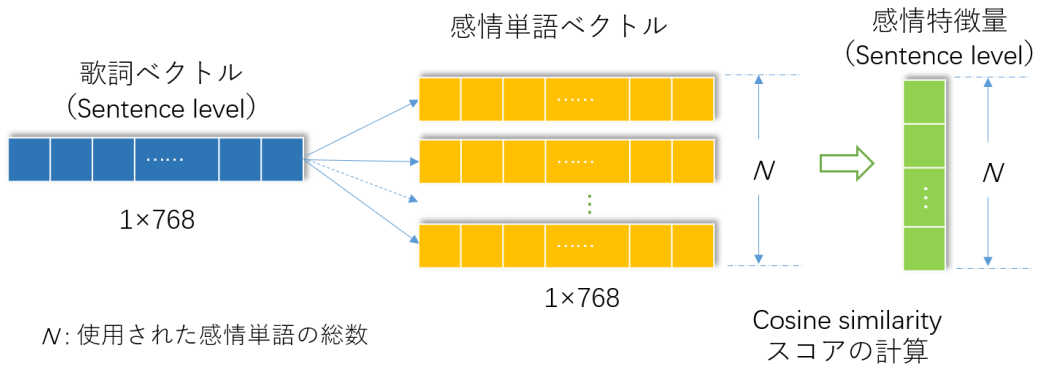


図 4.4.1 感情特徴量の作成

セグメント化を行わない場合には、一曲分の歌詞に含まれる全ての行に対して感情特徴量を算出し、それらの平均値を該当する歌詞の特徴ベクトルとする。

4.4.2 セグメント化ありの場合

前項で述べたセグメント化なしの手法は、歌詞全体感情特徴を着目し、平均的な特徴量を算出している。しかし、楽曲には感情の変化が存在している。そこで本実験は、歌詞に於いて各部分の感情を解析するため、もう一つの解析パターンとして歌詞のセグメント化を追加する。解析対象とした歌詞を若干のセグメントに分割し、それぞれのセグメントに対して前項の手法で感情特徴量の平均値を算出する。その後、得られた平均値ベクトルをセグメント順に結合し、新たな特徴ベクトルを作成する。

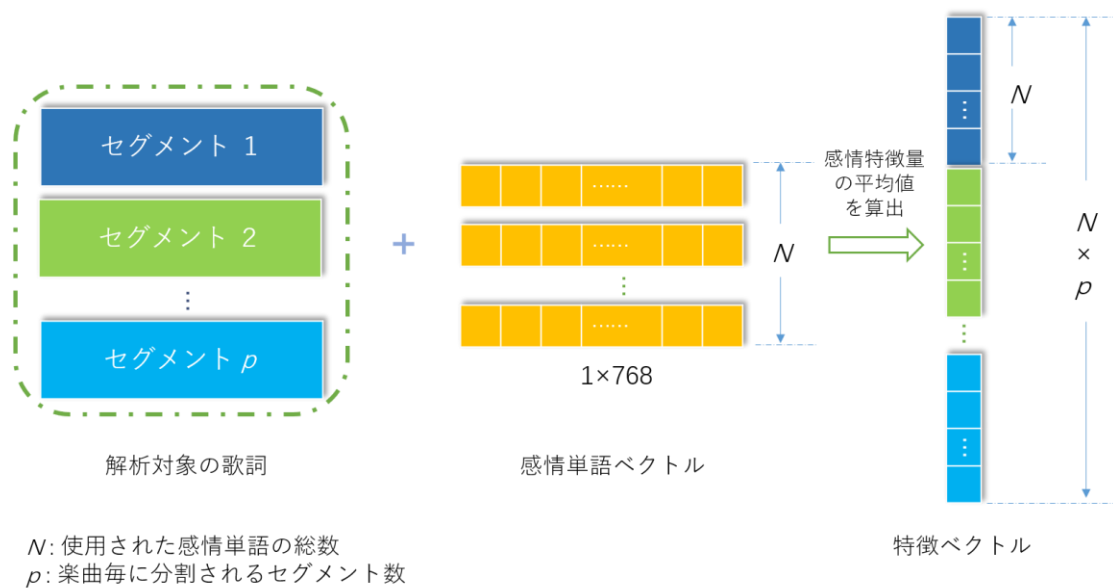


図 4.4.2 特徴ベクトルの作成 (セグメント化あり)

図 4.4.2 には、セグメント化を追加した特徴ベクトルの作成手法を示している。ここでは、 N を解析に使用された感情単語数、 p を楽曲毎に分割されるセグメント数を表している。セグメント化ありの場合には、作成した特徴ベクトルの次元数が $N \times p$ となる。

4.5 クラスタリング

前節の手法で作成した特徴ベクトルを学習データとして、**k-means** 法でクラスタリングを行う。実験に設定するクラスタ数に関しては、**SSE** 曲線が比較的平滑であるため、エルボー法では明確な最適値が判明できなかった。図 4.5.1 には、“セグメント化なし”かつ“感情単語のバランス調整なし”の実験条件での **SSE** 曲線を示している。本実験では、 k の値を 2, 4, 6, 8 にした結果をそれぞれに解析し、最終的に $k = 4$ とした。

クラスタリングを行った結果、5115 曲分の学習データのクラスタ番号が得られた。それらの番号を用いて、続きに分類器の学習を行う。

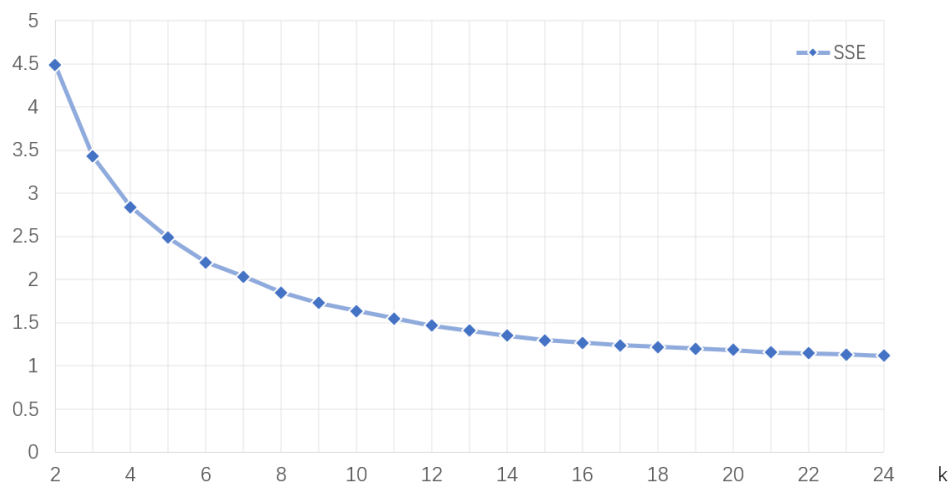


図 4.5.1 SSE 曲線

4.6 分類手法

クラスタリングの結果を学習するには、**Random Forest Classifier** を使用している。クラスタリング実験により取得したクラスタ番号は、正解ラベルとして **Random Forest** に入力される。学習のパラメーターについては、表 4.6.1 の設定値をデフォルトパラメーターとして使用している。

表 4.6.1 Random Forest Classifier のパラメーター

| パラメーター | 設定値 |
|--------------------|-----------|
| n_estimators | 64 |
| criterion | 'entropy' |
| max_depth | 20 |
| Train : Validation | 7 : 3 |

第5章 実験結果及び考察

5.1 セグメント化有無の比較

本節では、4.4節で述べた2つの特徴ベクトルの作成手法による実験結果を示し、考察を行う。

5.1.1 セグメント化なしの実験結果

図5.1.1には、セグメント化なしの場合の実験結果を示している。検証曲中では、クラスタ2に分類されたものはなかった。また、クラスタ0については、特定の感情カテゴリに属する傾向が判明できなかった。

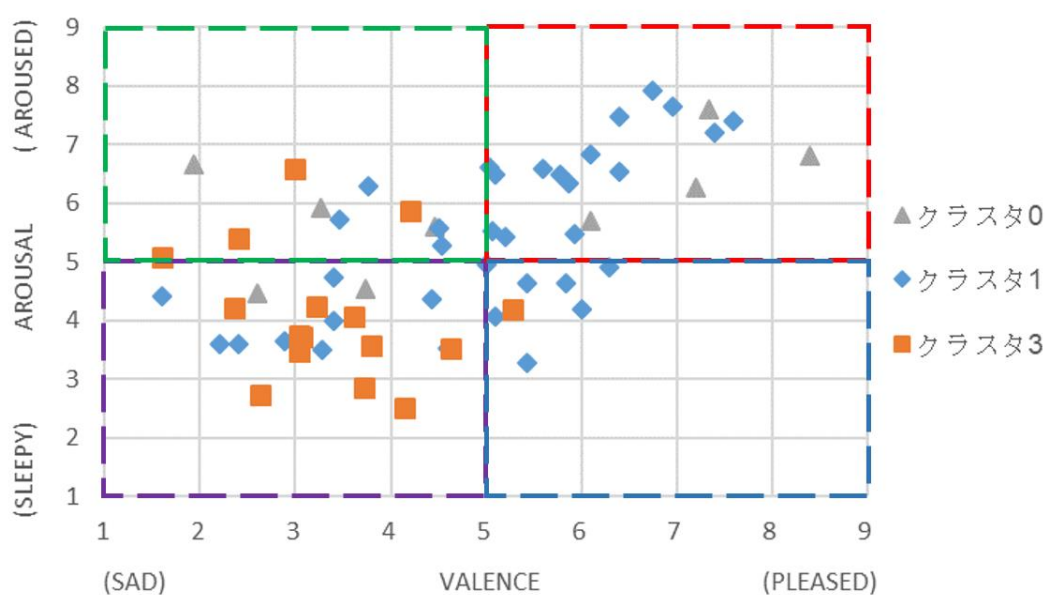


図 5.1.1 結果分布（セグメント化なし）

表5.1.1には、クラスタ1が“V+A+（図中の赤の領域に対応）”，クラスタ3が“V-A-（図中の紫の領域に対応）”，クラスタ0が“その他の感情（図中の青と緑の領域に対応）”と仮定した場合の分類精度を示している。算出した結果から、クラスタ2とクラスタ3では比較的に高い精度が得られ、クラスタ0は適切に分類されていないことがわかった。

表 5.1.1 分類精度（セグメント化なし）

| カテゴリ | 再現率 | 適合率 | F 値 |
|------|-------|-------|-------|
| V+A+ | 78.9% | 42.9% | 55.6% |
| V-A- | 47.8% | 68.8% | 56.4% |
| その他 | 16.7% | 33.3% | 22.2% |

5.1.2 セグメント化ありの実験結果

図 5.1.2 には、セグメント化ありの場合の実験結果を示している。セグメント化なしの結果と比べれば、“V-A+”の領域で2箇所の分類結果に変更があった。

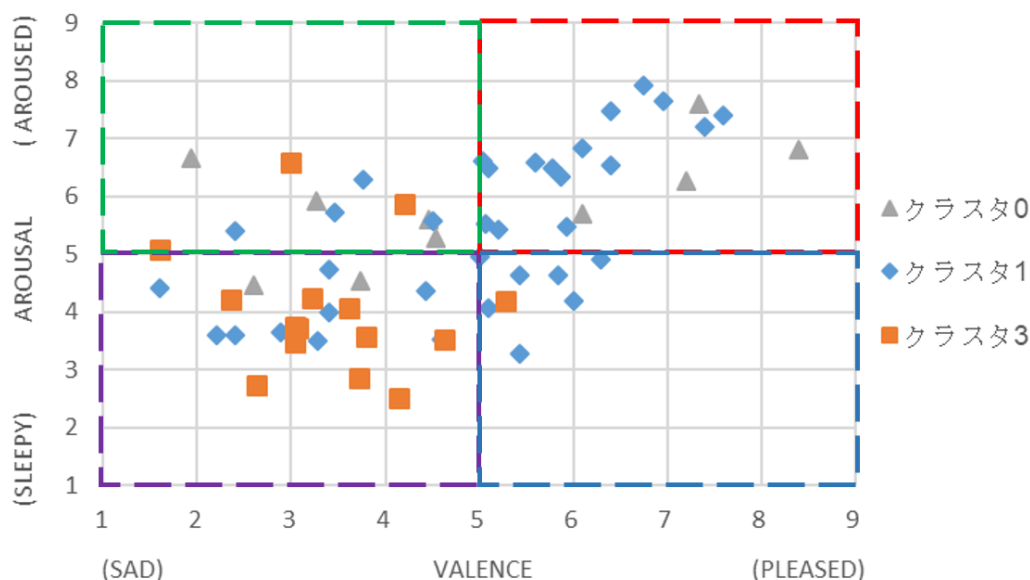


図 5.1.2 結果分布（セグメント化あり）

表 5.1.2 にも表 5.1.1 と同様に、クラスタ 1 が“V+A+”，クラスタ 3 が“V-A-”，クラスタ 0 が“その他の感情”と仮定した場合の分類精度を示している。

表 5.1.2 セグメント化ありの分類精度

| カテゴリ | 再現率 | 適合率 | F 値 |
|------|-------|-------|-------|
| V+A+ | 78.9% | 42.9% | 55.6% |
| V-A- | 47.8% | 73.3% | 57.9% |
| その他 | 22.2% | 40.0% | 28.6% |

5.1.3 考察

検証に使用する 60 曲の歌詞データのうち、58 曲の分類結果に対してセグメント化の影響が見られなかった。分類結果のクラスタが異なった 2 曲には、誤ったクラスタから正確なクラスタへ変化したものが 1 曲あった。その結果、セグメント化ありの場合に、“V-A-”と“その他の感情”の分類精度がセグメント化なしの場合より向上した。

また、ほとんどのデータに対してクラスタの変化がないということから、感情類似度による特徴を用いる分類手法で取得した結果は比較的安定していると示唆された。

5.2 感情単語のバランスを調整した解析

本節では、4.2節で述べた感情単語数のバランス調整を実行する。さらに、調整による実験結果を示し、特徴量の重要度に踏まえて考察を行う。

5.2.1 実験結果

図 5.2.1 には、“セグメント化なし”かつ“バランス調整あり”の場合の実験結果を示している。依然として、クラスタ 2 に分類されたものはなかった。また、“V-A+”の領域では、5.1.1 項の結果と比べて、5.1.2 項の 2 箇所とは異なった 1 箇所の分類結果に変更があった。

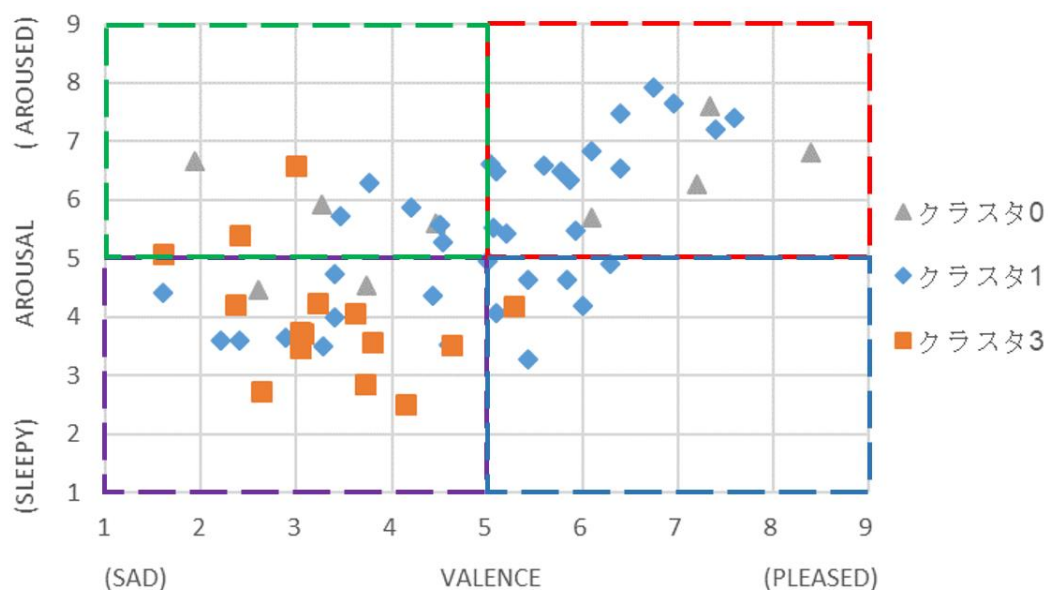


図 5.2.1 結果分布 (セグメント化なし&バランス調整あり)

表 5.2.1 には、前節と同様な条件に基づき、“セグメント化なし”かつ“バランス調整あり”の場合の分類精度を示している。

表 5.2.1 分類精度 (セグメント化なし&バランス調整あり)

| カテゴリ | 再現率 | 適合率 | F 値 |
|------|-------|-------|-------|
| V+A+ | 78.9% | 41.7% | 54.5% |
| V-A- | 47.8% | 73.3% | 57.9% |
| その他 | 16.7% | 33.3% | 22.2% |

また、セグメント化ありの実験条件に対し、バランス調整を追加した検証も行った。その結果、バランス調整を行わない場合と一致している分類結果を得た。

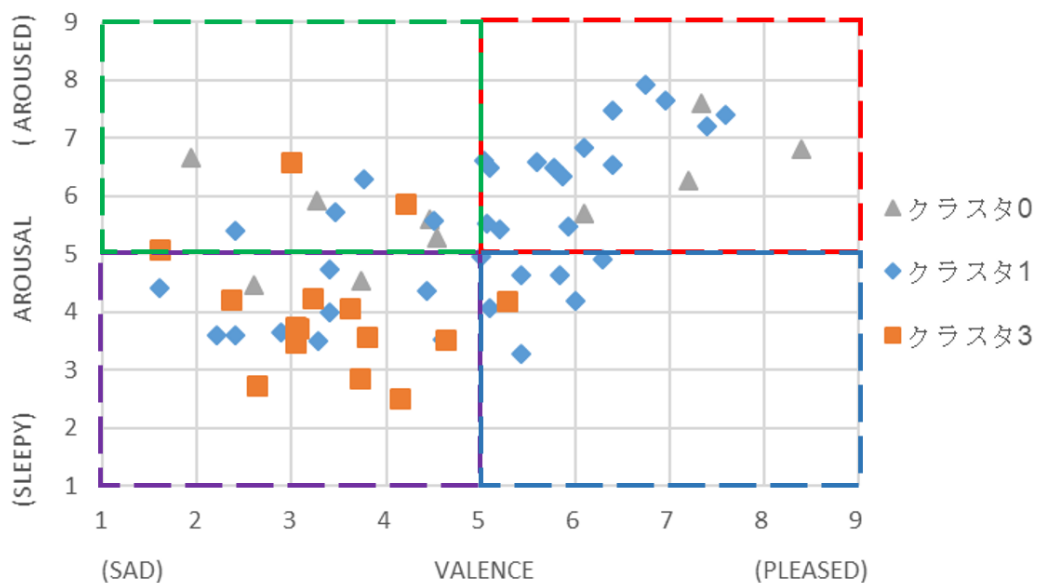


図 5.2.2 結果分布 (セグメント化あり&バランス調整あり)

5.2.2 特徴量の重要度

本実験に使用されている機械学習ライブラリ `scikit-learn` では、ジニ不純度に基づいて特徴量の重要度 (Gini importance) を算出している [14]。セグメント化なしの実験条件下で、感情単語リストの二種類のパターンに対して特徴量の重要度を調べた。

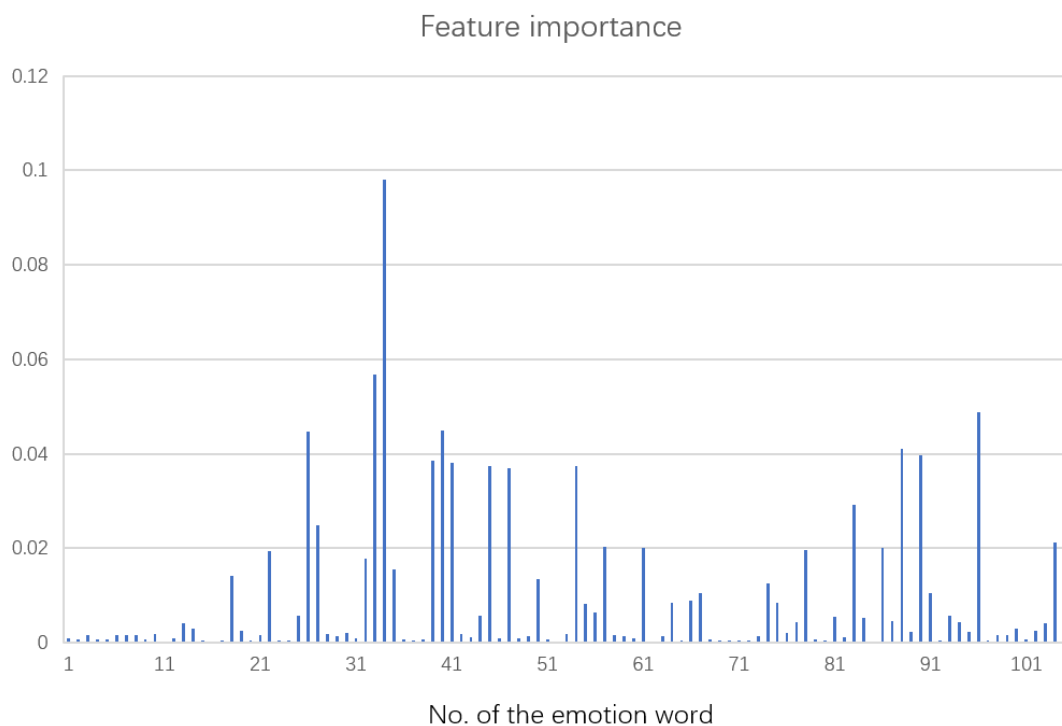


図 5.2.3 特徴量の重要度 (バランス調整なし)

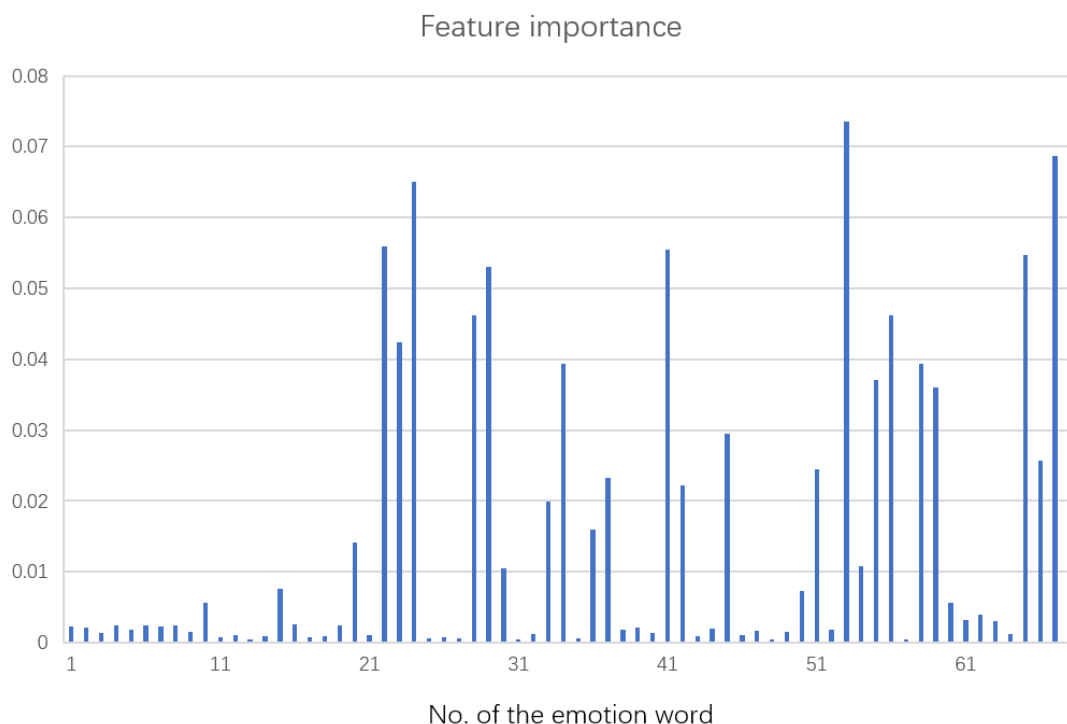


図 5.2.4 特徴量の重要度（バランス調整あり）

図 5.2.3 と図 5.2.4 には、それぞれ“バランス調整なし”と“バランス調整あり”の条件下で取得した特徴量の重要度を示している。また、表 5.2.2 と表 5.2.3 には、特徴量の重要度順位に関する情報を上位の十位まで示している。ここでは、バランス調整の有無に関わらず、両方とも高い重要度を示した感情単語を赤字で表している。

表 5.2.2 重要度の順位（バランス調整なし）

| 重要度順位 | 感情単語 | 感情カテゴリ | 特徴量の重要度 |
|-------|------|--------|----------|
| 1 | 冷たい | 哀 | 0.097996 |
| 2 | 虚しい | 哀 | 0.056841 |
| 3 | 打たれる | 昂 | 0.048830 |
| 4 | 戦く | 怖 | 0.044842 |
| 5 | 咽ぶ | 哀 | 0.044629 |
| 6 | 悩む | 厭 | 0.041018 |
| 7 | 苦しい | 厭 | 0.039721 |
| 8 | 怯える | 怖 | 0.038440 |
| 9 | 怯む | 怖 | 0.038073 |
| 10 | 赤らむ | 恥 | 0.037424 |

表 5.2.3 重要度の順位 (バランス調整あり)

| 重要度順位 | 感情単語 | 感情カテゴリ | 特徴量の重要度 |
|-------|------|--------|----------|
| 1 | 悩む | 厭 | 0.073618 |
| 2 | 慌てる | 驚 | 0.068673 |
| 3 | 凍る | 哀 | 0.065090 |
| 4 | 虚しい | 哀 | 0.055869 |
| 5 | 渴く | 好 | 0.055418 |
| 6 | 鎮める | 安 | 0.054705 |
| 7 | 戦く | 怖 | 0.053086 |
| 8 | 焦る | 昂 | 0.046206 |
| 9 | 怯える | 怖 | 0.046180 |
| 10 | 冷たい | 哀 | 0.042460 |

さらに、各感情カテゴリの重要度を調べるため、感情表現辞典の各感情カテゴリに抽出した感情単語に対して、平均重要度の算出を行った。算出した結果は、表 5.2.4 に示している。

表 5.2.4 感情カテゴリ別の平均重要度

| 感情カテゴリ | 特徴量の平均重要度 (バランス調整なし) | 特徴量の平均重要度 (バランス調整あり) |
|--------|-------------------------|-------------------------|
| 喜 | 0.001466 | 0.002450 |
| 怒 | 0.002853 | 0.002193 |
| 哀 | 0.019365 | 0.018588 |
| 怖 | 0.024522 | 0.022236 |
| 恥 | 0.013983 | 0.012344 |
| 好 | 0.007357 | 0.015453 |
| 厭 | 0.008514 | 0.015998 |
| 昂 | 0.010391 | 0.025547 |
| 安 | 0.001822 | 0.013234 |
| 驚 | 0.011414 | 0.032445 |

5.2.3 考察

まず、検証結果から判断すれば、本実験での感情単語に対するバランス調整は分類精度に大きな影響を与えることがなかった。その原因については、調整後の 65 次元の特徴ベクトルが未調整の 105 次元の特徴ベクトルとほぼ同じレベルの感情情報量を持っていると推測している。特徴量の重要度を調べることにより、バランス調整前後の特徴で作成した分類器は似たような基準で分類していると示唆した。ただし、バランス調整を行うことで、行わない場合より“驚”の感情カテゴリから選出した、数少ない感情単語の特徴量重要度が大幅に向上した。感情単語の少ない感情カテゴリに対しては、バランス調整がそれらの感情を重要視する傾向があると考えられる。

さらに、感情カテゴリ別の平均重要度から、本来 Valence 軸の端にある喜びの感情が低

い重要度を持っていることが判った。また、怒りは典型的な“V-A+”のカテゴリに分類する感情と考えられているが、それに関する特徴量の重要度も低下している。同時に、解析パターンを変更する際には“V-A+”の領域で不安定な分類結果が生じている。その現象からは、“V-A+”の特徴を正確に識別できなかったことが“その他の感情”での精度低下を引き起こしているのではないかと考えている。

5.3 ノイズ除去を追加した解析

これまでの検証結果では、クラスタ 2 に分類されるデータがなかった。その問題を解決するためには、学習データに存在するクラスタ 2 の歌詞について調べた。その結果、クラスタ 2 にクラスタリングされた歌詞はノイズであることが判った。本節では、学習データでのノイズ除去について説明する。さらに、ノイズを除去した上で、クラスタ数 k を増加する結果に対して検証を行う。

5.3.1 学習データでのノイズ

最初に、実験でクラスタ 2 にクラスタリングされるノイズを発見した。それらのノイズの内容については、表 5.3.1 に示している。

表 5.3.1 ノイズの内容 (クラスタ 2)

| データ番号 | 内容 |
|-------|-------------------|
| 76 | 連続的な 10 曲分の歌詞 |
| 3346 | 意味不明な擬音語で書かれている歌詞 |
| 3575 | アニメのセリフ |
| 3576 | アニメのセリフ |
| 3577 | アニメのセリフ |
| 4235 | ラジオドラマなどの内容 |

以上のノイズは、歌詞であるものがあり、歌詞でないものも混在している。また、それらのノイズの共通点は、テキスト量が歌詞の通常テキスト量を遥かに上回っていることである。今迄の前処理では、歌詞テキストが日本語で書かれていることに着目していたが、内容についての選別を行うことがなかった。

クラスタリング実験で k の値を調整し、その他の学習データに対して再確認した結果、ローマ字、顔文字、特殊文字と記号等の混在が稀に発見されている。そこで、学習データに対し、ノイズ除去を行う。

5.3.2 ノイズ除去の手法

まず、顔文字、特殊文字と記号等に対し、個別的に削除と置換を行う。ローマ字は、英語のアルファベットで表現しているため、英語の文字数の割合をさらに制限する。テキストの内容について、歌詞と他のテキストを区別することが困難ではあるが、テキストの長さに着目し、長さが閾値未満の短いテキストデータを除外する。そして、学習データ全体の平均長さの二倍を超えたデータも解析対象外とする。

このような手法でノイズ除去を行い、学習用のテキストデータを 5115 本から 4497 本まで削減した。

5.3.3 実験結果

図 5.3.1 には, “セグメント化あり” と “バランス調整あり” の実験条件で, ノイズ除去を追加した結果を示す.

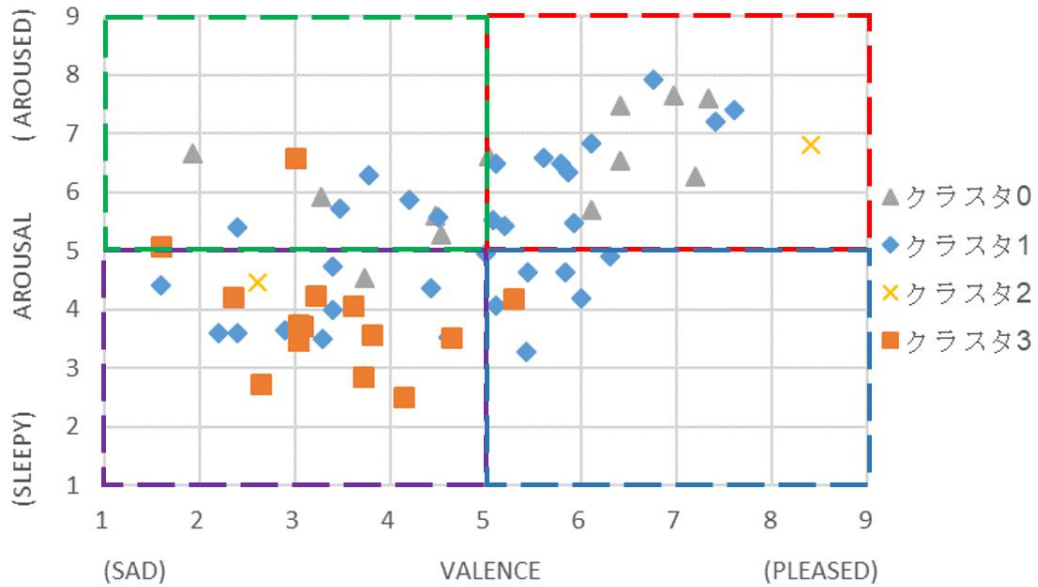


図 5.3.1 ノイズ除去の実験結果 ($k=4$)

k の値を 6 と 8 まで増加した結果は, 図 5.3.2 と図 5.3.3 に示す.

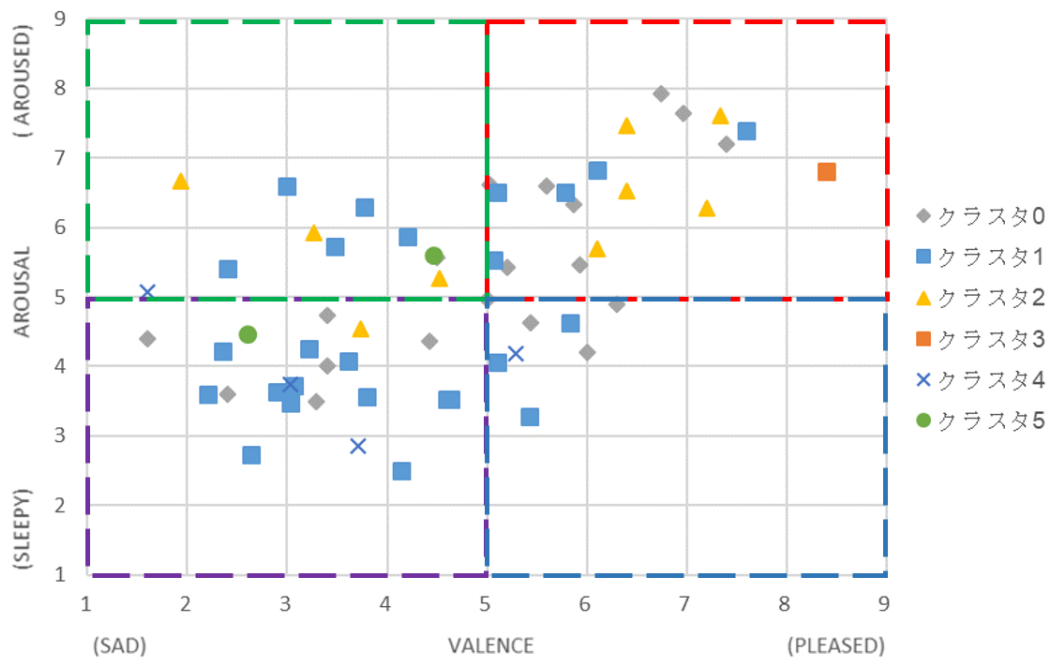


図 5.3.2 ノイズ除去の実験結果 ($k=6$)

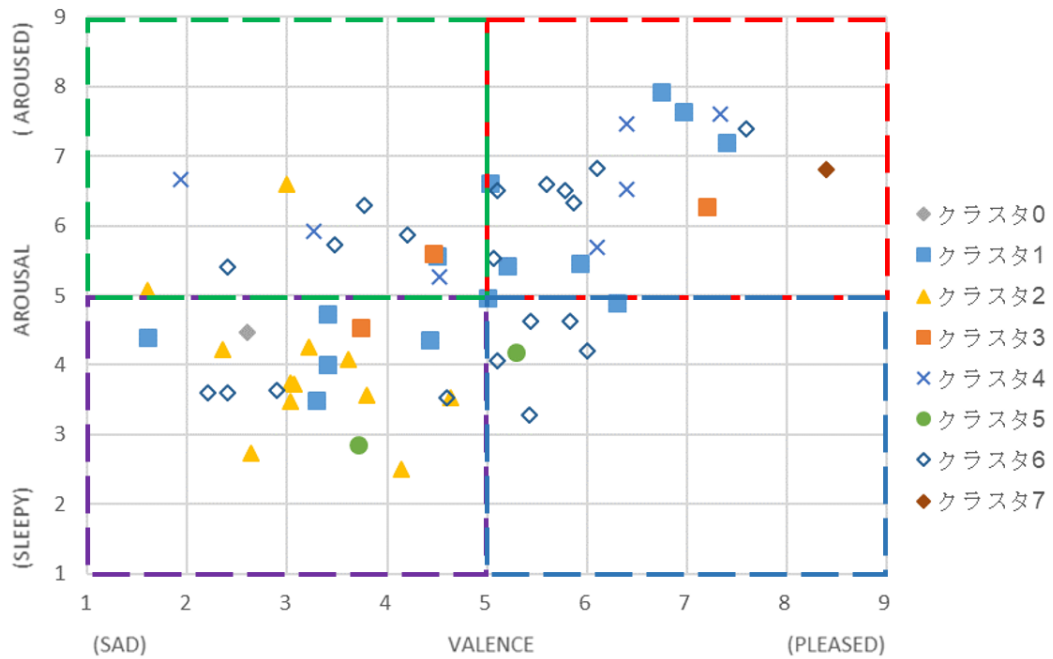


図 5.3.3 ノイズ除去の実験結果 (k=8)

5.3.4 考察

ノイズ除去を行うことで、クラスタ 2 に分類されるデータが出現した。しかし、クラスタ 2 に分類されたデータは 2 本しかなかった。また、その 2 本のデータには、明確なカテゴリ傾向がなかった。

k の値を 6 と 8 まで増加した結果、感情傾向を特定できないカテゴリも増加した。比較的に安定している分布としては、k=6 の場合のクラスタ 2 と k=8 の場合のクラスタ 4 (k=4 の場合のクラスタ 0 に対応)、k=8 の場合のクラスタ 2 (k=4 の場合のクラスタ 3 に対応) が挙げられる。また、k=4 の場合にクラスタ 2 に分類された 2 本のデータは、k の値を増加しても単独のクラスタとして現れる状況が多い。このような現象から、解析する感情カテゴリを増加する場合には解析手法の改善が必要と考える。

第 6 章 結論

6.1 まとめ

本研究では、日本語歌詞楽曲に対して、歌詞テキストと感情単語の類似度に基づき、教師なしで楽曲のムード分類を行うことを目的とした。

実験では、歌詞と感情単語リストの類似度を計算し、それによって作成した特徴ベクトルを用いることで、楽曲のムードを教師なしで分類を行った。クラスタ数を 4 に設定する場合に算出した結果から、“楽しいかつ興奮”（クラスタ 2）と“悲しいかつ落ち着く”（クラスタ 3）のカテゴリでは比較的に高い精度が得られ、“その他の感情”（クラスタ 0）は適切に分類されていないことがわかった。また、セグメント化ありの場合に、“悲しいかつ落ち着く”と“その他の感情”の分類精度がセグメント化の場合より向上していることが確認できた。さらに、バランス調整と特徴重要度の確認により、“喜”と“怒”の感情が低い重要度を持つ傾向があると判明した。クラスタ数 k の増加に関しては、解析する感情カテゴリの増加に伴い、解析手法の改善が必要であることがわかった。

6.2 今後の課題

本研究における今後の課題としては、解析する感情カテゴリの増加に伴った解析手法の改善、クラスタと感情の関係をより正確に表すための各種パラメータの最適化が挙げられる。

また、結果の検証に関して、アンケート評価で取得した検証データには、カテゴリの偏りが生じることが不可避である。その上、ラッセル円環モデルのような連続的な感情モデルに対し、隣接している感情カテゴリを区別する方法を改めて考えなければならない。特にデータ量の少ない場合には、如何なる手法でより合理的な検証を行うことが重要な課題であると考えている。

謝辞

本研究を進めるにあたり，度々ご指導をしていただいた亀山渉教授，菅沼睦客人主任研究員に心より感謝申し上げます。

また，亀山研究室の皆様には，有益なご助言をいただきました。心より感謝申し上げます。

最後に，実験とアンケートに協力してくださった実験協力者の皆様に感謝致します。

参考文献

- [1] Xiao Hu, “Improving music mood classification using lyrics, audio and social tags,” University of Illinois at Urbana-Champaign, 2010, https://www.ideals.illinois.edu/bitstream/handle/2142/18435/Hu_Xiao.pdf?..1 (2021年2月19日最終確認)
- [2] WHISSELL, C., “Dictionary of Affect in Language In PLUTCHIK AND KELLERMAN (Eds.) Emotion: Theory, Research and Experience,” Vol 4 Academic Press, NY. 1989.
- [3] Dan Yang, Won-Sook Lee, “Music Emotion Identification from Lyrics Multimedia,” ISM ’ 09. 11th IEEE International Symposium, pp. 624-629, 2009.
- [4] STONE, P.J., DUNPHY, D.C., SMITH, M.S., AND OGILVIE, D.M. (EDS.), “The General Inquirer: A computer approach to content analysis,” MIT Press, Cambridge, MA, USA, 1966.
- [5] Erion Çano, and M. Morisio . “Music Mood Dataset Creation Based on Last FM Tags,” AIAP 2017, Fourth International Conference on Artificial Intelligence and Applications, 2017.
- [6] Lan, Z. , Chen, M. , Goodman, S. , Gimpel, K. , Sharma, P. , & Soricut, R. , “Albert: a lite bert for self-supervised learning of language representations,” 2019.
- [7] 杉浦 康友, 木本 晴夫, “歌詞を用いた音楽の自動分類,” 情報科学技術フォーラム一般講演論文集 2002(2), pp. 109-110, 2002年.
- [8] 古屋 瑞生, 黄 宏軒, 川越 恭二, “歌詞情報に基づく聴取目的に応じた楽曲推薦システムの提案,” 第76回全国大会講演論文集, pp. 625-626, 2014年.
- [9] Hugging Face, <https://huggingface.co/sonois/sentence-bert-base-ja-mean-tokens> (2021年8月30日最終確認) .
- [10] Russell, J. A., “A circumplex model of affect. Journal of Personality and Social Psychology,” 39(6), pp. 1161-1178, 1980.
- [11] NetEase Cloud Music, <https://music.163.com> (2021年8月30日最終確認)
- [12] 歌詞検索サイト うたてん, <https://utaten.com> (2021年8月30日最終確認)
- [13] 感情表現辞典, 中村 明, 東京堂出版, 1979年.
- [14] scikit-learn 1.0.2 , <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier> (2021年12月21日最終確認)

図一覧

| | |
|--|----|
| 図 2.1.1 二値分類の学習曲線 (Happy) | 4 |
| 図 2.1.2 二値分類の学習曲線 (Angry) | 4 |
| 図 2.1.3 二値分類の学習曲線 (Sad) | 5 |
| 図 2.1.4 二値分類の学習曲線 (Relaxed) | 5 |
| 図 2.1.5 ALBERT による分類精度..... | 6 |
| 図 2.4.1 感情単語のベクトル空間..... | 8 |
| 図 2.4.2 Russell の円環モデル[10]..... | 8 |
| 図 3.3.1 アンケートに使用する感情指標 | 10 |
| 図 3.3.2 アンケート項目の一例..... | 11 |
| 図 3.3.3 アンケート評価の感情カテゴリ | 11 |
| 図 3.5.1 実験手順 | 12 |
| 図 4.4.1 感情特徴量の作成..... | 14 |
| 図 4.4.2 特徴ベクトルの作成 (セグメント化あり) | 15 |
| 図 4.5.1 SSE 曲線..... | 16 |
| 図 5.1.1 結果分布 (セグメント化なし) | 17 |
| 図 5.1.2 結果分布 (セグメント化あり) | 18 |
| 図 5.2.1 結果分布 (セグメント化なし&バランス調整あり) | 19 |
| 図 5.2.2 結果分布 (セグメント化あり&バランス調整あり) | 20 |
| 図 5.2.3 特徴量の重要度 (バランス調整なし) | 20 |
| 図 5.2.4 特徴量の重要度 (バランス調整あり) | 21 |
| 図 5.3.1 ノイズ除去の実験結果 (k=4) | 24 |
| 図 5.3.2 ノイズ除去の実験結果 (k=6) | 24 |
| 図 5.3.3 ノイズ除去の実験結果 (k=8) | 25 |

表一覧

| | |
|--|----|
| 表 2.1.1 ALBERT のパラメーター (Finetuning) | 3 |
| 表 2.2.1 杉浦らの実験結果 (文献[7]より引用) | 6 |
| 表 2.3.1 古屋らの実験結果 (文献[8]より引用) | 7 |
| 表 2.4.1 例文における類似度順位 | 9 |
| 表 3.3.1 アンケート結果 | 11 |
| 表 4.2.1 感情単語リストのバランス調整 | 14 |
| 表 4.6.1 Random Forest Classifier のパラメーター | 16 |
| 表 5.1.1 分類精度 (セグメント化なし) | 17 |
| 表 5.1.2 セグメント化ありの分類精度 | 18 |
| 表 5.2.1 分類精度 (セグメント化なし&バランス調整あり) | 19 |
| 表 5.2.2 重要度の順位 (バランス調整なし) | 21 |
| 表 5.2.3 重要度の順位 (バランス調整あり) | 22 |
| 表 5.2.4 感情カテゴリ別の平均重要度 | 22 |
| 表 5.3.1 ノイズの内容 (クラスタ 2) | 23 |

付録

本実験で作成した二種類の感情単語リストを表Ⅰと表Ⅱに示す。

表Ⅰ 感情単語リスト（バランス調整なし）

| 感情表現辞典でのカテゴリ | 感情単語 |
|--------------|--|
| 喜 | 喜ばしい、嬉しい、有難い、幸せ、誇らしい、楽しい、面白い、明るい、晴れる、快い、踊る、弾む、燥ぐ、笑う |
| 怒 | 怒る、憤る、苦い、震える、叱る、怒鳴る、睨む |
| 哀 | 悲しい、嘆く、泣く、啜る、咽ぶ、憐れむ、愛おしい、淋しい、侘しい、憂い、懶い、虚しい 冷たい、凍る、沈む |
| 怖 | 怖い、恐ろしい、怯える、戦く、怯む |
| 恥 | 恥ずかしい、照れる、眩しい 赤らむ、恥じる、俯く |
| 好 | 熱い、燃える、慕わしい、恋しい、愛しい、惚れる、酔う、渴く、重んじる、仰ぐ、敬う、好き、親しむ、迷う、溺れる、懐かしい、慈しむ |
| 厭 | 厭らしい、顰める、疎ましい、嫌い、嘲る、憎い、恨む、惜しい、悔しい、妬ましい、羨ましい、困る、惑う、飽きる、だれる、怠い、曇る、暗い、痛い、萎える、諦める、辛い、悶える、悩む、煩わしい、苦しい |
| 昂 | 焦る、上がる、揺れる、乱れる、昂ぶる、打たれる、褒める |
| 安 | 救われる、寛ぐ、安らぐ、冷める、鎮める |
| 驚 | 驚く、慌てる、呆れる |

表II 感情単語リスト (バランス調整あり)

| 感情表現辞典でのカテゴリ | 感情単語 |
|--------------|--|
| 喜 | 喜ばしい、嬉しい、有難い、幸せ、 誇らしい、楽しい、明るい、 快い、踊る、笑う |
| 怒 | 怒る、憤る、叱る、 怒鳴る、睨む |
| 哀 | 悲しい、嘆く、泣く、愛おしい、 淋しい、憂い、虚しい、 冷たい、凍る、沈む |
| 怖 | 怖い、恐ろしい、怯える、 戦く、怯む |
| 恥 | 恥ずかしい、照れる、眩しい 赤らむ、恥じる |
| 好 | 燃える、慕わしい、恋しい、愛しい、 惚れる、渴く、仰ぐ、 好き、親しむ、懐かしい |
| 厭 | 嫌い、憎い、惜しい、悔しい、 羨ましい、痛い、辛い、 悩む、煩わしい、苦しい |
| 昂 | 焦る、上がる、揺れる、 乱れる、昂ぶる |
| 安 | 救われる、寛ぐ、安らぐ、 冷める、鎮める |
| 驚 | 驚く、慌てる、呆れる |

研究業績

| 題目 | 発表年月 | 発表 | 連名者 |
|--|-------------------|------------------|--------------|
| 感情単語ベクトルと 歌詞ベクトルの類似度による 日本語歌詞楽曲の ムード分類の検討 | 2022年3月 (発表予定) | 電子情報通信学会 総合大会 | 亀山 渉 菅沼 睦 |