修士論文

トピックモデルと感情推定を活用したネガティブ感情軽減のための音楽推薦システムの開発

Development of a Music Recommendation System to Reduce Negative Emotions Using Topic Models and Emotion Estimation

富山県立大学大学院 工学研究科 電子・情報工学専攻
2355020 水上和秀
指導教員 António Oliveira Nzinga René 講師
提出令和7年 (2025年) 2月

目 次

図一覧

表一覧

記号一覧

以下に本論文において用いられる用語と記号の対応表を示す.

用語	記号
実数の集合	$\mathbb{R}$
BERT の入力における語彙サイズ	$\|J\|$
BERT における埋め込みベクトルの次元	d
BERT における各トークン i のトークン埋め込みベクトル	$\mathbf{E}_{\text{token}n,i}$
BERT における各トークン i の位置埋め込みベクトル	$\mathbf{E}_{\text{poaition},i}$
BERT における各トークン i のセグメント埋め込みベクトル	$\mathbf{E}_{\text{segment},i}$
BERT における入力ベクトルトークン i における入力ベクトル	$\mathbf{E}_{\text{input},i}$
Attention における Query の重み行列	$\mathbf{W}^Q$
Attention における Key の重み行列	$\mathbf{W}^K$
Attention における Value の重み行列	$\mathbf{W}^V$
Attention における入力ベクトルに $\mathbf{W}^Q$ を掛けたもの	$\mathbf{Q}$
Attention における入力ベクトルに $\mathbf{W}^K$ を掛けたもの	$\mathbf{K}$
Attention における入力ベクトルに $\mathbf{W}^V$ を掛けたもの	$\mathbf{V}$
MultiHead Attention における次元調整するための線形変換行列	$\mathbf{W}^O$
入力画像	X
畳み込み層のフィルタ k における (i,j) の出力	$F_{i,j,k}$
畳み込み層のフィルタ k の (m,n) におけるチャネル c に対する重み	$W_{m,n,c,k}$
入力画像 (i,j) におけるピクセル値	$X_{i,j}$
畳み込み層におけるフィルタ k に対応するバイアス項	b_k
入力画像のサイズ	W_{size}
畳み込み層におけるフィルタのサイズ	K_{size}
パディングサイズ	P_{size}
フィルタが画像上を移動するステップ数	S
畳み込み層における特徴マップの出力サイズ	H_{size}
ローカル結合層におけるチャネル k における (i,j) の出力特徴マップ	$h_{i,j,k}$
全結合層の出力ノード i	u_i
感情カテゴリ k におけるスコア	O_k
感情カテゴリ k に対するスコアの確率	p_k
クロスエントロピーにおける正解の確率分布	y_k
クロスエントロピーの損失関数	L

用語	記号
LDA におけるトピック数	K
LDA における文章の数	M
LDA における単語の数	N
LDA におけるトピック分布のハイパーパラメータ	α
LDA における単語分布のハイパーパラメータ	β
LDA における文章 D がもつトピック分布	θ_D
LDA におけるトピック k がもつ単語分布	k
ガンマ関数	Γ
LDA における文章 D 内の n 番目のに対応するトピック	$z_{D,n}$
LDA における文章 D 内の n 番目の単語	$w_{D,n}$
LDA におけるトピック分布 θ を近似するディリクレ分布のパラメータ	γ
LDA における対数事後確率の下限	$\mathcal{L}$
ディガンマ関数	Ψ

第1章

はじめに

§ 1.1　本研究の背景

現代社会において，個人の幸福感や健康は感情のバランスに大きく影響される．ポジティブな感情は幸福感の向上やストレスの緩和，自尊心の向上に寄与する一方で，ネガティブな感情が強まると心理的な問題を引き起こす可能性が高まる．特に，ネガティブな感情が過剰になる場合，うつ病や不安障害といった心理的健康問題のリスクが高まり，日常生活や社会生活におけるパフォーマンスや満足度の低下を招くと考えられている [1]．このような背景から，ネガティブな感情を抑制する方法に注目が集まっている．

心理的ストレスを軽減する方法の一つとして，音楽が注目されている．音楽は古くから人々の心や体に癒しを与える手段として親しまれており，その音楽が持つ感情調整の効果を活用した治療法が近年さらに広がりを見せている [2]．音楽は人間心理に与える影響には生理的な反応から心理的な効果まで，さまざまな側面がある．例えば，音楽を聴くことで集中力や記憶力が向上し，不安が和らぎ，気分が改善されることが研究によって示されている．それに加え，音楽はテンポ，リズム，歌詞などの楽曲の特徴量などが感情に影響を与えるとされている．例えば，テンポの速い音楽は活力やエネルギーの向上に貢献し，穏やかなリズムを持つ音楽はリラクゼーションを促進するといった効果が確認されている．他にも，歌詞のテーマやポジティブさが感情的共感を引き起こすことも研究で示されている．

また，過去の再生履歴や楽曲のデータをもとにした関連のある楽曲を推薦する音楽推薦システムが存在する．例えば，Spotify では，ユーザの再生履歴や楽曲への「いいね」などの行動データを分析し協調フィルタリングを用いて，似たような嗜好を持つユーザが気に入った楽曲を推薦しているシステムや，楽曲の特徴量を用いて閲覧した楽曲と似た特徴量を持つ楽曲を推薦するシステムなどを実装している．

一方で，現在普及している音楽推薦システムは，ユーザの感情状態を十分に考慮していない．音楽推薦システムにおいて感情を考慮しない場合，ユーザが抱えるネガティブな感情を増幅させる可能性がある．例えば，深刻な悲しみや落ち込みを感じている人が明るく活発な音楽を聴いた場合，逆に不快に感じてしまう場合がある．この現象は「同質の原理」と呼ばれる．同質の原理とは，人は自分の感情状態に似た特徴を持つ音楽を選び，その音楽を聴くことで感情が安定するという効果がある原理のことであり，気持ちが落ち込んでいる場合は落ち着いた音楽を聴いた方がよく，気分がよいときは明るい曲を聴く方がよいことを表している．そのため，既存の音楽推薦システムで気分転換を行おうとした場合，感情状態が考慮されておらず，推薦された楽曲によってはストレスを感じてしまう場合がある．

§ 1.2　本研究の目的

1.1 節では，ストレスを軽減する方法の一つとして音楽を聞くことがあり，既存の音楽推薦システムの課題について述べた．これに対し，ネガティブ感情を軽減させるために音楽を推薦するには，ユーザの感情状態に応じた音楽を推薦する必要がある．

一方で，情報処理技術の発展に伴い感情分析の活用事例が増えてきている．感情分析とは，データから人の感情的な状態を抽出し，定量的に評価する技術である．この技術は，自然言語処理や画像処理，音声処理，生体情報解析など，さまざまなデータ形式を対象に分析されている．たとえば，テキストデータを対象とした感情分析では，レビューや SNS 投稿などからポジティブ，ネガティブ，中立といった感情を分類することが可能である．また，画像や映像からは顔の表情を解析して感情状態を推定することが可能であり，音声データでは声の抑揚や大きさをもとに感情を推定することができる．感情分析を活用することで，人の感情状態に合わせた物やサービスを提供でき，感情の改善を図ることが可能である．そこで，本研究では，テキストと表情を対象とした感情分析を活用したユーザの現在の感情状態を考慮した音楽の推薦システムを開発する．本研究のシステムは，Web から楽曲の特徴量および歌詞を取得し，ユーザの感情に影響する特楽曲の特徴量を取得する．また，各トピックにあったプレイリストを作成する．また，カメラ映像から表情を認識し，表情によりユーザの感情状態をリアルタイムで推定し，その感情状態にあったプレイリストを提示する．

具体的な手法として，まず，楽曲の特徴量を Spotify から取得する．楽曲のテンポ，アコースティック性，エネルギー，曲調など，感情に影響を与える特徴量を Spotify API を使用して取得する．また，楽曲名やアーティスト名も同時に取得する．ただし，Spotify API では歌詞の取得ができないため，取得したアーティスト名と楽曲名を基に Genius から歌詞を取得する．そして，Genius から取得した歌詞をもとに，BERT モデルを用いて感情分析を行う．BERT は自然言語処理のための深層学習モデルであり，文章分類，言語翻訳，感情分類など様々な自然言語処理タスクに応用される．BERT を用いた感情分析を行うことで歌詞の感情値を算出することができる．また，取得した歌詞をもとに，ガイド付き LDA を用いて楽曲をトピックを抽出し，そのトピックごとにクラスタリングを行う．そして，クラスタリングを行った楽曲に対して楽曲の特徴量をもとに，トピックごとのネガティブ度合いが高い，ネガティブ度合いが中程度，ネガティブ度合いが低い人のためのプレイリストを作成する．

そして，DeepFace を用い，カメラ映像からユーザの感情値を推定する．DeepFace とは顔認識を行うための深層学習モデルであり，照明の変化やカメラに対する顔の向きなどに左右されることなく顔を認識できる．DeepFace は年齢，性別，人種，感情の推定を行うことができる．DeepFace を用いることによりユーザの感情をリアルタイムに分析することができる．そして，カメラ映像からリアルタイムでユーザの感情分析を行い，推定された感情値に基づき，プレイリストを表示する．この時提示されるプレイリストはその時のネガティブ度合いに応じた楽曲の特徴量が入ったものになっている．そのプレイリストには別のトピックを選択できるようになっており，ユーザが自由に選択できる仕組みを設ける．これにより，カメラ映像にの感情状態を考慮した音楽を提示することで，ネガティブ感情を軽減を図る．最後に，複数人に実際にシステムを使用してもらい，楽曲を聴く前後で感情値の変化を分析し，開発したシステムの有効性を示す．

§ 1.3　本論文の概要

本論文は次のように構成される．

第 1 章 本研究の背景と目的について説明する．背景ではストレスを軽減するための音楽の重要性および既存の音楽推薦システムの課題について述べた．目的ではネガティブ強度に応じた楽曲の特徴量を持つ音楽を推薦するシステムを開発することを述べた．

第 2 章 音楽による感情や心理に与える効果および音楽の特徴量がもつに感情への影響について解説する．また，一般的な推薦システムの事例と課題について述べる．加えて，感情分析について，代表的な手法についてその概要を示す．

第 3 章 本研究の提案手法を構築するにあたって参考とした感情およびトピック分析手法の先行研究を挙げ，一般的な理論について解説する．

第 4 章 本システムにおける楽曲特徴量の取得方法について述べる．また，第 3 章で挙げた分析手法を用いてテキストと画像からの感情分析の流れや，プレイリストの作成方法を述べる．加えて，感情分析を活用した楽曲推薦システムの開発について，システムの概要と画面遷移を説明する．

第 5 章 提案手法における有効性の検証を目的として行った検証実験について，その概要と結果を示す．また，結果に対して考察を行い，提案手法において考慮すべき事柄について言及する．

第 6 章 本研究で述べている提案手法をまとめて説明する．また，今後の課題について述べる．

第2章

音楽と感情に基づく推薦システムのアプローチ

§ 2.1　音楽が感情・心理に与える影響

現在の社会は，ストレスの多い社会といわれており，そのような中で，リラクゼーションに関する関心が高まっている．その方法の一つとして音楽があげられる．音楽は古くから，人々の心や体に癒しを与える手段として親しまれており，その音楽が持つ感情調整の効果を活用した治療法が，近年さらに広がりをみせている．その代表的なものが音楽療法である．

音楽療法とは，音楽を用いて心身の健康を促進し，治療やリハビリテーションに役立てる療法のことである．音楽のリズム，歌詞，テンポなどが感情や身体的な状態に影響を与える力を活かし，感情的，精神的，身体的な問題を改善することを目的としている [3]．音楽療法の主な効果を図 2.1 に示す．

音楽療法は，主に「能動的音楽療法」と「受動的音楽療法」に分類される．それぞれのイメージを図 2.2 に示す．能動的音楽療法は，音楽を自分で演奏したり歌ったりすることを通じて，感情や身体の表現を促す手法である．患者が音楽を作り出す過程で，自己表現や創造性を発揮でき，感情の解放や心理的な安定を得ることができる．この方法は，特に心的外傷を抱えた患者やストレスを感じている人々に効果的とされており，音楽を演奏することで，感情の整理が進み，身体的な緊張も緩和されることが多い．

一方，受動的音楽療法は，音楽を聴くことによってリラクゼーションや心理的な回復を促進する方法である．音楽を聞くことによって，リラックスや集中を促し，感情のバランスを整えることができる．受動的音楽療法は，身体的なリラックスを促進し，精神的な安定を得るために広く活用されている [5]．

本研究では受動的音楽療法の観点から，ユーザのネガティブな感情の強度に基づいて音楽を推薦するシステムを提案する．ユーザの感情状態に対応した音楽を推薦することで，効果的に気分の向上を図る．以下に，音楽を聴くことによる生理的・心理的影響を述べたのち，本研究で考慮する音楽の特徴量とそれによる感情の影響を述べる．

音楽は人間の感情や心理状況に大きな影響を与えることが音楽心理学の研究において広く示されている．音楽が人間心理に与える影響には，生理的な反応から心理的な効果まで，さまざまな側面がある．

集中力と記憶力の向上

音楽を聴くことにより，認知能力を向上させることが研究によって示されている．特に，バックグラウンドミュージックは認知能力を強化することが示されている．高齢

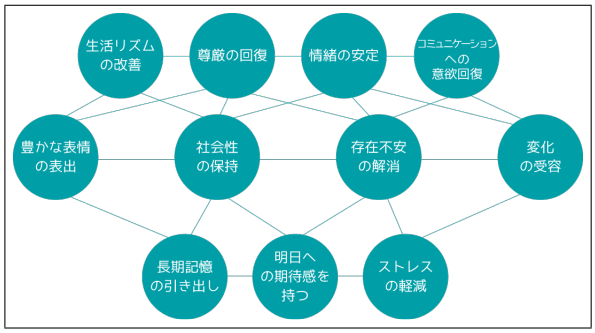

図 2.1: 音楽療法の効果 [4]

図 2.2: 音楽療法の種類 [6]

者を対象にした研究では，明るい音楽が処理速度を向上させ，穏やかな音楽が記憶力を改善する効果があることが示された．また，学生を対象にした研究では，クラシック音楽を聴きながら勉強することで，記憶保持力が高まり，試験の成績が向上したという結果が報告されている．[7]

ストレスの軽減

音楽を聴くことは，ストレスホルモンであるコルチゾールの値が減少するだけでなく，心拍数や血圧を調整することがわかっている．手術前の患者を対象とした研究では，音楽を聴くことでコルチゾールレベルが低下し，リラックス効果が確認された．日常生活でのストレスにおいても，リラックスできる音楽を聞くことがストレス軽減に貢献するとされている [8].

気分の向上

音楽を聴くことにより，不安を和らげ，気分を向上させることができる．明るくポジティブなメロディーや歌詞は，幸福感や満足感を促進し，気分を向上させる効果がある．アップテンポの音楽やポジティブな歌詞を聴くことで，セロトニンやドーパミンなどの「幸福ホルモン」が分泌される [9]

音楽の感情的効果はいくつかの特徴量に影響される．主な特徴量として，テンポ，曲調（メジャー，マイナーなど），エネルギー，アコースティック性，そして歌詞の内容が挙げられる．これらの特徴量はその大きさによって感情に与える影響が異なっている．以下より，それぞれの特徴量の大きさによる感情に与える影響の違いを説明する．

テンポにおける感情的効果

テンポは 1 分間の拍数 (BPM) で測定される．テンポは，その違い (早いテンポ，遅いテンポ) によって感情に影響を与える．速いテンポの曲は喜びや興奮，活気を喚起され，中程度のテンポは安定感と穏やかな気分に誘導し，リラックスしながら集中力を保てる状態をもたらす．また，遅いテンポの曲は副交感神経を活性化させ，深いリラクゼーションを促進させる．このような効果は脳の活動にも影響され，早いテンポでは側頭葉の活性化が強まり，遅いテンポでは心拍数や血圧の低下がみられる．また，中程度のテンポの音楽は，特に聴覚皮質や感情記憶に関連する脳領域を活性化し，強い感情的覚醒をもたらすとされる [10].

曲調における感情的効果

音楽の曲調 (メジャー・マイナー) によって感情に影響を与えることがいくつかの研究によって示されている．軽度のネガティブ感情を抱える人々に対しては，長調 (メジャーキー) な曲が感情を改善し，ポジティブな感情を喚起することが示されている．例えば，メジャー調や高テンポの音楽が，リラックス効果やエネルギーの向上をもたらし，ネガティブな感情を軽減することが確認されている．長調の音楽は活力を与え，心理的なストレスを緩和する効果が期待される．

一方で，深刻な悲しみやうつ病状態のような極度にネガティブな感情を抱えている場合，長調よりも短調の方がリラックスできることがわかっている．一般的に，長調の音楽は明るく快活な印象を与え，ポジティブな感情を喚起しやすく，短調の音楽は，より内省的で落ち着いた雰囲気を持ち，感情を穏やかに整える効果があるとされている．特に，深刻な悲しみやうつ病のような状況では，過度にポジティブな音楽が逆に感情の圧迫感や不安感を引き起こす可能性があるため，感情の急激な変化を避けるためにも，リラックスできる短調の音楽が適しているとされている [11].

また，他の研究によれば，短調の音楽は心理的な安定感を高め，過度の感情的な変化を防ぐ効果があるとされている．具体的には，短調の音楽が聴く人の心を落ち着かせ，心拍数や呼吸数を安定させることが示されており，これにより，リラクゼーションやストレス軽減が促進されるとされている．また，悲しみや絶望感といった強いネガティブ感情に直面した人々に対しては，短調の音楽が感情の調整をスムーズに行い，無理なく心を落ち着ける助けとなる [12].

エネルギーにおける感情的効果

エネルギーは音楽が持つ「活力」や「力強さ」を示す指標であり，音楽のリズム，音量，テンポなどに関連している．高エネルギーの音楽は，活気に満ちたリズムとメロディーによって気分を明るくし，エネルギッシュな感情を引き起こす．これにより，喜び，興奮，活力といった感情を感じやすくなり，ポジティブな心情を促進する．一方，低エネルギーの音楽は，リズムが遅く，メロディーやハーモニーが穏やかで，音量も控えめである．低エネルギーの音楽は，心拍数を下げ，リラックス効果を高めるため，ストレスや緊張を緩和する効果があるとされている．

アコースティック性おける感情的効果

アコースティック性とは，曲がどれだけ自然な楽器の音，または生音に近いかを示す指標のことである．アコースティック性が高い音楽は，ギター，ピアノ，弦楽器，ドラムなど，生楽器を主体とした音が使用されており，自然で心地よい音色であることが特徴である．また，多くの研究において，アコースティック性が高い音楽が感情のバランスをとる助けになることが示されている．例えば，アコースティック性の高い音楽はストレスホルモンであるコルチゾールの分泌を減少させ，心理的な安定感を高める効果があることが示されている [13]．また，ほかの研究でも，アコースティック性が高い音楽が感情の調整に有効であることが確認されている．これらの研究からも，アコースティック音楽が心理的な安定や感情の調整に有効であることが示唆されている [14]．特に，アコースティック性の高い音楽はリラクゼーション効果が高く，心拍数や血圧の低下を促すため，ネガティブな感情が強い状態にある場合に聴くことでリ

ラックスを促し，気分が改善されることが期待される．一方で，アコースティック性が低い音楽は，エレクトロニック音楽やシンセサイザーを多く使用した音楽を指し，リズムやビートが強調され，感情に強い影響を与える．このような音楽は，活力や興奮を促進し，身体的にも精神的にもエネルギーを増加させる効果がある．特に，ネガティブな感情が少ない人は，アコースティック性が低い音楽のエネルギーを受け入れやすく，ポジティブな感情を高めることができる．逆に，ネガティブ度合いが高い人には，アコースティック性が低い音楽を聴くことはあまり効果的ではないとされている．特に，心の状態が不安定である場合，このような音楽は感情をさらに乱す可能性があり，リラックスや心理的な安定を促進するには適していない．

歌詞における感情的効果

曲の歌詞は，感情に影響を与えることが示されており，その影響の度合いは感情値により異なる．ポジティブな歌詞の音楽を聴くことにより，ネガティブな感情を軽減する効果について，多くの研究が示している．例えば，ある研究では，ポジティブな感情を表現する歌詞（「愛」「幸せ」など）が幸福感や満足感を引き起こし，心の状態を改善することを示している [15]．歌詞の感情的な内容が感情に与える影響は大きく，ポジティブな感情を持つ歌詞は，気分を向上させ，ストレスや不安を軽減する効果があることが示されている．

一方，ネガティブな歌詞は悲しみを軽減することが実験により示されている．実験によると，悲しい音楽が気分にどのような影響を及ぼすか実験的に検討した結果，やや悲しい場合に悲しい音楽を聴くと，悲しみは低下しないが，非常に悲しい場合に悲しい音楽を聴くと，悲しみが軽減することが示唆された [16]．

ここまで，音楽の特徴量が感情に与える影響について述べてきたが，音楽の効果は，その時に聴く人の現在の感情状態に大きく依存する．音楽による気分の変化について，同質の原理（Iso-principle）という考え方が提唱されている [17]．

同質の原理

同質の原理とは，人が自分の感情状態に似た特徴を持つ音楽を選び，その音楽を聴くことで感情を安定させる効果を持つ原理である．例えば，ネガティブな感情を抱えている場合には，その感情に同調するような音楽を聴くことで，感情的な一致感が生まれ，リラクゼーションや感情の調整が進むとされる．

同質の原理は，ネガティブな感情状態にある場合には，ポジティブな音楽よりもネガティブな感情に同調する音楽が効果的であることを示唆している．例えば，深い悲しみや落ち込みを感じている人が，明るく活発な音楽を聴いても，その音楽が感情的に異質に感じられることがある．この場合，感情的に明るい音楽は逆に不快に感じ，気分転換の効果が薄れる可能性がある．逆に，悲しい気分を持っている場合，暗い音楽を聴くことで感情的な一致感が生まれ，その後にポジティブな感情に移行する助けになると考えられる．

同質の原理は，音楽推薦による気分誘導にとって重要な概念である．ネガティブな感情を持つユーザに対して，無理にポジティブな音楽を薦めるのではなく，感情的に一致する音楽を選ぶことで，感情の調整が促され，最終的にポジティブな感情への移行をサポート

することができる．このように，感情の強さや状態に合わせた音楽推薦を行うことで，効果的に感情の調整を行うことが可能になる．

本研究では同質の原理に基づき，ユーザのネガティブ度合に応じた音楽の推薦を行う．具体的には，ネガティブ強度が強い場合には感情的に一致する特徴量を持つ音楽を推薦することで，ユーザが自分の感情と調和し，リラックスや心の安定を促進できるようにする．また，ネガティブ度合が中程度または低い場合には，少しポジティブな特徴量を持つ音楽を選択し，少しずつ気分を高めるような音楽を推薦する．これにより，最終的に気分の向上をサポートする．

§ 2.2　音楽および他分野における推薦システム

近年，ソーシャルメディアの発展やスマートフォンなどのICT機器の普及により，多くの人々がインターネットに触れる機会が増えている．その中でも，YouTube[1]をはじめとする動画配信サイトや，Amazon[2]をはじめとするEC（Electronic Commerce：電子商取引）サイトなどを利用する機会が増加している．これらのサービスでは，動画や商品を閲覧している際に，関連性の高い別の動画や商品を提示されることが一般的である．このような技術は「情報推薦システム」と呼ばれ，現在その重要性が高まっている．また，レコメンドシステムのイメージを図 2.3 に示す．

推薦システムにはいくつかの手法が存在し，主に情報推薦における嗜好の予測方法としてルールベース，コンテンツ（内容）ベースフィルタリング，協調フィルタリング，ハイブリッド法の 4 つに分けることができる．以下にそれぞれの手法について述べ，その後，他分野及び音楽分野における情報推薦の事例，その課題について紹介する．また，図2.4はコ推薦システムとして広く使われているコンテンツベースフィルタリングと協調フィルタリングの概要であり，表2.1はコンテンツベースフィルタリングと協調フィルタリングの比較を示す表である．

ルールベース推薦

ルールベース推薦とは，システムの運営者があらかじめ「特定の行動をとった人や特定の属性を持つ人に対してどのような商品や情報を提供するか」というルールを設定し，そのルールにしたがってレコメンドを行う手法である．例えば，パスタを購入しようとしている人にチーズやソースを推薦するように，利用者の行動や嗜好を予測して適切だと思われるルールを設定する．この手法は比較的低コストで容易に実装可能であり，透明性が高いという利点がある．運営者が設定したルールに基づいて商品を推薦するだけなので，推薦の結果を簡単に説明できる．しかし，あらかじめ設定したルールに依存するため，利用者の多様なニーズや嗜好を細かく反映することが難しいという欠点がある．また，利用者の行動パターンが変化した場合に対応するには，ルールを手動で更新する必要があるため，運用コストが高くなる可能性がある．

コンテンツベースフィルタリング

コンテンツベースフィルタリングは，アイテム（利用者に推薦する対象やコンテンツ）

[1] https://www.youtube.com
[2] https://www.amazon.co.jp

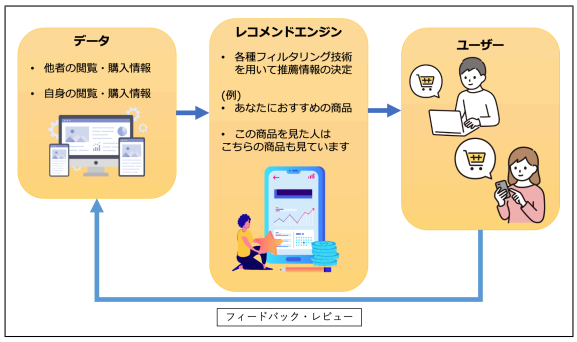

図 2.3: レコメンドシステムのイメージ

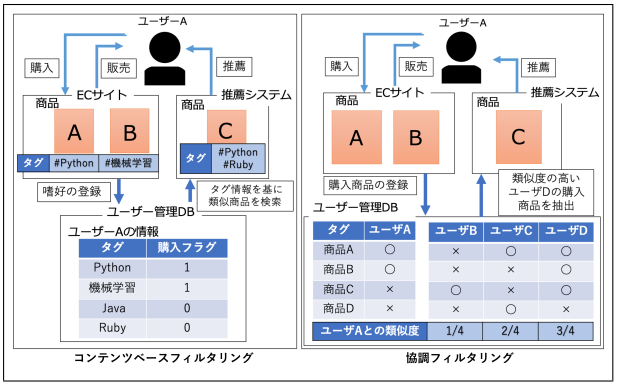

図 2.4: コンテンツベースフィルタリングと協調フィルタリング

の特徴量を使用し，利用者の嗜好に近いアイテムを推薦する手法である．例えば，映画を推薦するシステムを考えてみる．この場合，利用者が「アクション映画」や「サスペンス」といったジャンルを検索すると，システムはそれらの特徴量を分析し，ジャンルやテーマが類似する映画を選び出して推薦する．また，過去に視聴した映画のデータを基に，類似した内容や雰囲気の映画を提案することも可能である．この手法は，利用者の過去の嗜好に基づき，個人に特化した精度の高い推薦ができるという利点がある．利用者が以前に好んだアイテムと特徴の類似度を計算しアイテムを優先的に推薦するため，個々の利用者に特化した推薦が可能である．一方，この手法は過去の嗜好に基づいて推薦を行うため，利用者に同じようなアイテムが繰り返し推薦され，新しいアイテムの発見が難しいという欠点がある．コンテンツベースフィルタリングには，利用者が自分の好むものを直接指定する「直接指定コンテンツベースフィルタリング」と，利用者の嗜好データからプロファイルを作成し，アイテムデータベースと比較することで嗜好を測る「間接指定コンテンツベースフィルタリング」が存在する．

協調フィルタリング

協調フィルタリングは，システム使用前に利用者の嗜好データを収集し，それを基に利用者の好みや傾向（例えば，どのアイテムが好まれ，どれが嫌われるか）を分析する手法である．この手法では，嗜好が類似している利用者同士が，共通して好むアイテムや嫌うアイテムがあると仮定し，似たような嗜好を持つユーザを見つけ出す．その後，そのユーザが好むと思われるアイテムを推薦する．

協調フィルタリングは大きく分けてユーザベース協調フィルタリングとアイテムベース協調フィルタリングの二つに分類される．ユーザベース協調フィルタリングとは似た嗜好を持つ他のユーザを見つけ，そのユーザが高く評価したアイテムを推薦する手法のことであり，アイテムベース協調フィルタリングとはアイテム同士の類似性に基づき，ユーザが過去に好んだアイテムに似たアイテムを推薦手法のことである．これら手法は他の利用者の行動データを活用するため，利用者の過去の評価や行動に基づいて適切な推薦ができるという特徴がある．しかし，協調フィルタリングはシステムに新規ユーザや新規アイテムが追加された場合に過去のデータがないため推薦することが難しくなるという「コールドスタート問題」やユーザとアイテムの評価データが非常に少ないためにゆーザー間の類似性

表 2.1: 協調フィルタリングとコンテンツベースフィルタリングの比較 [18]

分類	協調フィルタリング	コンテンツベーフィルタリング
多様性	○	×
ドメイン知識	○	×
スタートアップ問題	×	△
利用者数	×	○
被覆率	×	○
類似アイテム	×	○
少数派の利用者	×	○

を正確に算出できず推薦精度が低下してしまう「スパース性問題」などの課題がある．これらの問題を解決するために，コンテンツベースフィルタリングやハイブリッドフィルタリングと組み合わせて使われることが多い．

ハイブリッドフィルタリング

ハイブリッドフィルタリングとは，コンテンツベースフィルタリングと協調フィルタリングを組み合わせた手法である．コンテンツベースフィルタリングは推薦がパターン化されやすく，協調フィルタリングは，新規ユーザや新規アイテムに対して機能しにくいといった欠点があるが，ハイブリッドフィルタリングはそれらの欠点が補われ，幅広いユーザに適切な推薦を行うことができる．

推薦システムは，現在，様々な分野で活用されており，その適用事例も増えている．例えば，EC サイトの分野では商品の推薦，教育分野では学習コンテンツの提示，利用者のニーズや好みに応じた提案を行う仕組みが使われている．こうした他分野での具体的な推薦システムの事例について紹介し，その後，音楽分野における推薦システムの事例を紹介する．

他分野での推薦システムの事例

1. Amazon の推薦システムの事例
   Amazon の推薦システムは，協調フィルタリング，コンテンツベースフィルタリング，機械学習を組み合わせて推薦を行っている．協調フィルタリングでは，過去のユーザの行動を分析し，他のユーザの行動と類似したアイテムを推薦している．コンテンツベースフィルタリングでは，商品を購入した人に対して，商品の特徴や属性（ジャンル，ブランドなど）を分析し，類似した商品を推薦している．また，機械学習により，履歴データでモデルをトレーニングすることで予測精度を向上させている [19].

2. Coursera における推薦システムの事例
   Coursera は，オンライン学習プラットフォームで，世界中の大学や企業と提携し，専門的なコースや専門職の資格を提供しているサイトである．Cousera は学習者の過去の履歴や興味に基づき，関連するコースを推薦している．ユーザのプロファイル情報（学習分野，スキルレベル，過去に受講したコースなど）を元に，次に学ぶべきコースを予測している．また，機械学習を利用して，ユーザに合わせた最適なコースを提案することで，学習体験の向上を図っている．[20]

音楽分野でのシステムの事例

1. YouTubeMusic での推薦システムの事例
   YouTube Music では，Transformer モデルを活用して音楽推薦を行っている．Transformer モデルは，言語翻訳や分類タスクに使用されていた自然言語処理技術であり，ユーザの音楽の視聴，スキップ，いいねなどの行動を分析している．さらに，既存のランキングモデルと Transformer を組み合わせることで，現在のユーザの行動と過去のリスニング履歴を反映したランキングを学習している．この方法により，ユーザのスキップ率が減少し，音楽を聴く時間が増加する結果が得られ，ユーザの満足度が向上してている [21].

2. Spotify での推薦システムの事例
   Spotify では，協調フィルタリング，自然言語処理，音声解析の技術を組み合わせて個人に合った音楽を推薦している．Spotify はユーザのプレイリスト，再生履歴，楽曲への「いいね」などの行動データを分析し，似たような嗜好を持つユーザが気に入った楽曲を推薦している．また，楽曲のテンポ，音程，音量などの音楽的特徴を分析し，類似した特徴を持つ楽曲を識別し，推薦に活用している．さらに，アーティストや楽曲に関連する記事，レビュー，ブログ投稿などのテキストデータを収集・分析し，楽曲がどのように言及されているかを理解し，ユーザーの関心に合った曲を推薦している [22].

このように，音楽分野及び他分野では推薦システムが活用されている．また，ユーザ一人ひとりの嗜好性に合ったコンテンツを推薦する手法は，すべてのユーザに同じコンテンツを推薦する手法に比べて次の 2 つの効果が期待される．

購買・利用促進

ユーザが興味を持ちやすい商品やサービスを推薦することで，購入率や利用率を向上させることができる．YouTube の推薦システムの研究によると，協調フィルタリングに基づいた推薦システムを YouTube に適用した結果，視聴回数による人気ランキングを表示した場合と比較し視聴数の増加やクリック率の向上が確認され，満足度も向上したとされている [23].

多様性と満足度の向上

推薦システムを利用することによって，新しいアイテムやジャンルを知ることができ，それにより満足度を高めることがわかっている．ある研究によると，推薦システムはユーザに関連性の高い情報を提示することで，新しい商品やジャンルを発見でき，より充実した体験を得ることができるとされている．また，推薦システムの多様性と正確性にはユーザの満足度に対する相互作用効果があり，これは，ユーザが探していたものと関連性が高く，かつ新しい商品を推薦することで，よりユーザの満足度が向上することが示されている [24].

推薦システムを導入することによってさまざまな効果があり，多くの分野で使用されている．しかし，既存の推薦システムは，主に過去のユーザデータや他のユーザの行動履歴

に基づいてアイテムを推薦することが多い．これらのシステムは過去のデータに大きく依存しており，リアルタイムで変化するユーザの感情や心理状態を十分に反映できていない．そこで本研究ではユーザの感情状態を考慮し，その感情に合った特徴量を持つ音楽の推薦を行う．本研究のシステムは，従来の感情を考慮しない推薦システムと比較し，ユーザの心理的満足度や体験価値を大幅に向上させる可能性があり，ネガティブ感情を緩和しポジティブ感情を促進する音楽の影響を活用することで，メンタルヘルスの向上にも寄与することが期待される．

§ 2.3　感情分析の手法および活用事例

感情分析とはデータから人々の感情的な状態や意図を抽出し，それを定量的・定性的に分析する技術のことである．感情分析の対象は，テキストや音声，映像，生体情報などがある．各感情分析の比較を表 2.2 に示す．分析テキストによる分析では文章に含まれている単語や表現を分析することで，書き手の感情分析する．音声による感情分析では声の抑揚や声の大きさを分析することで話し手の感情を分析する．人間の感情表現は非言語による部分が大きいため，音声分析では，テキストでは読み取れない情報を読み取れることができる．

また，映像による分析では顔の表情や顔の筋肉の動きを分析することで対象の感情を分析する．映像による分析では音声に頼らずに感情をとらえるため，視線やジェスチャーなどの言葉を使わない感情表現を分析することができる．生体による分析では脳波や脈拍などの生体データを分析する．生体反応は無意識的に起こるため，生体による分析は感情を客観的に分析することができる．

感情分析の手法は大きく分けて「ルールベースによる手法」,「機械学習による手法」,「深層学習による手法」3 つのアプローチがある [25]．以下にそれぞれの手法の概要及び特徴について説明する．

ルールベースによる手法

ルールベースによる手法は，前持って定義された一連のルールを用いて感情を分析する手法である．ルールの定義自体は，基本的に人間が手動で行う．特徴としては，比較的簡単に実装できる一方，複雑な感情表現や大規模データへの対応が難しいという欠点がある．

テキストにおけるルールベースによる手法は，感情極性辞書を用いて感情を分析する．感情極性辞書とは，単語に対して感情のスコア (「良い:0.999995，悪い:-1 など」) が割り振られた辞書であり，この辞書をもとにテキスト全体の感情スコアを集計し感情を推定する．日本語の感情分析で使われる感情極性辞書は単語感情極性対応表 [27] と日本語感情極性辞書がある [28]．音声データでは，音響特徴（例: ピッチ，テンポ，音量）に基づく閾値ルールを用いる．例えば，高音量で急なテンポ変化が「怒り」を示し，低音量で緩やかなテンポが「悲しみ」を示すといったルールを構築し，抽出された音声特徴とルールに基づいて感情カテゴリを分類する [29]．

画像データ，特に顔画像においては，眉や口角の位置といった表情特徴に基づくルールを設定する．有名な画像の感情分析におけるルールとして，Facial Action Coding System (FACS) に基づくルール方式がある．FACS には，顔の筋肉の動きを定義し，それらの組み

表 2.2: 各感情分析の比較 [26]

感情分析の種類	分析対象	活用例
音声	声の抑揚 声の大きさ	電話口での怒りの感情から対応方法を判断
テキスト	単語	口コミから「ネガティブ」「ポジティブ」に分類して集計、商品改善に活かす
画像	顔の表情	ドライバーの感情を表情で判定、あおり運転や居眠り運転などを未然に回避
生体情報	脳波 脈拍 発汗	授業中の生徒の脈拍から集中度をチェックし、生徒の状態に合わせて授業を実施

合わせで特定の感情を表現するというものである．例えば，喜びは口角の上昇と頬の上がりを伴う表情であり，これらの筋肉の動きを観察することによって「喜び」と識別できる．また，怒りは眉が下がり，中央に寄るとともに，唇が強く閉じられるという特徴を示し，これをもとに「怒り」の感情を分類することができる [30]．

生体データでは，脳波測定や心拍センサにより取得できた特徴量に対して閾値を設定する．例えば，ストレス状態の場合はコルチゾールの値を，リラックスしている場合はオキシトシンの数値や心拍の変動をもとに感情を数値化する [31]．ルールベースの手法は，感情分析における簡易的なアプローチとして，基本的でわかりやすい特徴を基にしているが，画像データや音声データにおける多様で複雑な表現には限界があり，機械学習や深層学習などのより高度な手法と組み合わせて使用されることが多い．

機械学習による手法

機械学習による手法は，データからパターンや規則を学習し，未知のデータに対して予測を行う教師あり学習を用いた手法である．教師あり学習とは，入力データとその対応する正解ラベル（教師データ）がペアで与えられ，分類器が入力データの特徴と正解ラベルの関連性を学び，未見のデータに対しても正しい予測を行えるように訓練される手法である．

機械学習による手法においては，まずデータから適切な特徴量を抽出し，それを分析器に入力して分析を行う．特徴量とは，データが持つ重要な情報を数値やベクトル形式で表現したものである．例えば，テキストデータであれば，TF-IDF や Bag of Words などの方法を使用して単語の頻度を特徴量として抽出し，感情分析やトピック分類を行う．

画像データの場合，Haar-Like 特徴量を用いて画像における明暗差の集合を特徴量として抽出し，顔認識や物体分類，感情の表情分析を行う．音声データにおいては，声の高さや速さ，周波数，強度などを特徴量とする．声の周波数，強度はフーリエ変換を用いることにより波長へ変換され特徴量として抽出される [29]．また，生体データでは，デルタ波，シータ波，アルファ波などの周波数帯域別のパワーや心拍数などの特徴量を使用する．特徴量抽出には脳波測定機器や心拍センサなどを用いる．

これらの特徴量を基に，サポートベクターマシン (Support Vector Machine:SVM) や ランダムフォレスト ，k-近傍法 (k-nearest neighbor algorithm:K-NN) などの機械学習アルゴリズムを使用して感情分析が行われる．これらのアルゴリズムは，与えられた特徴量と対応する感情ラベルを学習し，新たなデータに対して感情を予測するために用いられる．

SVM

SVM は，異なるクラスを分けるための最適な境界（決定境界）を見つけ，データの

分類を行う．この決定境界は，クラス間のマージンを最大化するように設計されており，データがどのクラスに属するかを予測するために使用される．

ランダムフォレスト

ランダムフォレストは，多数の決定木を組み合わせて分類を行う手法で，個々の決定木の予測を集約することで高い予測精度を実現する．複数の木を使うため，過学習を防ぐことができ，特に複雑なデータに有効である．

k-NN

k-NN，データポイントを特徴量に基づいて他のデータポイントと比較し，最も近い k 個のデータのラベルを基に予測を行うシンプルなアルゴリズムである．感情分析では，特徴量に基づいて感情ラベルを決定するために使用される．

感情分析の場合，この教師あり学習を活用して，感情ラベル（例えば，ネガティブ，ニュートラルなどの 2 値クラスラベルや喜び，驚き，悲しみ，怒り，恐怖，嫌悪などの多クラスラベルなど）が付与された大量のデータを使ってモデルを訓練し，入力データを各ラベルに分類する．

深層学習による手法

深層学習による手法は，人工神経ネットワークを基盤にした手法で，特に大規模なデータセットを使って非常に高精度な感情分析を実現する方法である．従来の機械学習手法と比較して，深層学習は特徴量抽出とモデル学習を同時に行う能力があり，これにより手作業で特徴量を設計する必要がなくなる．また，深層学習は複雑で非線形な関係を学習することができ，テキスト，音声，画像，そして生体データにおいて高い性能を発揮する．深層学習による感情分析には，主に以下の 3 つのアーキテクチャが使用される．

RNN

Recurrent Neural Network(RNN) は，音声などの時系列データや文章など，順序が重要なデータを扱うために設計されたニューラルネットワークである．従来のニューラルネットワークでは，入力データの長さが固定である必要があったが，RNN は各時点の出力を次の時点の入力にフィードバックする構造を持つため，データ間の時系列的な依存関係を捉えることが可能である．ただし，長い時系列データにおいては，情報が時間の経過とともに失われる「勾配消失問題」が生じるという欠点がある．

CNN

Convolutional Neural Network(CNN) は，畳み込みを使用するニューラルネットワークであり，画像や音声などに適用可能である．畳み込みとは，入力データに対してフィルター（カーネル）を適用する処理である．フィルターは，データの局所的な領域（例えば，画像の小さな部分や音声の一部）に対してスライドし，その領域内の特徴を抽出する．畳み込み操作は，入力データとフィルターを重ね合わせ，出力を生成する．この操作によって，フィルターは画像のエッジや音声の特定の音の特徴を捉え，ネットワークが重要なパターンを学習できるようにする．

Transformer

近年，翻訳などの入力文章を別の文章で出力するというモデルは，Attention を用いたエンコーダー，デコーダ形式の RNN が主流であった．しかし RNN や LSTM は逐次的に単語を処理しているため，訓練時に並列処理ができないという欠点があった．それに対し Transformer は，RNN や CNN を用いず Attention のみを用いている．Transformer は，再帰も畳み込みも一切行わないので並列化が容易であり，他のタスクにも汎用性が高いという特徴がある．

機械学習と深層学習の主な違いは，特徴量の抽出方法とモデル学習のアプローチにある．機械学習では，データから有用な特徴量を手作業で設計・抽出し，それらの特徴量を基にモデルを訓練する．一方，深層学習では，特徴量の抽出とモデル学習が一体となって行われ，データから直接重要な特徴を自動的に学習する．

また，感情分析を用いた活用事例が増えており，様々な分野で活用されている．以下に，その事例および効果を説明する．

感情分析の応用事例

1. ベネッセにおける感情分析の活用事例
   ベネッセは画像認識を用いた感情分析を活用して教育現場の改善を行っている．具体的には，授業中の生徒の表情をリアルタイムで分析し，授業の質を評価する取り組みを行っている．画像認識では顔の表情により生徒の感情を解析し，教師は授業の進行中に正との反応を把握し，必要に応じて授業内容や方法を調整している．この取り組みにより，効率的な教育を提供している [32].

2. Amazon における感情分析事例
   Amazon では，テキストの感情分析を活用して，カスタマーレビューの感情を分析している．これにより，どのレビューがどれくらいポジティブ・ネガティブかを判別することができ，顧客の意見を理解し，製品改良やマーケティングに役立てている [33].

3. 日立における感情分析事例
   日立は HondaMotor と共同で，SNS やブログなどの顧客の生の声を分析し，サービス改善に生かす感情分析サービスを開発した．このサービスは，消費者の感情やレビューをリアルタイムで解析し，製品やサービスに対する意見を収集している．これにより，新製品の発売時やキャンペーンの効果測定に活用されている．また，SNS 上の反応を分析し，迅速にフィードバックを得ることで，製品開発やブランド価値の向上，イノベーティブなサービス創出に繋げることができる [34].

このように感情分析は様々な分野で活用されており，注目が集まっている．感情分析を活用することで人の感情状態に合わせた，ものやサービスを提供することができ，これにより感情の改善を行うことが可能である．

本研究ではユーザの感情を推定する手法として表情からの感情推定の深層学習におけるアプローチを活用する．また，歌詞の感情値を分析する方法として，テキストからの感情推定の深層学習におけるアプローチを活用する．また，これによりユーザの感情にあった音楽を推薦することで，ネガティブ強度の減少をサポートする．

第3章

感情推定と歌詞のトピック分類

§ 3.1　テキストからの感情分析

2.2 説でも述べた通り，曲の歌詞は感情に影響を与えることが示されており，その影響は感情値によって異なっている．例えば，明るい歌詞の楽曲は気分が落ち込んでいる人にとっては抵抗感を覚え，気分が落ち込んでいない人にとっては気分を向上させることがあげられる．このように，楽曲を推薦するにあたり，ユーザの感情にあった感情値を持つ歌詞を提示することが必要である．

本研究では歌詞を Bidirectional Encoder Representations from Transformers(BERT) を用いて感情分析を行い，歌詞の感情値を推定する [35]．BERT は，文章の意味を理解する能力に優れているため，自然言語処理の様々なタスクに活用されている．

BERT

BERT は，Google が提案した自然言語処理モデルの一つであり，Transformer の Encorder 部分を基盤とし，Attention メカニズムを用いて単語間の関係性をモデル化している．BERT による処理の流れを図 3.2 に示す．BERT は双方向の Transformer アーキテクチャを持っていることが最大の特徴であり，文脈を考慮して単語の意味を理解することができる．また，BERT は，テキスト内の単語の位置情報を学習することができるため，単語の順序を考慮した文脈理解が可能である．これにより，文書分類，質問応答，言語翻訳など，様々な自然言語処理タスクに応用されている．BERT は，大規模なコーパスから事前に学習されたモデルであり，学習済みのモデルを転移学習することで，より小規模なデータセットでも高い精度で処理することができる．

BERT のモデルである Transformer はベクトル化された文章を入力とし，Attention を多数並列に配置した Multi-Head Attention が用いて分析を行っている．以下に，BERT における入力部分の説明を行う．

BERT における入力部分

言語を用いたタスクを解く際には，モデルが言語を扱えるように数値化する必要がある．BERT では，まず MeCab を使用して文を単語に分割し，その後 WordPiece を用いて単語をさらにトークンに分割する．BERT の日本語モデルでは，32,000 個のトークンが定義されており，各トークンには固有の ID が割り振られている．BERT への入力時には，このトークン ID が使用される．トークン ID に変換されたデータは，BERT モデルに入力される前

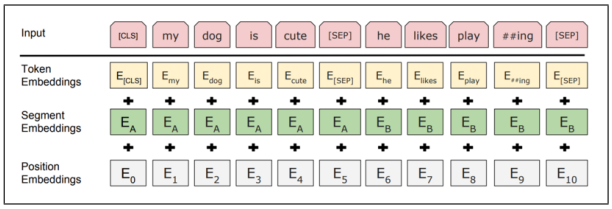

図 3.1: BERT の入力部分 [35]

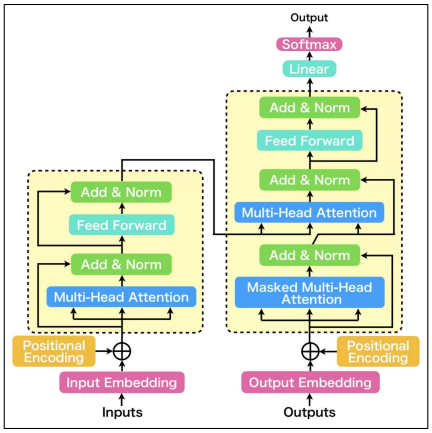

図 3.2: BERT の流れ [35]

に，Input の先頭に [CLS] トークンを，文の終わりに [SEP] トークンを追加する．そして，以下の 3 種類の埋め込みを加えることで，モデルに適した数値ベクトルとして表現される．また，BERT の入力部分を図 3.1 に示す．

トークン埋め込み

トークンごとに，事前学習された埋め込み行列 $\mathbf{E} \in \mathbb{R}^{|J| \times d}$ を用いてベクトル表現に変換する．ここで，$|J|$ は語彙サイズ，d は埋め込みベクトルの次元である．各トークン ID t_i に対応するトークン埋め込みベクトル $\mathbf{E}_{\text{token},i}$ は次のように定義される．

$$\mathbf{E}_{\text{token},i} = \mathbf{E}[t_i] \tag{3.1}$$

位置埋め込み

トークン埋め込みを行っただけでは入力の順序に関する情報を持たないため，文章を正しく扱えない．そのため，トークン列内での順序情報をモデルに追加するため，位置埋め込みを加える．各位置 i に対応する位置ベクトル $\mathbf{E}_{\text{position},i}$ は次のように計算される．ここで，pos はトークンの位置（0, 1, 2, ...），i は埋め込み次元のインデックスである．これにより，固定長の埋め込みベクトルを用いながら，トークンの相対的な順序や距離をモデルが学習可能となる．

$$\mathbf{E}_{\text{position},i} = \begin{cases} \sin\left(\dfrac{\text{pos}}{10000^{\frac{2i}{d_{\text{model}}}}}\right) & \text{if } (i \mod 2 = 0) \\ \cos\left(\dfrac{\text{pos}}{10000^{\frac{2i}{d_{\text{model}}}}}\right) & \text{if } (i \mod 2 = 1) \end{cases} \tag{3.2}$$

セグメント埋め込み

BERT では，1 つの入力が単一文か複数文かを区別するために，セグメント埋め込みを使用する．各トークンには，対応するセグメント ID s_i（文 1 なら 0，文 2 なら 1）が割り振られ，セグメント埋め込みベクトルは以下のように表される．

$$\mathbf{E}_{\text{segment},i} = \mathbf{S}[s_i] \tag{3.3}$$

ここで，$\mathbf{S} \in \mathbb{R}^{2 \times d}$ はセグメント埋め込み行列であり，s_i が 0 または 1 に応じて適切なベクトルが選択される．

埋め込みベクトルの統合

最終的に，トークン埋め込み，位置埋め込み，セグメント埋め込みを加算して，モデルに入力するベクトル $\mathbf{E}_{\text{input},i}$ を生成する．

$$\mathbf{E}_{\text{input},i} = \mathbf{E}_{\text{token},i} + \mathbf{E}_{\text{position},i} + \mathbf{E}_{\text{segment},i} \tag{3.4}$$

これにより，各トークンには，その意味（トークン埋め込み），位置（位置埋め込み），文区別（セグメント埋め込み）の情報が含まれるベクトルが追加される．ベクトル化されたトークンは Attention を多数並列に配置した Multi-Head Attentiont と Feed-Forward Neural Network(FFNN) によって分析され，分析結果を文章における特徴量としている．以下に Attention および Multi-Head Attentiont，FFNN における説明を行う．また，BERT における特徴量取得の流れを図 3.2 に示す．

Attention

Attention は，入力の各トークンが他のすべてのトークンにどれだけ関連しているかを学習するメカニズムであり，この機構は，一般的に自己注意（Self-Attention）といわれている．Self-Attention では，各トークンの埋め込みベクトルを Query，Key，および Value という 3 つのベクトルに変換し，その相関関係を計算して，重み付けされた値を集約する．Query，Key，および Value は入力単語 $\mathbf{E}_{\text{input}}$ にそれぞれの重み行列 $\mathbf{W}^Q$，$\mathbf{W}^K$，$\mathbf{W}^V$ を用いて以下の式で定式化される．

$$\mathbf{Q} = \mathbf{W}^Q \cdot \mathbf{E}_{\text{input}}, \qquad \mathbf{K} = \mathbf{W}^K \cdot \mathbf{E}_{\text{input}}, \qquad \mathbf{V} = \mathbf{W}^V \cdot \mathbf{E}_{\text{input}} \tag{3.5}$$

そして，各トークンのクエリと全キーとの内積にソフトマックス関数を適応し，$\mathbf{V}$ を付加重することによって Attention を求めることができる．

$$\mathit{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_{\text{model}}}}\right)\mathbf{V} \tag{3.6}$$

Multi-Head Attention

Attention は一つの計算を逐次的に行っているため，1 つの視点からしか文脈を読み取ることができない．これに対し，Multi-Head Attention は複数の Attention を並列に連結して出力を得る．これにより，情報の多様な側面を同時に捉えることができる．次元調整するための線形変換行列を $\mathbf{W}^O$ とすると Multi-Head Attention は以下の式で定式化される．

$$\mathit{Multi\text{-}HeadAttention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathit{Concat}(\mathit{head}_1, \mathit{head}_2, \cdots, \mathit{head}_h)\mathbf{W^O} \tag{3.7}$$

$$\text{where } \mathit{head}_i = \mathit{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \tag{3.8}$$

FFNN

Transformer の Encoder と Decoder の最後の層にあるサブレイヤーである．FFNN は，入力データをより高次元の表現に変換するために使用され，モデル内で非線形な変換を行う．FFNN は全結合層，活性化関数 (ReLU)，全結合層の 3 つの層で構成される．

第 1 層（全結合層）

Multi-Head Attention から入力ベクトルを受け取り，高次元の中間層（通常は 2048 次元など）に変換する．高次元に変換することで，モデルが学習する表現空間が広がり，より複雑で多様な関数を学習することができる．

第 2 層 (ReLU 活性化関数)

第 1 層の出力に非線形活性化関数である ReLU を適用する．これにより，非線形性を導入し，モデルがより複雑な関数を学習できるようにする．

第 3 層（全結合層）

高次元の中間層を，元の入力と同じ次元に戻す．この層は，元の入力次元数に戻すことによって，最終的な出力を得る役割がある

FFNN は入力された情報を高次元で表現し，非線形活性化関数を経て，最終的に入力次元に戻すことで，モデルが表現できる情報の範囲を広げることができる．Attention 層は文全体の位置関係を学習するのに優れているが，各トークンの詳細な特徴量を捉えることに欠けている．FFNN によって各トークンが持つ情報量を増やすことができ，入力された特徴量から複雑な関係性を学習することができる [36]．FFNN を通過した後，最終層では学習されたベクトルが出力される．このベクトルを用い，感情分析などをの様々タスクへ応用させる．

モデルの学習方法

BERT は，事前学習を行うことで，一般的な言語の文脈を理解する．ここでは，BERT は大規模なテキストコーパス（Wikipedia や BooksCorpus など）を使って，自己教師あり学習を行う．事前学習では，主に 2 つのタスクが用いられる．

事前学習

事前学習では，BERT は Masked Language Modeling (MLM) と Next Sentence Prediction (NSP) という 2 つのタスクを使ってモデルを訓練する．MLM では，入力文の中でランダムに選ばれた単語を「[MASK]」トークンで置き換え，モデルにその単語を予測させる．MLM において，Multi-Head Attention は文脈を双方向的に考慮し，マスクされた単語を予測する．NSP では，2 つの文が連続しているかどうかを予測する．このタスクによって，BERT は文間の関係を理解し，文章全体の構造を把握できるようになる．例えば，文 A と文 B が連続している場合は「1」を，連続していない場合は「0」を予測する．

ファインチューニング

事前学習後，BERT は特定のタスクにファインチューニングされる．ファインチューニングとは，事前学習で得られた重みをベースに，特定のタスク（感情分析や質問応答）に最適化する工程である．

最終層の出力と感情予測

最後の層で出力する特徴量は解くタスクによって異なる．感情分析においては，BERT の最終層から [CLS] トークンの特徴量を使用する．[CLS] は文章全体の埋め込みベクトルであり，文全体の特徴を含んだベクトルである．出力されたベクトルを，全結合層に入力し，シグモイド関数を用いることで各クラスのスコアを 0〜1 の範囲に正規化され，感情スコアとして出力される [37]．本研究では，BERT を用いて歌詞の感情値 (0〜1 の連続値) を分析する．また，その分析結果を歌詞の感情値とし，楽曲の特徴量として加える．

§ 3.2　画像からの感情分析

ユーザの感情を分析する技術として，顔認証技術を用いた手法がある．顔認証とは，顔の特徴を解析して，特定の人物を識別したり，感情を推定したりする技術のことである．顔認証は，物理的な特徴（目，鼻，口，顔の輪郭など）をもとに，個人を識別するために使用されることが多いが，感情分析では顔の表情に基づいて感情を推定する．本研究では DeepFace を用い，カメラ映像からユーザの 7 つの感情極性 (「怒り」「悲しみ」「喜び」「驚き」「恐れ」「嫌悪」「ニュートラル」) を検出する．そして，それぞれのスコアをもとにユーザのネガティブ強度を算出する．以下より，Deepface のモデルおよび DeepFace の感情分析の流れを説明する．

DeepFace

DeepFace は，Facebook が開発したディープラーニングモデルであり，9 層の畳み込み CNN を基盤を使用して顔の識別を行っている．また，モデルの学習の際に facebook の大量の画像データを学習させることにより，人と同等レベルの認識精度を実現した [38].

Deepface は「顔検出」,「3D アライメント」,「顔特徴の抽出」,「パラメータの学習」の 4 つの工程により感情分析を行うことができる．以下にその流れについて述べる．

顔検出

顔検出は入力映像から顔を検出する処理である．顔検出では，画像内の顔の位置と大きさを特定し，検出された顔をトリミングして次の処理（顔特徴抽出や感情推定）を行う．DeepFace では複数の既存の顔検出モデルを選択することができ (図 3.3),「Haar カスケード分類器」と「Multi-task Cascaded Convolutional Networks(MTCNN)」が一般的に使われる [39].

3D アライメント

顔検出によって得られた顔部分は，位置や角度によってバラバラな場合がある．このバラバラになった顔を正面に向くように補正する処理を顔アライメントという．従来の 2D アライメントでは，顔が回転している場合や非正面向きの場合に十分な補正ができないという課題があった．この課題を解決するため，DeepFace では 3D アライメント技術を導入し

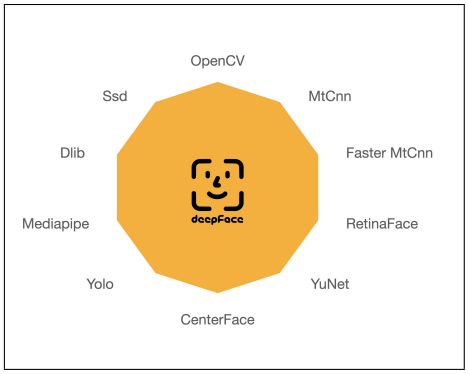

図 3.3: DeepFace の顔検出の種類 [39]

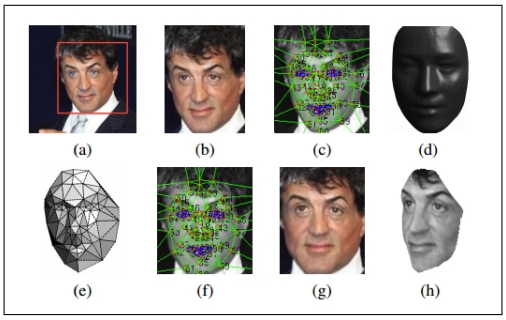

図 3.4: 3D アライメント [38]

ている [38]．3D アライメントの例を図 3.4 に示す．3D アライメントでは，まず顔画像内から 67 個の特徴点を検出する (図 3.4.(c))．平均的な顔の 3D モデル (図 3.4.(d)) を使用し，画像内の特徴量と対応付けを行う (3.4.(f))．対応付けを行った 3D モデルをアフィン変換を行い，2D 平面に投影する (3.4.(g))．アフィン変換とは，画像の拡大縮小，回転，平行移動などを行列を使って座標を変換することである [40]．この処理によって顔認識や感情分析がより高精度に行うことができる．

特徴量の抽出

ここでは，DeepFace の 9 層の CNN を用いて，前工程の処理で抽出した顔の画像から高次元の特徴ベクトルを抽出する工程である．DeepFace は入力層，畳み込み層 (C1)，マックスプーリング層 (M2)，畳み込み層 (C3)，ローカル接続層 (L4，L5，L6)，全結合層 (F7，F8) の 9 層で構成されている [41]．

入力層は，顔画像が読み込まれる層である．入力は，3D アライメントで抽出した正面化された顔画像である．入力の際，顔画像のサイズを 152 × 152 ピクセルに固定され，RGB 画像として 3 チャネルのデータとして入力される．入力画像を X とすると以下のようにあらわされる．

$$X \in \mathbb{R}^{152\times152\times3} \tag{3.9}$$

畳み込み層は，入力画像から局所的な特徴を抽出する処理を行う層である．畳み込み層では，入力画像に対してフィルタ (カーネル) をスライドさせながら畳み込み演算を適応し，特徴マップを作成する．畳み込み層の出力（フィルタ k の位置 (i,j) の値）を $F_{i,j,k}$，畳み込み層のフィルタ k における (m,n) におけるの重みを $W_{m,n}$，入力画像 (i,j) におけるピクセル値を $X_{i,j}$，フィルタ k に対応するバイアス項を b_k と畳み込み演算は以下のようであらわされる．

$$F_{i,j,k} = \sigma\left(\sum_{c=1}^{C}\sum_{m=1}^{H}\sum_{n=1}^{W} W_{m,n,c,k}\cdot X_{i+m,j+n,c} + b_k\right) \tag{3.10}$$

C1 層は，RGB 画像（サイズ $152\times152\times3$）に適用される最初の畳み込み層である．フィルタサイズは 11×11，フィルタ数は 32 である．ここでは，$152\times152\times3$ の画像に対して

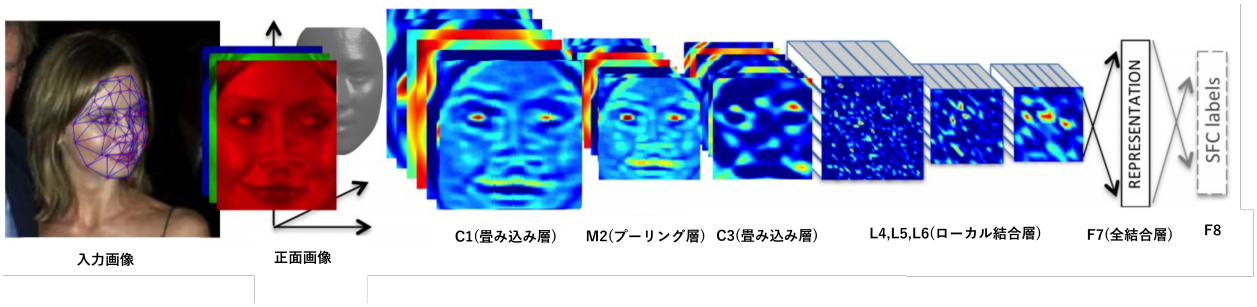

図 3.5: DeepFace のアーキテクチャ[38]

11×11 サイズの 32 個のフィルタをそれぞれ 1 ピクセルごとにスライドさせ，32 個の異なる特徴量を抽出している．また，畳み込み処理を行ったとき，入力画像のサイズを W_{size}，フィルタのサイズを K_{size}，パディングサイズを P_{size}(画像の境界に追加するピクセルの数)，ストライド (フィルタが画像上を移動するステップ数) を S，出力サイズを H_{out} は以下の式で求められる [42].

$$H_{\text{out}} = \frac{W_{\text{size}} - K_{\text{size}} + 2P_{\text{size}}}{S} + 1 \tag{3.11}$$

式 (3.12) により，C1 層では 142×142 の 32 個の異なる特徴マップが出力される [42].

畳み込み層によって得られた出力は活性化関数が適応される．DeepFace では ReLU が用いられている．ReLU は，入力が正の値であればそのまま出力し，負の値に対しては 0 を出力する非線形活性化関数である．これにより，ネットワークが複雑な特徴を学習できるようにし，勾配消失問題を避け，計算効率を向上させる．

マックスプーリング層は C1 層からの出力に対してプーリング操作を行う層である．プーリングは，特徴マップのサイズを縮小する操作であり，計算コストを削減しながら，空間的な情報を効率よく保持するために使用される．M2 層では，主にマックスプーリングという手法が用いられる．マックスプーリングとは，特徴マップを一定のサイズの領域でスライドさせ，その領域内の最大値を選び出して特徴マップを縮小する手法である．

プーリングサイズは 3×3，ストライドは 2 である．ここでは，C1 層からの特徴マップ $142 \times 142 \times 32$ を 3×3 の領域でスライドさせ，各領域内の最大値を選択することで，出力の特徴マップのサイズを縮小する．縮小サイズは式 (3.12) により $71 \times 71 \times 32$ に縮小する．C3 層は，C1 層から得られた特徴マップ適用される畳み込み層である．フィルタサイズは 9×9，フィルタ数は 16 である．これを入力特徴マップ (サイズ $71 \times 71 \times 32$) に対して局所的な特徴を抽出する．縮小サイズは式 (3.12) により $63 \times 63 \times 16$ となる．これにより，前の層で得られた特徴をさらに複雑なパターンに変換する．

ローカル結合層は各入力位置に対して異なるセットのフィルタを適応する層である．畳み込み層は，全ての位置で同じフィルタを適用するが，ローカル結合層では，位置に対応するフィルタ重みがその位置固有のものになる． 出力特徴マップの位置 (i, j) とチャネル k における値を $h_{i,j,k}$，入力特徴マップの位置 $(i+m, j+n)$ とチャネル c における値を $X_{i+m,j+n,c}$，各空間位置 (i, j) に固有のフィルタ重みを $W_{i,j,m,n,c,k}$，各空間位置 (i, j) に固有のバイアス項を $b_{i,j,k}$，フィルタの高さと幅を H_f, W_f，入力特徴マップのチャンネル数をとすると以下の式で求められる．

$$h_{i,j,k} = \sum_{m=1}^{H_f} \sum_{n=1}^{W_f} \sum_{c=1}^{C} X_{i+m,j+n,c} \cdot W_{i,j,m,n,c,k} + b_{i,j,k} \tag{3.12}$$

L4 のフィルタは 9×9 であり出力の特徴マップのサイズは縮小サイズは式 (3.12) により 55×55 となる．また，L5 のフィルタは 7×7 であり，出力の特徴マップのサイズは縮小サイズは式 (3.12) により 25×25 となる．L6 のフィルタは 5×5 であり，縮小サイズは式 (3.12) により 21×21 となる．

全結合層は特徴抽出後のデータを一つの固定長ベクトルに変換する層である．これにより，次のタスク (分類タスクや識別タスク) に適した形に変換される．F7 層は，C1〜L6 層を通じて抽出された高次元の特徴を統合し，顔画像を表現する 4096 次元の特徴ベクトルを生成する層である．このベクトルは，入力画像の識別に必要な情報を凝縮したものであり，次の F8 層の分類や顔画像間の類似度計算に利用される．F7 層の出力ノード i の値を u_i，L6 層からの入力ノード j の値を h_j，入力ノード j と出力ノード i を結ぶ重みを $W_{j,i}$，出力ノード i に対応するバイアスを b_i，L6 層の出力をフラット化した入力ノードの総数を N とすると以下のようにあらわされる．

$$u_i = \sum_{j=1}^{N} h_j \cdot W_{j,i} + b_i \tag{3.13}$$

出力は 4096 次元のベクトルであり，このベクトルは，成分ごとの最大値正規化と L2 正規化法を用いて 0〜1 の値に正規化される．これによりモデルの過学習を防ぐことができ，精度を高めることができる．

F8 層は 7 層の出力をもとに 7 つのカテゴリ (怒り，嫌悪，恐怖，幸せ，悲しみ，驚き，中立) のいずれかに分類する層である．7 層で出力された 4096 次元のベクトルを入力とし，各感情に対するスコアを求め，softomax 関数により確率に変換している．F7 層からの特徴ベクトルの成分 i を u_i，F7 層の出力ノード i と感情カテゴリ k を結ぶ重みを $W_{i,k}$，感情カテゴリ k に対応するバイアスを b_k，感情カテゴリのインデックスを k とした時，感情カテゴリ k に対応するスコアを O_k は以下のようにあらわされる

$$O_k = \sum_{i=1}^{4096} u_i \cdot W_{i,k} + b_k \tag{3.14}$$

各感情のスコア O_k を感情の確率 p_k に変換するため，softmax 関数を適応する．確率 p_k が最も高いカテゴリ k が，感情分析の最終的な出力となる．

$$p_k = \frac{\exp(O_k)}{\sum_{k=1}^{7} \exp(O_j)} \tag{3.15}$$

パラメータの学習

感情分析のパラメータの学習には，クロスエントロピー損失を使用する．クロスエントロピー損失とは，モデルの予測値と正解ラベルとの間の誤差を定量化する関数のことであり，この関数を最小化することによりモデルの予測精度を向上させている．正解の確率分布を y_k とするとクロスエントロピー損失関数 L は以下で求められる．

$$L = -\sum_{k=1}^{7} y_k \log(p_k) \tag{3.16}$$

以上のように，DeepFace により，顔映像から顔を検出し，9 層の CNN によって顔を 4098 次元の特徴ベクトルに変換され，この特徴量をもとに各感情の確率を出力している．本研究では，DeepFace を用い，カメラ映像によりユーザの各感情の確率を感情スコアとして，各感情に重みを付けることによりネガティブ強度を算出する．また，ネガティブ強度に応じた楽曲を提示し，ネガティブ強度の推移を計測する．

§ 3.3　トピックモデルによるトピック分類

本節では，歌詞の主題を推定するために用いるトピックモデルについて説明する．トピックモデルとは，文章中の単語は文章の潜在的な意味 (トピック) に依存して出現すると仮定し，文章中に出現する単語の頻度からトピックを推定する手法である．トピックモデルの代表として潜在ディリクレ配分法（Latent Dirichlet Allocation：LDA）がある．本研究では，歌詞のトピックを LDA を拡張した，ガイド付き LDA(Guided Latent Dirichlet Allocation) 用いて推定する．また，推定したトピックごとに楽曲をクラスタリングし，トピックごとのプレイリストを作成する．LDA のイメージを図 3.6 に示す．

LDA

LDA は，文書の潜在的なトピック構造を発見するための確率的生成モデルである [43]. LDA は，文章がどのように生成されるかを仮定し，文書中に出現する単語の頻度からそのトピックを推定し，トピックを分類する手法である．

LDA のグラフィカルモデルは図 3.7 に示す．グラフィカルモデルとは確率変数の関係をグラフで表現したモデルであり，図 3.7 において W は観測される文章の単語を表し，K はトピック数，M は文章の数 (本研究では歌詞の数であり，楽曲の数でもある)，N は文章内の単語数，ϕ はトピックが持つ単語分布，θ は文章が持つトピック分布を表す．α，β はディリクレ事前分布のパラメータであり，ハイパーパラメータと呼ぶ．LDA の生成モデルは，以下の過程を経て文章が生成されると仮定する．

トピックの分布

各文章 D_i は，いくつかのトピックから成り立っていると仮定する．例えば，ニュース記事では,『政治』『経済』『スポーツ』などのトピックが含まれ，それぞれのトピックが文書に対してどれくらい関係しているかを示す分布が存在する．それらのトピックは，ディリクレ分布から得られると仮定する．

単語の分布

各トピック k は，そのトピックに関連する特定の単語の分布を持っていると仮定する．たとえば,「スポーツ」のトピックでは「試合」「選手」「チーム」などの単語が頻繁に現れる．トピックごとの単語分布もディリクレ分布から生成される．

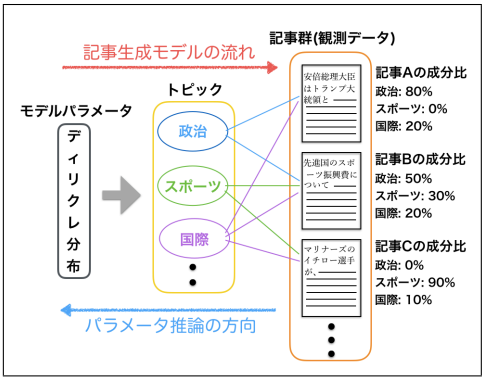

図 3.6: LDA のイメージ [44]

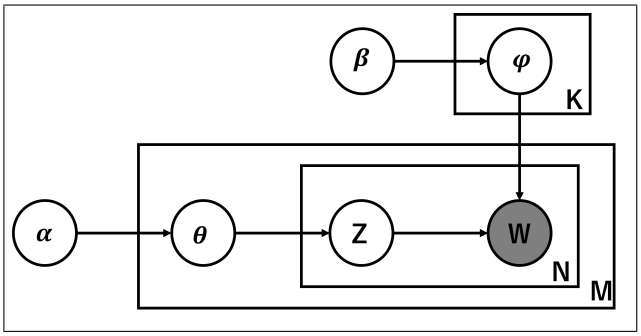

図 3.7: LDA におけるグラフィカルモデル

単語の生成

各文書の単語は，まずその単語が関連するトピックを確率的に選び，そのトピックに基づく単語分布からサンプリングされて生成される．たとえば,「試合」という単語が文書内で出現した場合，それが「スポーツ」トピックから生成される確率が高いと推測される．

LDA ではディリクレ分布を用いて確率分布（トピック分布，単語分布）を生成し，それらをもとに多項分布を使い，単語やトピックを生成する．そして，実際の文書中の単語の頻度をもとに，各文書のトピック分布やトピックごとの単語分布を逆推定し，文章のクラスタリングを行う．文章生成のプロセスは以下の流れで行われている．

トピック分布と単語分布の作成

各文章 D について，トピックの数を K，トピック分布のハイパーパラメータを $\alpha = (\alpha_1, \alpha_2, ..., \alpha_K)$ としたとき，文章 D が持つトピック分布 θ_D はディリクレ分布 (Dirichlet) からランダムに生成される．

$$\theta_D \sim Dirichlet(\alpha) \tag{3.17}$$

同様に，単語分布のハイパーパラメータを β としたとき，トピック k が持つ単語分布 ϕ_k はディリクレ分布からランダムに生成される．

$$\phi_k \sim Dirichlet(\beta) \tag{3.18}$$

ディリクレ分布とは各要素が確率値をとる（非負であり和が 1 となる）K 次元ベクトルに対する分布である．また，ハイパーパラメータ α におけるディリクレ分布の確率密度関数は以下のようにあらわされる．

$$p(x \mid \alpha) = \frac{1}{B(\alpha)} \prod_{k=1}^{K} x_k^{\alpha_k - 1} \tag{3.19}$$

ここで $B(\alpha)$ はベータ関数であり，ディリクレ分布を確率分布（確率の合計が 1 になる分布）として正規化する定数である．ガンマ関数を Γ としたとき，ベータ関数は以下のようにあ

らわされる．

$$B(\alpha) = \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)} \tag{3.20}$$

このようにして各文書が複数のトピックにどの程度関連しているかを示すトピック分布を，ディリクレ分布を使ってランダムに決定する．

単語の生成

LDA のモデルは，各文章が複数のトピックから成り立っているという仮定のもとで文章内の単語がどのように生成されるかをモデル化する．文章内の単語を生成する流れは,「トピックの選択」,「単語の生成」の流れで行われる．文章 D に含まれる単語がどのトピックから生成されるかは，文章全体のトピック分布 θ_D に基づいて確率的に決まる．文章 D 内の n 番目の単語に対応するトピックを $z_{D,n}$ としたとき，$z_{D,n}$ は多項分布 (Multinomial) をもとに選択される．

$$z_{D,n} \sim Multinomial(\theta_D) \tag{3.21}$$

トピックが選ばれた後，選択されたトピックに基づいて，単語を生成する．トピック $z_{D,n}$ に対応する単語分布を $\phi_{z_{D,n}}$ としたとき，文章 D の n 番目の単語 $w_{D,n}$ は多項分布をもとに生成される．

$$w_{D,n} \sim Multinomial(\phi_{z_{D,n}}) \tag{3.22}$$

多項分布とは，確率論における離散的な確率分布の一つで，ある事象が複数回の試行にわたって発生する場合に，各結果が発生する回数の確率をモデル化するである．LDA においては，文書内の単語がどのトピックから生成されるかの確率や，選ばれたトピックから単語が生成される確率を，多項分布を用いてモデル化する．文章内の単語の総数を N，カテゴリー k に属する単語の出現回数を x_k，トピック k が選ばれる確率を p_k としたとき，多項分布の確率質量関数は以下のようにあらわされる．

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_K = x_K) = \frac{N!}{x_1! x_2! \ldots x_K!} p_1^{x_1} p_2^{x_2} \ldots p_K^{x_K} \tag{3.23}$$

LDA は，この生成過程を逆方向に実行することで，実際の文書からトピック分布 θ および単語分布 ϕ，$z_{D,n}$ を推定する．

パラメータの推定方法

LDA のパラメータ推定では，実際の文章集合 w からトピック生成で仮定した θ，ϕ，$z_{D,n}$ を推定する．θ，ϕ，$z_{D,n}$ は直接観測することができない潜在変数であるため，観測データ w に基づいた θ，z の事後分布によって求めることができる．事後分布とは観測データを考慮したうえで，未知のパラメータがどのような分布を持つかあらわしたものある．$p(w, \theta, \phi, z \mid \alpha, \beta)$ を結合確率とすると以下の式であらわされる．

$$p(\theta, z, \phi \mid w, \alpha, \beta) = \frac{p(w, \theta, \phi, z \mid \alpha, \beta)}{p(w \mid \alpha, \beta)} \tag{3.24}$$

$p(w \mid \alpha, \beta)$ は観測データの周辺尤度であり，モデルから実測データ w が出現する尤もらしさである．$p(w \mid \alpha, \beta)$ における以下の式であらわされる．

$$p(w \mid \alpha, \beta) = \int \sum_z p(\theta, z, w \mid \alpha, \beta)\, d\theta \tag{3.25}$$

パラメータは，$p(w \mid \alpha, \beta)$ の対数をとった以下の式

$$\log p(w \mid \alpha, \beta) = \log \int \sum_z p(\theta, z, w \mid \alpha, \beta)\, d\theta \tag{3.26}$$

を最大化することで推定を行うことができる．対数周辺尤度を最大化することで，観測データ w を生成する尤もらしさが最大化され，そのデータを最もよく説明するパラメータが得られるからである．しかし，$p(w \mid, \alpha, \beta)$ は高次元であるため，直接計算するのが非常に困難である．そのため，イェンゼンの不等式を利用し，その下限を最大化することで求める．イェンゼンの不等式とは上に凸な関数 $f(x)$ について以下の不等式が成り立つ式である．

$$f(E[x]) \geqq E[f(x)] \tag{3.27}$$

イェンゼンの不等式を用いると，式 (3.26) は以下の通りに表される．

$$\begin{aligned} \log p(w \mid \alpha, \beta) &= \log \int \sum_z p(\theta, z, w \mid \alpha, \beta)\, d\theta \\ &= \log \int \sum_z q(\theta, z \mid \gamma, \phi) \frac{p(\theta, z, w \mid \alpha, \beta)}{q(\theta, z \mid \gamma, \phi)}\, d\theta \\ &\geqq \int \sum_z q(\theta, z \mid \gamma, \phi) \log \frac{p(\theta, z, w \mid \alpha, \beta)}{q(\theta, z \mid \gamma, \phi)}\, d\theta \end{aligned} \tag{3.28}$$

ここで，事後分布を近似するために，近似分布 $q(\theta, z \mid \gamma, \phi)$ を導入する．この近似分布は，事後分布に近い分布を選び，そのパラメータを最適化することで，推定を行う．LDA の場合，γ を各文書のトピック分布 θ を近似するディリクレ分布のパラメータ，ϕ_n を単語 w_n のトピック割り当て確率を近似するカテゴリ分布のパラメータとし，近似分布は以下のように設定する．

$$q(\theta, z \mid \gamma, \phi) \simeq p(w \mid \alpha, \beta) \tag{3.29}$$

$$q(\theta, z \mid \gamma, \phi) = q(\theta \mid \gamma) \prod_{n=1}^{N} q(z_n \mid \phi_n) \tag{3.30}$$

式 (3.28) の右辺は対数事後確率の下限であり ELBO（Evidence Lower Bound，証拠下限）と呼ばれ，ELBO を $\mathcal{L}(q)$ とすると，$\mathcal{L}(q)$ は以下の通りとなる．

$$\begin{aligned} \mathcal{L}(q) &= \int \sum_z q(\theta, z \mid \gamma, \phi) \log \frac{p(\theta, z, w \mid \alpha, \beta)}{q(\theta, z \mid \gamma, \phi)}\, d\theta \\ &= \mathbb{E}_q[\log p(\theta, z, w | \alpha, \beta)] - \mathbb{E}_q[\log q(\theta, z \mid \gamma, \phi)] \end{aligned} \tag{3.31}$$

ここで，$p(\theta, z, w \mid \alpha, \beta)$ は LDA の生成モデルの確率，$q(\theta, z \mid \gamma, \phi)$ は近似分布であり，$\mathbb{E}_q$ は近似分布 q に基づく期待値を表す．$\mathcal{L}(q)$ を最大化することで，近似分布 q が事後分布に近づく．

そして，近似分布 $q(\theta, z \mid \gamma, \phi)$ のパラメータは反復的に更新される．単語 w_n が特定のトピック i に属する確率 ϕ_{ni} は ディガンマ関数を Ψ とすると以下のように更新される．

$$\phi_{ni} \propto \beta_{i,w_n} \exp(\Psi(\gamma_i) - \Psi(\sum_k \gamma_k)). \tag{3.32}$$

また，トピック分布 θ の近似分布 γ_i は，以下のように更新される

$$\gamma_i = \alpha_i + \sum_{n=1}^{N} \phi_{ni}. \tag{3.33}$$

$\mathcal{L}(q)$ を最大化するために，近似分布 $q(\theta, z \mid \gamma, \phi)$ のパラメータを繰り返し更新する．収束判定には，$\mathcal{L}(q)$ の増加量が小さくなった時点で更新を停止する．収束後，得られた変分分布から，トピック分布 θ や単語分布 ϕ の推定値を得ることができる．また，この推定値より $z_{d,n}$ を算出することができる．

ガイド付き LDA

通常の LDA では，トピック分布や単語分布が文章の単語の頻度に基づいて自動的に学習される．トピックはデータに基づいて自動的に抽出されるため，必ずしも解釈しやすい結果が得られるとは限らない．

この問題を解決するために半教師あり LDA であるガイド付き LDA を用いる．ガイド付き LDA は，特定のトピックに関連する単語を予約語事前に指定することでトピック抽出を制御する LDA である．これにより，各トピックに代表語として分類される単語は，ガイド単語とそれに共起する単語である可能性が高くなる．そのため，トピックの分類結果に明確な視点を導入することが可能になる．

ガイド付き LDA も LDA と同様に変分ベイズ法を使ってトピック分布と単語分布を推定するが，大きな違いは，各トピックの単語分布の初期値を予約語に基づいて設定する点である．この初期値に基づいてモデルを学習することで，予約語に関連する歌詞が高い確率で指定されたトピックに関連付けられるようになる．この手法を使うことで，完全な教師付き歌詞データを準備しなくても，歌詞を任意のトピックに分類できる．

本研究では，歌詞のトピック推定にガイド付き LDA を活用する．具体的には，事前に設定した予約語を使って歌詞を任意のトピックに分類する．このトピック分類を活用することで，特定のテーマに基づいたプレイリストを作成する．

第4章

提案手法

§ 4.1　楽曲特徴量の取得の流れ

本研究では，ユーザのネガティブ強度に合った楽曲を推薦するため，Web 上から楽曲の特徴量を取得する．使用する特徴量はテンポ，曲調，アコースティック性，楽曲の感情値である．本研究ではテンポ，曲調，アコースティック性は SpotifyAPI を用いて取得し，その後 Genius API を用いて楽曲の歌詞データを取得する．また，取得した歌詞データをもとに BERT を用いて歌詞の感情スコア計算し，それを歌詞の感情値とする．

SpotifyAPI を用いた楽曲特徴量の取得

SpotifyAPI とは，音楽ストリーミングサービス Spotify の豊富なデータをプログラムでアクセスできる API である．API とは，異なるソフトウェアや Web サービスを接続するインターフェースのことである．API を用いることに，システムやサービスが提供する機能やデータにアクセスできるようにし，他のシステムがその機能やデータを利用できるようになる [47]．SpotifyAPI を用いることにより，楽曲やアーティスト，プレイリスト，アルバムをはじめとした楽曲のメタデータだけでなく，楽曲の音楽的な特徴量を取得することができる．

具体的な取得の流れについて説明する．まず，SpotifyAPI を利用するため，Spotify Developer Dashboard でアプリを登録し，クライアント ID とクライアントシークレットを取得する．取得ができたら，Python のモジュールである「spotipy」を利用して楽曲の特徴量を取得する．spotipy とは，SpotifyAPI を Python から利用するためのライブラリであり，クライアント ID とクライアントシークレットを使用することでユーザ認証，データアクセス，楽曲検索，プレイリストの管理が可能となる．ここでは，取得したクライアント ID とクライアントシークレットを使用し，楽曲の特徴量を取得する．

SpotifyAPI で取得できる楽曲特徴量として,「テンポ」,「曲調」,「アコースティック性」,「エネルギー」,「音楽長」,「ダンス性」,「ライブ性」などがある．取得できる楽曲特徴量の例を図 4.2 に示す．テンポとは，曲の 1 分あたりの拍数であり，1 分間に何回の拍が楽節の中で起こるかを数値化したもである．曲調とは，楽曲全体の雰囲気や性格を表す．曲調は主に長調 (メジャーキー) と短調 (マイナーキー) に分類され，メジャーキーは明るく，ポジティブな雰囲気があり，マイナーキーは暗く，落ち着いた雰囲気がある．メジャーキーであれば 1，マイナーキーであれば 0 で取得することができる．アコースティック性とは曲がどれだけ自然な楽器の音，または生音に近いかを示す指標である．アコースティック性が高い音楽はアコースティックギター，ピアノ，弦楽器などの生楽器を主体とした編成となって

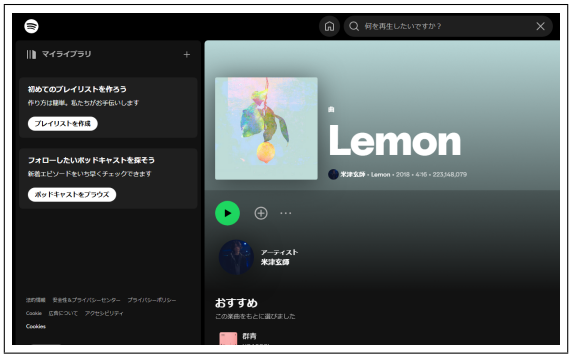

図 4.1: Spotify のホームページ [48]


```
{'acousticness': 0.2239,
 'analysis_url': 'https://api.spotify.com/v1/audio-analysis/6KDmk0EF3Qk7zJfDDtRoF2',
 'danceability': 0.873,
 'duration_ms': 241680,
 'energy': 0.755,
 'id': '6KDmk0EF3Qk7zJfDDtRoF2',
 'instrumentalness': 0,
 'key': 7,
 'liveness': 0.336,
 'loudness': -3.352,
 'mode': 0,
 'speechiness': 0.135,
 'tempo': 122.01,
 'time_signature': 4,
 'track_href': 'https://api.spotify.com/v1/tracks/6KDmk0EF3Qk7zJfDDtRoF2',
 'type': 'audio_features',
 'uri': 'spotify:track:6KDmk0EF3Qk7zJfDDtRoF2'
}
```


図 4.2: 取得できる楽曲特徴量の例

おり，アコースティック性が低い音楽はシンセサイザー，ドラムマシン，エレクトロニックサウンドを主体とした編成であり，デジタル音源が中心となっている．アコースティック性は 0〜1 の連続値で取得することができ，1 に近いほどアコースティック性が高い楽曲であることを示す．

エネルギーは音楽が持つ「活力」や「力強さ」を示す指標であり，0〜1 の連続値で取得することができ，1 に近いほど活発で躍動感のある楽曲であることを示す．音楽長は音楽の長さであり，ミリセカンド単位での時間長で取得することができる．ダンス性はその楽曲がダンスに適した音楽家を示す特徴量であり，0〜1 の連続値で取得することができ，1 に近いほどダンスに適している楽曲であることを示す．ライブ性はライブ音源らしさを示す特徴量であり，0〜1 の連続値で取得することができ，1 に近いほどライブで演奏されている可能性が高い楽曲であることを示す．

2.1 節でも述べた通り，楽曲が持つ特徴量の中でも，特にテンポ，曲調，アコースティック性，エネルギーの特徴量が感情に影響を与えるとされているため，それらの特徴量を取得した．また，SpotifyAPI では楽曲の歌詞を取得することができない．そのため，取得した楽曲の特徴量を取得すると同時に，アーティスト名と楽曲名を取得した．また，取得したデータを加工しやすくするように CSV ファイルに保存した．

Genius API を用いた歌詞データの取得

先ほども述べたように，SpotifyAPI では楽曲の歌詞を取得することができない．そのため，GeniusAPI を用い，SpotifyAPI で取得したアーティスト名と楽曲名をもとに楽曲の歌詞を取得する．Genius は楽曲の歌詞や関連情報を提供するサービスであり，この API を用いることで指定されたアーティスト名と楽曲名に基づき，対応する歌詞を取得することができる．

具体的な流れを説明する．まず，GeniusAPI を利用するためには，Genius のアカウントが必要である．そのため，Genius にアクセスし，アカウントを作成する．そこで API キーを取得する．次に，取得した API キーを使用し，SpotifyAPI で取得した楽曲名とアーティスト名をもとに，GeniusAPI へリクエストを送信する．リクエストは Python の requests モジュールを用い行う．requests とは Python で使うためのモジュールであり，HTTP リクエストを簡単に送信するためのライブラリである．Spotify では Spotify から直接データを取得することができる spotipy という専用のモジュールがあったが，Genius には専用のモジュールがないため，requests を用い，歌詞の検索結果の HTML データをそのまま抽出する．

図 4.3: Genius のホームページ [49]

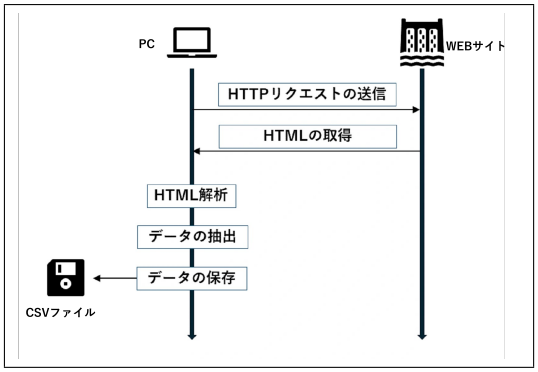

図 4.4: スクレイピングの流れ [50]

また，GeniusAPI は直接歌詞データを提供しないため，取得した歌詞ページ URL をもとにスクレイピングを行う．スクレイピングには，Python の BeautifulSoup4 というライブラリを使用し，HTML データから歌詞部分を抽出する．スクレイピングのイメージを図 4.4 に示す．BeautifulSoup4 とは，Python 上で実装できるスクレイピング用ライブラリの 1 つである．BeautifulSoup4 に HTML データを渡すと，それらのドキュメントをパースし，要素を検索，操作，抽出することができる．タグ指定やテキスト指定で自分が必要とする情報を取得できる．ここでは BeautifulSoup4 を用い，Genius の歌詞が含まれる HTML タグ (class_=”lyrics”) を指定し，歌詞部分を取得する．このとき同時に取得される [Intro] や [Verse 1] などの歌詞に関係ないデータは除去する．

スクレイングにおいて，一度に大量のリクエストを要求するとリクエスト制限に達してしまい，歌詞を取得できなくなってしまう．そこで，一度に歌詞の HTML データを取得してから次の HTML データを取得するまでに 5 秒間待機するように設定した．このようにして抽出した歌詞データを，楽曲の特徴量が保存されている CSV ファイルに追加する．

BERT を用いた歌詞の感情値の取得

次に，GeniusAPI で取得した歌詞データをもとに，BERT を用い，歌詞の感情値を取得する．具体的には，取得した歌詞を BERT モデルに入力し，その出力を基に感情分析を行う．BERT の出力は 0〜1 の連続値であり，1 に近いほどポジティブな歌詞であり，0 に近いほどネガティブな歌詞であると分析できる．この BERT による出力を歌詞の感情値として取得する．

具体的な取得の流れを説明する．まず，感情分析を行うためにラベル付き正解データを収集する．感情分析におけるラベル付き正解データは感情ラベルが付与された文章のデータセットである．本研究では正解データセットとして「chABSA-dataset」を使用し，モデルを学習させる．chABSA-dataset は TIS が提供している感情分析を行うためのデータセットであり，上場企業の有価証券報告書 (2016 年度) をベースに作成されたものである．各文に対してネガティブ・ポジティブの感情分類だけでなく,「何が」ネガティブ・ポジティブなのかという観点を表す情報が含まれいる．こうした観点単位の感情分類を BERT モデルに学習させることで，より高度な分析を行うことができる [51] [52].

次に，収集したデータセットを用いて，BERT モデルを学習させる．事前学習モデルは Hugging Face や GitHub などのサイト公開されており，東京大学や京都大学なども独自の

曲名	アーティス	楽曲url	id	アコーステ	エネルギー	コード	テンポ	感情値	歌詞
夏祭り	Jitterin' Ji	https://a	07h39XEN	0.00842	0.875	0	140.897	0.957	君がいた夏は遠い夢の中 あぁ 空に消えて
どんなときも。	Noriyuki I	https://a	4feDzcxt	0.344	0.809	1	123.827	0.915	僕の背中は自分が 思うより正直かい？ 誰
僕の彼女はウコ	Noriyuki I	https://a	6yDwkmi	0.13	0.787	1	120.017	0.8	留守番電話のメッセージ バイトの途中耳
Rain	Senri Oe	https://a	1vIqve7dH	0.31	0.632	1	96.033	0.685	言葉にできず凍えたままで 人前ではやさ
風	Yuji Oda	https://a	5734o6Ba	0.0137	0.886	1	122.935	0.479	すみれ色のまま夕暮れを止めて 新しい自
ロマンスの神様	Kohmi Hi	https://a	0xYSgQH	0.149	0.802	1	125.91	0.556	勇気と愛が世界を救う 絶対いつか出会え
微笑みの爆弾	Matsuko	https://a	4J91rrMG	0.0873	0.827	0	128.5	0.808	都会の人ごみ　肩がぶつかって　ひとりぼ
Choo Choo TI	ZOO	https://a	1XS2YCE	0.0132	0.968	1	115.036	0.879	Fun Fun We hit the step step 同じ風の
恋はくえすちょ	Onyanko	https://a	3QIqA4Xi	0.168	0.572	1	152.307	0.913	く・く・く・く・くえすちょん 恋のしくみ
Dear	Kohmi Hi	https://a	500nTtsT	0.383	0.694	1	89.838	0.324	To read the lyrics, click on the song you
アンブレラ・コ	Onyanko	https://a	4NmJAWI	0.114	0.712	0	95.202	0.918	急に振り出した雨 天気予報もはずれ 逃げ
渚の『・・・	Ushiroyub	https://a	3UirTmI6	0.106	0.743	1	149.855	0.887	やってくれますね やってくれますね あな
うしろゆびさ	Ushiroyub	https://a	2zA6aHqc	0.146	0.933	1	160.004	0.663	長い渡り廊下で あの人と すれ違う度 心
夏休みは終わ	Onyanko	https://a	0IsvV1Wi	0.0697	0.899	1	135.255	0.656	走るバスの窓から 君は身を乗り出し ずっ
セーラー服を	Onyanko	https://a	5qcLX6W	0.264	0.742	1	121.02	0.964	セーラー服を脱がさないで 今はダメよ 我
LIKE A CHER	Onyanko	https://a	3U9xPmJ	0.154	0.746	1	165.06	0.584	眠い目をこすり 息を弾ませて 滑り込む 朝

図 4.5: 取得した楽曲特徴量

モデルを公開している．本研究では Hugging Face に登録されている「cl-tohoku/bert-base-japanese-v2 」を用いる．bert-base-japanese-v2 は東北大学の NLP が公開している BERT の事前学習モデルであり，CC-100 データセットと日本語版 Wikipedia でトレーニングされている．アーキテクチャはオリジナルの BERT モデルと同じ 12 層の中間層，768 次元の隠れ層で構成されており，12 個の AttentionHead で構成されている [53].

以下より，BERT モデルによる歌詞の感情値推定の分析の流れを説明する．まず，収集した歌詞データの前処理を行う．具体的には，〇や☆，無駄なスペースなどの分析に不必要な特殊文字を取り除く．また，BERT は一度の分析に最大で 512 トークンしか分析できないため，1 つの歌詞を複数に分割する．

BERT はトークンという単位でテキストを処理するため，歌詞のテキストを BERT のトークナイザーで変換する．本研究では，BertJapaneseTokenizer を用いてトークン化する．BertJapaneseTokenizer では Mecab を使用した形態素解析を行ない，歌詞を単語に分割した後，WordPiece でサブワード分割し，トークン番号に変換する．トークナイザーは，歌詞テキストを ID のリストに変換し，これが BERT の入力となる．

BERT のモデルは，先述した通り，事前学習済みの BERT モデルである「cl-tohoku/bert-base-japanese-v2」を使用した．このモデルに，感情分析用のデータセット「chABSA-dataset」を使用しファインチューニングさせることによって感情分析に適応させる．

最後に，ファインチューニングし，感情分析に適応させた BERT モデルを用いて，歌詞の感情を分析した．各歌詞に対してポジティブクラスの確率を歌詞の感情値として出力した．先述した通り，長い歌詞は BERT では一度に分析することができない．そのため，分割した歌詞の分析結果の平均値を歌詞の全体的な感情値とした．以上のように取得できた歌詞のメタデータおよび楽曲特徴量を CSV ファイルに保存し，パラメータとして活用した．取得した楽曲特徴量を図 4.5 に示す．

§ 4.2　半教師ありLDAと楽曲特徴量によるプレイリスト作成

本研究では楽曲の特徴量を用いた楽曲推薦に加えて，プレイリストを作成し，ユーザが任意に選択できるようにした．具体的には，ガイド付きLDAを用いて歌詞を予約語に基づいてトピックに分類し，分類したトピックごとにプレイリストを作成した．ガイド付きLDAは，あらかじめ重要な単語を予約語として各トピックに割り当てておくLDA手法であり，これにより，各トピックに代表語として分類される単語は，ガイド単語とそれに共起する単語である可能性が高くなる．そのため，トピックの分類結果に明確な視点を持たせることができる．以下に，ガイド付きLDAによる分析方法とプレイリスト作成の流れを示す．

データの前処理

分析した歌詞は，先述した通りGeniusAPIを用いて取得した歌詞を利用する．ここで，日本語と英語が混在している歌詞は，英語詞を除去し，日本語詞のみを分析対象とした．収集した歌詞データを日本語の形態素解析ツールを用いて単語単位のリストに変換する．本研究ではjanomeというモジュールを用いて単語ごとに分割を行った．janomeは形態素解析を行うPythonのライブラリである．同じ形態素解析ツールとしてMeCabが存在するが，janomeと同等の精度を持ち，janomeの方が容易に使用できるため，janomeを採用した．単語単位に分割した歌詞を分析する前に，前処理を行う．まずはじめに，シンボルの除去を行う．使用する歌詞には,「?」や「!」等の記号や,「”」や「’」等の符号に加え，絵文字や顔文字なども含まれている場合があるため，除外していく．

次に，ストップワードを定義し除去を行う．ストップワードとは頻繁に現れ，テキスト検索処理に関連する内容をもたない単語のことであり，除去を行うことで分析精度を向上させることができる．本研究で設定したストップワードは，ストップワードを設定せずにLDAを適用した際，頻出した単語のうちトピック生成に関係しないものを設定した．本研究で設定した具体的なストップワードを表4.1に示す．

辞書とコーパスの作成

前処理が終わった後，LDAのモデルに必要な辞書とコーパスを作成する．辞書とは歌詞に含まれる全ての単語を一意のIDに対応させたものであり，コーパスは各歌詞を単語のIDとその出現頻度に変換したものである．辞書とコーパスのイメージを図4.6に示す．本研究では，Bag of Words(BoW)形式を採用して，歌詞のコーパスを作成した．BoWは，文章内の単語の頻度を用いて文章を数値ベクトルに変換する方法であり，このようにして得られた辞書とコーパスを用いてガイド付きLDAに利用する．

トピック数と予約語の設定

最初に，歌詞を任意のトピックに分類するためにトピックのカテゴリを事前に決定する．本研究では，一般的によく使われるテーマである「恋愛」,「青春」,「応援」,「応援」,「友情」,「孤独・絶望」,「夢・目標」,「人生」,「ストーリー」,「季節・風景描写」,「地名場所」とした．

「恋愛」は，愛や恋，思いといった人間関係における感情を中心に描かれるトピックであり，歌詞では，恋の喜びや悲しみ，切なさ，運命的な出会いなどが主題になることが多

表 4.1: 設定したストップワード

する	ある	ない	いる	なる	俺	あの	それ	もう	この
僕	あなた	君	私	僕ら	から	れる	られる	せる	その

い.「青春」は若さ特有のエネルギーや希望，挑戦を象徴するトピックであり，青春を題材にした曲は，成長の瞬間や忘れられない記憶に焦点を当てるものが多い.「応援」は聞き手にエールを送るようなポジティブなメッセージが含まれるトピックであり挑戦を乗り越える勇気や努力への励ましを描く.「友情」は友達や仲間との関係性や絆を強調するトピックであり，互いを支え合う信頼感や共感が中心的なテーマとなる.

「孤独・絶望」は悲しみや切なさ，心の痛みを表現するトピックであり，人生の中で感じる孤独や失望が歌詞の主題となる.「夢・目標」は将来の目標や夢に向かう情熱や希望を描いたトピックである.「人生」は人生そのものの意味や選択，運命をテーマとしたトピックであり，生きることに対する哲学的な考えが含まれることが多い.「ストーリー」は歌詞の世界を１つの物語として捉えてそれをテーマとしたものである.「季節・情景描写」は自然や季節を描写し，その中に感情を重ねるトピック．四季折々の風景を通して心情を表現することが多い.「地名・場所」は地名や特定の場所を舞台にした歌詞に関連するトピック．具体的な場所を通して，感情やストーリーが展開される.

次に，各トピックに関連する予約語を設定する．予約語は，以下のような基準で選定する.

1. 各トピックに直接的な意味で関連する単語を選ぶ
2. データセット全体における単語の出現頻度を参考にする．頻度が高すぎる一般的な単語は除外する
3. トピックの内容を豊かにするため，類似した単語ばかりではなく，広がりのある単語を選ぶ

各トピックに対して，上記の基準を基に 10 個の予約語を選定した．設定した予約語は図 4.2 に示す．予約語はガイド付き LDA の入力パラメータとして使用される．各トピックに関連する単語を予約語として設定することで，トピックごとに分布が誘導される．これにより，トピックモデルが予約語を含む単語群を特定のトピックに割り当てやすくなり，結果として得られるトピックの意味が明確化される．この設定により，歌詞の内容に基づいた任意のトピックの分類を行うことができる.

初期単語分布の作成

通常の LDA では，トピックの単語分布がデータに基づいて完全に自動的に推定される．一方，ガイド付き LDA では，予約語が特定のトピックに関連付けられるように，初期単語分布を設定する．予約語による初期単語分布の作成方法を説明する．まず，事前に作成した辞書に基づいて，予約語がデータ内に存在するか確認した．次に，存在する予約語に対して，その単語の ID(辞書内で一意に定義された整数値) を取得し，トピックごとに対応する初期単語分布を更新した．この行列は，トピック数を行，辞書内の単語数を列とする二次元配列であり，各要素はそのトピックと単語間の初期関連度を示す.

具体的には，あるトピック k に対して予約語 w_i が辞書内で識別される場合，その単語に対応する列 j とトピックに対応する行 i の交差する位置に 1 を設定した．存在しない単語に

表 4.2: 設定した予約語

トピック	予約語									
恋愛	愛	恋	好き	ハート	思い	抱く	運命	感情	心	涙
青春	青春	夏	友情	希望	未来	夢	仲間	笑顔	時代	冒険
応援	頑張る	挑戦	力	勝利	支える	声援	負けない	全身	成功	エール
友情	友達	絆	信頼	仲間	助け合い	約束	支え	関係	共感	笑い
孤独・絶望	孤独	絶望	悲しみ	苦しみ	切ない	涙	痛み	空虚	寂しい	暗闇
夢・目標	夢	目標	挑戦	努力	願い	踏み出す	達成	叶う	ビジョン	挑む
人生	人生	生きる	道	選択	自由	終わり	意思	旅	出発点	経験
ストーリー	物語	ストーリー	冒険	過去	未来	伝説	出会い	探す	運命	証明
季節・風景	春	夏	秋	冬	季節	雪	櫻	雨	空	風
地名・場所	東京	大阪	故郷	海	山	街	空港	駅	夜	風景

ついては，対応する値を 0 のままとした．値が 1 の場合はその単語が対応するトピックのシードワードとして関連付けられていることを示す．値が 0 の場合はその単語がそのトピックに特に関連付けられていないことを示す．予約語とトピックの関連付けによって構築された初期トピック分布行列は，トピックモデルの学習プロセスにおいてガイドとして機能する．

トピックによる楽曲のクラスタリング

ガイド付き LDA のモデルの学習には gensim というモジュールを使用する．gensim は，トピックモデリングや文書類似度解析に特化した Python の自然言語処理ライブラリである．大規模なテキストコーパスを効率的に処理できる最適化されたアルゴリズムを提供しており，LDA に加えて，LSA や Doc2Vec などの様々なアルゴリズムをサポートしており，FastText や word2vec の学習済みモデルを読み込んで活用することも可能である．gensim の大きな特徴は，使いやすく直感的な API を備えていることであり，複雑な自然言語処理のタスクを手軽に実装できる点にある．また，大規模なデータセットを効率的に処理するためのメモリ効率や並列処理のサポートに優れ，高速な動作を実現している．さらに，scikit-learn など他のライブラリとの連携も容易であるため，様々なプロジェクトで活用しやすいという利点がある [54].

モデルの学習において，文章のトピック分布の初期パラメータとトピックごとの単語分布の初期値を決める必要がある．文章のトピック分布のパラメータは自動で調整するようにした．トピックごとの単語分布の初期値は先ほど作成した初期単語分布を活用する．これにより，トピックごとに任意の単語が配置されるようになる．LDA のモデル実行では，「models.LdaModel」関数を用いて LDA モデルを構築し，コーパスデータを用いて学習を行う．学習プロセスでは，指定したトピック数とパラメータを基に，トピック分布と単語分布が更新される．最終的に，確率が最大のトピック ID を選択し，そのトピックを文書の主要トピックとして分類する．

クラスタリング結果および楽曲特徴量におけるプレイリストの作成

2.1 節でも述べた通り，楽曲の特徴量の中でもテンポ，曲調，アコースティック性，エネルギー，歌詞が感情に影響を与え，その影響はその人の現在の感情状態に大きく依存する．例

辞書のイメージ

単語ID	単語
0	私
1	行く
2	それ
3	学校
4	徒歩
5	で
⋮	⋮

コーパスのイメージ

文章ID	出現回数	私	行く	それ	学校	徒歩	で	…
1	10	3	2	0	1	0	1	…
2	23	1	0	0	0	2	3	…
3	42	0	1	1	3	0	1	…

図 4.6: 辞書とコーパスのイメージ

図 4.7: 保存されたプレイリストの例

えば，ネガティブ強度が高い人はテンポが遅く，曲調がマイナー調で，アコースティック性が高く，エネルギーが低く，ネガティブな歌詞が含まれる楽曲を聴くことが効果的であり，ネガティブ強度が低い人はテンポが速く，曲調がメジャー調で，アコースティック性が低く，エネルギーが高く，ポジティブな歌詞が含まれる楽曲を聴くことが効果的であるとされる．そのため，各トピックに対して，テンポ，曲調，アコースティック性，エネルギー，歌詞の感情値の閾値を用いて，ネガティブ度合いに応じたプレイリストを分類する．

具体的には，各トピックに対して，ネガティブ強度が高い人に対して効果的な楽曲の条件としてテンポが 80 未満，エネルギーが 0.33 未満，アコースティック性が 0.66 以上，コードがマイナー長，歌詞の感情値が 0.33 未満の楽曲とした．ネガティブ強度が中程度の人に対して効果的な楽曲の条件の条件としてテンポが 80 以上 120 未満，エネルギーが 0.33 未満，アコースティック性が 0.33 以上 0.66 未満，歌詞の感情値が 0.33 以上 0.66 未満の楽曲とした．ネガティブ強度が低い人に対して効果的な楽曲の条件としてテンポが 120 以上，エネルギーが 0.66 以上，アコースティック性が 0.33 未満，コードがメジャー長，歌詞の感情値が 0.66 以上の楽曲とした．

以上の条件を満たす楽曲を各トピックごとに集め，トピックごとに高強度，中強度，低強度のプレイリストを作成した．保存されたプレイリストの例を図 4.7 に示す．

§ 4.3　提案システムの概要

4.1 節では，SpotifyAPI および GeniusAPI，BERT を用いた楽曲特徴量の取得方法について述べた．また，4.2 節では，取得した特徴量をもとにガイド付き LDA を用い，楽曲を設定した予約語に従ってクラスタリングを行い，プレイリストを作成する方法について述べた．

本節では，本研究で用いたデータや提案手法について再度整理し，システム全体の構成を述べる．システム全体の流れを図 4.8 に示す．また，以下では本研究におけるシステムを本システムと呼ぶ．そして，具体的な流れを以下に述べる．

Step1：楽曲のプレイリストの作成

ユーザがシステムを使用する前に，楽曲のプレイリストの HTML を作成する．HTML の作成の流れは以下の通りとなっている．まず，SpotifyAPI を用い楽曲の特徴量を取得する．

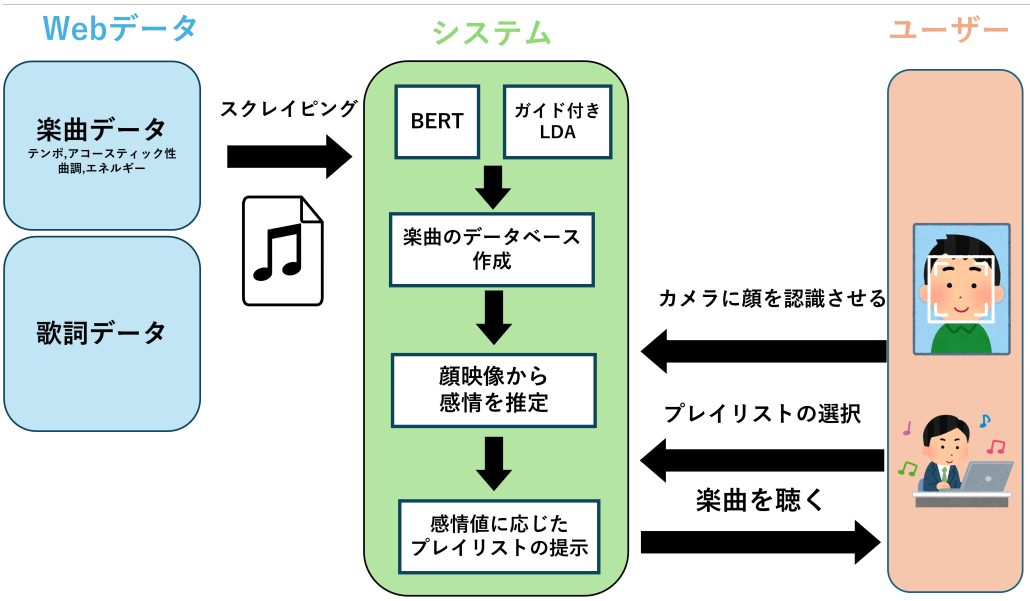

図 4.8: 提案システムの流れ

SpotifyAPI とは，音楽ストリーミングサービス Spotify の豊富なデータをプログラムでアクセスできる API であり，SpotifyAPI を用いることにより，楽曲やアーティスト，プレイリスト，アルバムをはじめとした楽曲のメタデータだけでなく，楽曲の音楽的な特徴量を取得することができる．本研究では楽曲特徴量として「テンポ」,「曲調」,「アコースティック性」,「エネルギー」を取得する．また，このときに楽曲のアーティスト名と楽曲名を取得する．

次に，GeniusAPI を用い，SpotifyAPI で取得したアーティスト名と楽曲名をもとに，楽曲の歌詞を取得する．Genius は楽曲の歌詞や関連情報を提供するサービスであり，この API を用いることで指定されたアーティスト名と楽曲名に基づき，対応する歌詞を取得することができる．楽曲の歌詞を取得する際，[Intro] や，[Verse 1] などの楽曲に関係ない部分が取得されるので，その部分は除去する．そして，BERT を用いて各歌詞の感情値を分析し，出力結果を歌詞の感情値として特徴量を追加する．

また，ガイド付き LDA を用い，歌詞のテーマに基づくプレイリストを作成する．想定したテーマは，歌詞の作曲でよく使われる「恋愛」「青春」「応援」「友情」「絶望・孤独」「夢・目標」「人生」「ストーリー」「季節・風景描写」「地名場所」とし，それに合った予約語を設定し，トピックを分類した．そして分類したトピックに対して,「テンポ」,「曲調」,「アコースティック性」,「エネルギー」による閾値を設定し，それぞれのネガティブ強度人向けのプレイリストを設定し，SpotifyAPI を活用しプレイリストを作成した．また，作成したプレイリストを Spotify の埋め込み機能を利用して HTML を作成する．

Step2：顔認証における感情推定

本システムでは，ユーザの感情状態をリアルタイムに推定するために顔認証技術を用いる．顔認証技術により，ネガティブ強度を定量化し，音楽推薦に反映させている．感情推定には，Python の顔認証ライブラリである DeepFace を使用する．DeepFace は Python ベースのオープンソースライブラリで，顔認証や顔分析（年齢，性別，感情推定など）に特化

	Word	Score
0	優れる	1.000000
1	良い	0.999995
2	喜ぶ	0.999979
3	褒める	0.999979
4	めでたい	0.999645
...	...	...
55120	ない	-0.999997
55121	酷い	-0.999997
55122	病気	-0.999998
55123	死ぬ	-0.999999
55124	悪い	-1.000000

図 4.9: 単語感情極性辞書

表 4.3: 感情と重み

単語	重み
喜び	0.998871
驚き	0.266928
嫌悪	-0.5892
恐怖	-0.71366
怒り	-0.99667
悲しみ	-0.99924

したものである．主要な深層学習フレームワークである TensorFlow や Keras を活用し，事前学習済みモデルを利用することで高精度な分析ができる．

顔認証におけるネガティブ強度の推定方法について説明する．DeepFace では，カメラ映像から顔を検知し，3D アライメントを用いて顔を正面化する．正面化された顔画像は 9 層の畳み込みニューラルネットワークの入力画像として用いられ，特徴量を抽出が抽出される．そして，その特徴量から 7 つの感情 (喜び:Joy，驚き:Surprise，恐怖:Fear，悲しみ:Sad，怒り:Angry，嫌悪:Disgust，中立:Neutral) の確率が出力される．本研究では，この確率値を感情スコアとし，それぞれの表情の感情値とする．そして，それぞれの感情値に対し，単語感情極性対応表を基に重みづけを行うことで，表情からネガティブ強度を推定する．

単語感情極性対応表

単語感情極性対応表は，東京工業大学の高村教授が公開した，スピンモデル (SPIN Model) を用いて作成された感情辞書である．スピンモデルとは物理学のスピン理論を応用し，単語の感情極性を文脈に基づいて推定する手法である．単語が同じ文脈で共起する場合，それらの単語は同じ感情極性を持つと仮定し，語彙ネットワークを構築する．このネットワークを用いて，単語同士の感情的な関係性を計算し，感情極性を決定している．

単語感情極性対応表は，各単語が持つ感情値を-1〜1 の範囲で割り振ったものであり，0 未満の単語はネガティブ，0 以上の単語はポジティブに分類される．ネガティブ強度は，各感情のうちネガティブに分類される嫌悪，恐怖，怒り，悲しみそれぞれの感情スコアとその重みの加重平均として計算する．

Step3：ネガティブ度に応じたプレイリストの提示および選択

ユーザのネガティブ強度はカメラが起動された瞬間から楽曲を表示するまでの平均値によって計算される．ネガティブ強度は，ネガティブスコアの平均値が 0.3 未満であれば「低ネガティブ度」，0.3 以上 0.6 未満であれば「中ネガティブ度」，0.6 以上であれば「高ネガティブ度」として分類し，それぞれのネガティブ度に合ったプレイリストの画面に推移する．

システムの処理

図 4.10: HTML のフロントページ

図 4.11: 表示されるプレイリストの例

本研究では，Flask を利用してカメラ映像からユーザのネガティブ強度を分析し，そのネガティブ強度に応じた特徴量を持つ楽曲を提示するアプリケーションを開発した．Flask は Python で Web アプリケーションを開発するための軽量なフレームワークであり，その主な特徴は，シンプルでコンパクトな設計であり，最小限の設定とコードでアプリケーションを素早く構築できる点である．また，Flask は非常に柔軟で拡張性が高く，必要に応じて外部ライブラリや機能を追加して，特定の要求に応じたアプリケーションを作成することができる．

システムの流れはフロントページとプレイリストのページの 2 つの部分で構成されている．HTML のフロント部分と表示されるプレリストのページを図 [?] および図 [?] に示す．HTML のフロント部分ではカメラから取得される映像と表情から 7 つの感情の確率を計算し，その時のネガティブスコアが表示される．また，ネガティブスコアの下には，カメラの起動から現在までのネガティブスコアの平均値が表示されている．

「楽曲を表示」というボタンを押すと，ネガティブスコアの閾値をもとに「高強度」,「中強度」,「低強度」の「恋愛」のプレイリストが埋め込まれた HTML ファイルに移動する．例えば，プログラム実行から「楽曲を表示」のボタンを押すまでに測定した時の平均ネガティブスコアが 0.6 以上であれば「恋愛_高ネガティブ度」の HTML ファイルへ移動し，0.3 以上 0.6 未満であれば「恋愛_中ネガティブ度」の HTML ファイルへ移動し，0.3 未満であれば「恋愛_低ネガティブ度」の HTML ファイルへ移動する．

また，移動する HTML にはプレイリストと他のトピックへ移動するボタンが配置されている．プレイリストの再生ボタンを押すと，プレイリストの先頭から順に曲が再生される．また，ほかのトピックへ移動するボタンを押すと同じ強度向けの，トピックに応じた，プレイリストが表示される．このボタンはユーザが現在聞いている曲が好みでなかった場合に任意で選択するためものである．

第5章

数値実験並びに考察

§ 5.1　数値実験の概要

本研究の数値実験として本システムによりユーザのネガティブ強度が減少したかどうかに注目して評価実験を行う．具体的には，本システムを実際に使用してもらい，システム使用中のネガティブ強度を測定することにより，ネガティブ強度が減少したかどうかを定量的に評価する．

本システムで設定した数値

まず，本システムの実装にあたり設定した数値を説明する．本システムは，Spotify，Genius からの楽曲特徴量および歌詞の取得，BERT による歌詞の感情分析，ガイド付き LDA による及び楽曲特徴量によるプレイリストの作成となっている．なお，本システムの実装に用いたプログラミング言語とパッケージは表 5.1 のとおりである．

楽曲データは，Spotify で配信されている楽曲のうち 1990 年〜2022 年までの人気の曲を中心とした 1500 曲とした．BERT のモデルは，事前学習済みの BERT モデルである「cl-tohoku/bert-base-japanese-v2」を使用した．このモデルに，感情分析用のデータセット「chABSA-dataset」を使用しファインチューニングさせることによって感情分析に適応させる．この時，トレーニングのパラメータとして学習率を $5e-5$，バッチサイズを 16，エポック数を 3，重み減衰を 0.01 とした．学習率はモデルが重みを更新する際のステップの大きさを制御する値であり，バッチサイズは一度にモデルへ入力されるデータサンプルの数であり，エポック数は学習データを全て 1 回モデルに入力するプロセスの回数を指し，重み減衰率は，モデルの過学習を防ぐためのパラメータである．

さらに，ガイド付き LDA においては，モデルのパラメータはトピック数は 10，パス数は 10，文章のトピック分布のパラメータは自動で調整するように設定した．ここで，パス数は LDA モデルがコーパスを何回反復するかを決めるパラメータである．また，各トピックごとの初期単語分布は，4.2 節の表 4.2 に示した予約語を用いて作成し，学習させた．

実験手順

次に，本研究の実験環境について説明する．調査の対象は同研究室の学生の合計 8 人に実際に開発したシステムを使用してもらった．実験を行う際，実験時の影響を排除するために，次の条件を設定した．まず，システムを利用する際，過去 1 時間以内に音楽を聴いていない状態とし，実験時の影響を少なくした．また，実験前に，リラックスした環境で待

表 5.1: 実装環境

プログラミング言語	Python 3.12.7
使用パッケージ	Deepface，Flask，TensorFlow，tf-keras，transformers，gensim，spotipy

機してもらい，感情の偏りを防いだ．さらに，周囲の感情の影響を軽減させるため，実験を行う際，周囲に何もない場所でシステムを使ってもらった．

次に，実験の手順を説明する．実験の手順は以下のとおりである．

1. まず，ユーザは顔をカメラで 15 分間撮影し，DeepFace を用いて 1 秒ごとのネガティブ度を推定する．
2. 15 分間のデータから平均ネガティブ度を算出する．
3. 15 分間の平均ネガティブ度に基づき，事前に設定した強度を持つプレイリストを推薦する．
4. ユーザは推薦された楽曲を 15 分間視聴する．この際，別のプレイリストへ任意に選択できるようにする．

以上の手順で実験を行い，楽曲推薦前後の 15 分平均ネガティブ強度をもとに，ネガティブ強度が減少したかを評価する．

評価方法

次に，楽曲を推薦したことによりネガティブ強度が減少したかどうか検証を行う．この検証では推薦前後のネガティブ度の平均値に有意な差があるかを検証する指標として t 検定を用いる．t 検定は，2 つのサンプルの平均値を比較する際に，統計的に意味のある差があるかを判断する際に用いられる分析手法である．t 検定には「対応のある t 検定」,「対応のない t 検定」があり，対応のない t 検定には「等分散を仮定した 2 標本による検定」と「分散が等しくないと仮定した 2 標本による検定」に区別される．

対応のある t 検定は，同じ被験者に対して 2 つの異なる時点でデータを取得し，その変化を評価する検定のことである．これは，データがペアになっているため，個々のデータの差分を求め，その平均値と標準誤差をもとに t 検定を行う．対応のない t 検定は，異なる 2 つのグループの平均値を比較する際に用いられる．

本研究においては同じユーザに対して，楽曲推薦前のネガティブ強度と楽曲推薦後のネガティブ強度の変化を検証するため，対応のある t 検定を用いる．具体的には，楽曲推薦前 15 分のネガティブ強度の平均値と楽曲推薦後 15 分のネガティブ強度の平均値を用いることで有意性を検証する．また，各ユーザの 1 分ごとのネガティブ強度の平均値の推移をもとに，システムに対する考察を行う．

§ 5.2　実験結果と考察

本節では，5.1 節で述べた数値実験について，結果とそれに対する考察を示す．表 5.2 はそれぞれのユーザの楽曲推薦前後の 15 分間のネガティブ強度の平均値であり，表 5.3 は平

表 5.2: 楽曲推薦前後のネガティブ強度の結果

	推薦前	推薦後	変化率
ユーザ1	0.72803	0.53875	-0.25999
ユーザ2	0.825086	0.756219	-0.08347
ユーザ3	0.811251	0.722563	-0.10932
ユーザ4	0.687888	0.47066	-0.31579
ユーザ5	0.466594	0.299254	-0.35864
ユーザ6	0.737784	0.691062	-0.06333
ユーザ7	0.670398	0.384993	-0.42572
ユーザ8	0.38685	0.319925	-0.173
平均値	0.664235	0.522928	-0.22366

表 5.3: t 検定の結果

	推薦前	推薦後
平均	0.664235	0.522928
分散	0.024795	0.033685
観測数	8	8
ピアソン相関	0.883232	
仮説平均との差異	0	
自由度	7	
t	4.637091	
P(T<=t) 片側	0.001189	
t 境界値 片側	1.894579	
P(T<=t) 両側	0.002378	
t 境界値 両側	2.364624	

均値をもとに t 検定を行った結果である．また，図 5.1，図 5.2，図 5.3，図 5.4，図 5.5，図 5.6，図 5.7，図 5.8 はそれぞれのユーザの 1 分ごとのネガティブ強度の平均値の推移である．また，各グラフの 15 分目にひかれている縦線は楽曲が提示されたタイミングを表し，この線以前は楽曲推薦前のネガティブ強度の推移を表し，以降は楽曲推薦後のネガティブ強度の推移を表す．

まず，t 検定の結果について考察する．本研究では，楽曲推薦前後のネガティブ強度の平均値に差があるかどうかを検証するために，以下の帰無仮説，対立仮説を立てる．帰無仮説とは検定を行うために立てる仮説のことであり，対立仮説とは帰無仮説に対する仮説のことである．

- 帰無仮説：楽曲推薦前後でネガティブ強度の平均値に有意に差がない
- 対立仮説：楽曲推薦前後でネガティブ強度の平均値が有意に減少する

表 5.3 において平均は推薦前後の全ユーザのネガティブ強度の平均値の表し，分散は全ユーザのネガティブ強度の分散を表す．また，ピアソン相関は 2 つのデータの間にどの程度の線形相関があるかを示す指標であり，自由度はある変数において自由な値をとることのできるデータの数である．

t 値は比較するデータに意味がある差があるかどうかを示す値であり，P(T<=t) 片側，P(T<=t) 両側はそれぞれ片側検定と両側検定における p 値を表す．p 値は帰無仮説が正しいと仮定したときに，計算された統計量以上の極端な結果が得られる確率のことである．t 境界線片側，t 境界線両側はそれぞれ片側検定と両側検定における t 分布において，有意水準の境界となる値であり，t 境界値よりも t 値が大きければ，帰無仮説が棄却できる．

本研究では，楽曲推薦前後でネガティブ強度の平均値が減少したかどうかを検証するため，片側検定を用いる．検定の結果，t 境界値片側は 1.894579 であり，t 値は 4.637091 である．t 値が t 境界線片側よりも大きいため，帰無仮説が棄却される．そして，片側検定における p 値は 0.001189 であり，これは有意水準 0.05 を下回るため，推薦前後でネガティブ強度の平均値が有意に減少したことが示される．これにより，楽曲を推薦したことにより，ネガティブ強度の平均値が有意に減少したことがわかった．

次に，表音楽視聴前後のネガティブ強度の結果および図 5.3，図 5.4，図 5.5，図 5.6，図 5.7，図 5.8 を用いて，各ユーザのネガティブ強度の変化について考察する．各ユーザのネガティブ強度の変化および推移に関してわかることが 3 つある．各ユーザのネガティブ強度の変化および推移に関してわかることの 1 つ目は，ネガティブ強度の減少の割合は個人差があるということである．表 5.2 のユーザそれぞれの変化率より，ユーザそれぞれの変化率が異なっていることがわかる．ユーザそれぞれの変化率が異なっている理由として以下のことが考えられる．まず，楽曲の嗜好性があげられる．本研究ではユーザのネガティブ強度を測定し，そのネガティブ強度に応じた特徴量を持つ楽曲を提示している．その際，異なるトピックのプレイリストをユーザが任意に変更できるように設定している．楽曲推薦時，ユーザによっては楽曲の好みと一致しネガティブ強度が大きく低下した可能性があると考える．逆に，推薦された楽曲が好みと一致しないものが多かった場合，効果が薄く，変化率が小さくなった可能性があると考えられる．次に表情の読み取りの精度の問題があげられる．本研究ではユーザの感情を推定する手法として表情からネガティブ強度を推定している．ユーザごとに，測定時の顔の向きや表情のとらえ方のばらつきがあると，同じ刺激でも推定する数値に差が生じてしまい，結果として変化率の違いが生じてしまった可能性がある．

各ユーザのネガティブ強度の変化および推移に関してわかることの 2 つ目は全体的にネガティブ強度が高めに測定されていることである．表 5.2 の推薦前の各ユーザのネガティブ強度の平均値の結果から，ユーザ 8 人のうちユーザ 1，ユーザ 2，ユーザ 3，ユーザ 4，ユーザ 6，ユーザ 7 の 6 人が高ネガティブ度に分類され，ユーザ 5，ユーザ 8 の 2 人が中ネガティブ度に分類されることがわかる．また，推薦前の全ユーザのネガティブ強度の平均値は 0.664235 であり，全体としてネガティブ強度が高い水準にあることがわかる．ユーザのネガティブ強度の平均値が高くなった原因として以下の可能性が考えられる．まず，環境や実験条件の影響があげられる．これは楽曲推薦前の静かな環境による心理的緊張状態などがネガティブ感情を一時的に増大させてしまった可能性がある．次に，ネガティブ強度の推定方法があげられる．本研究では，表情から 7 つの感情スコアを推定し，単語感情極性辞書を用いてネガティブ強度を変換した．そのため，単語感情極性辞書をもとにしたネガティブ強度の推定結果が，表情の感情と実際の心理状態を反映していない可能性がある．そのため，心拍数や脳波データなどを用いて，この推定方法が妥当であったか検証する必要がある．

各ユーザのネガティブ強度の変化および推移に関してわかることの 3 つ目は，ネガティブ強度の変化が一定でないことである．すべてのユーザにおいて，折れ線グラフのパターンが単調な下降曲線ではなく，一定期間ごとに上昇と下降を繰り返している．ネガティブ強度の測定値にずれが生じる理由として，測定誤差やノイズが考えられる．表情によるネガティブ強度の測定は，微細な顔の動きや瞬きにより感情の推定結果が変化する．これにより，短期的なネガティブ強度の上昇および下降が生じてしまった可能性がある．今後の検討として推定時の極端な外れ値の除去やデータの平滑化を行う必要があると考える．

以上の結果から，本システムによる楽曲推薦は，統計的に有意なネガティブ感情の低減効果を示す可能性があることが確認された．しかし，表情からネガティブ強度を推定するには限度があり，ユーザによって精度の差があることが分かった．今後は，他の感情指標との併用などにより，システムの効果をさらに明確にする必要があると考える．

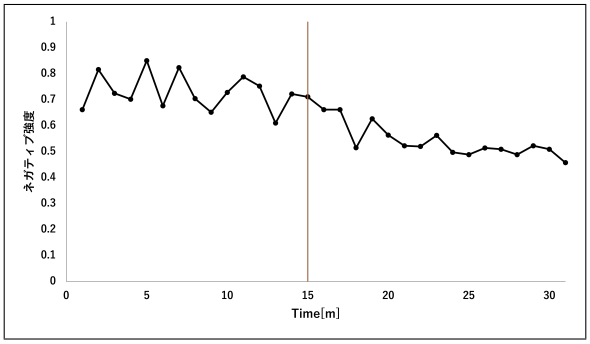

図 5.1: ユーザ 1 のネガティブ強度の平均値の推移

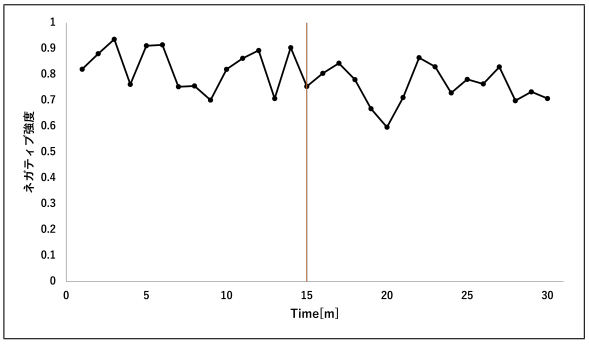

図 5.2: ユーザ 2 のネガティブ強度の平均値の推移

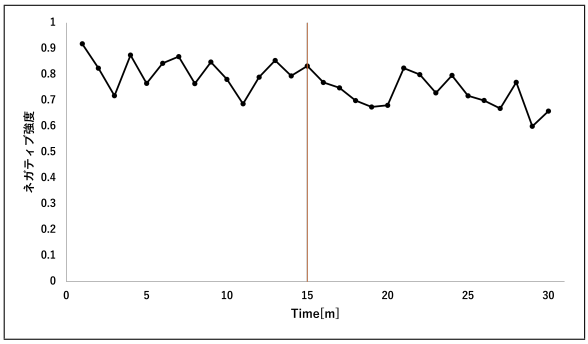

図 5.3: ユーザ 3 のネガティブ強度の平均値の推移

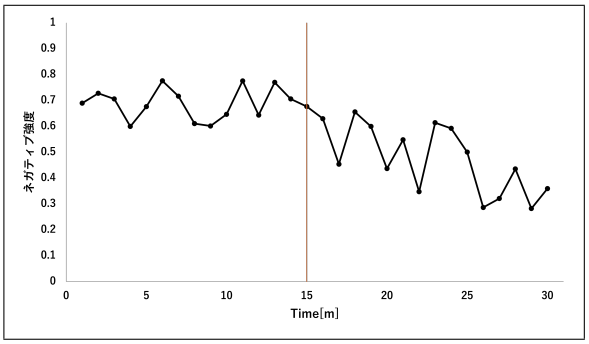

図 5.4: ユーザ 4 のネガティブ強度の平均値の推移

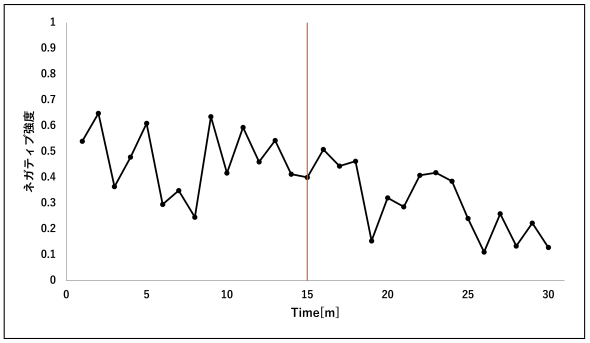

図 5.5: ユーザ 5 のネガティブ強度の平均値の推移

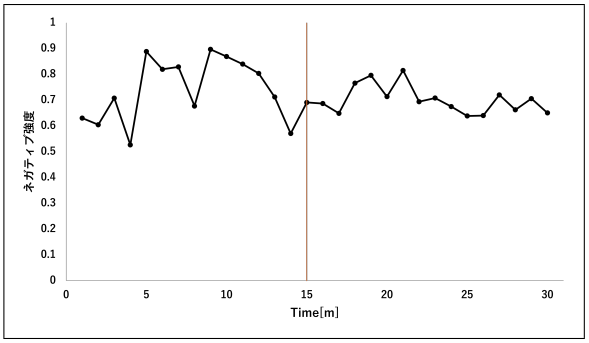

図 5.6: ユーザ 6 のネガティブ強度の平均値の推移

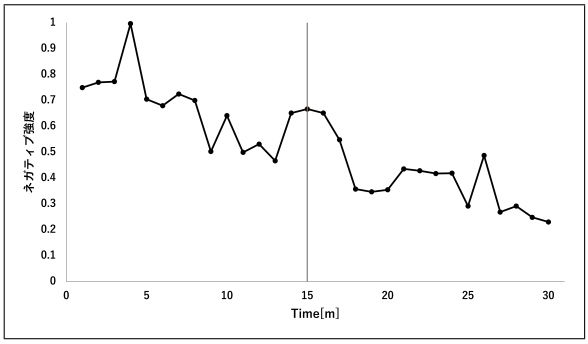

図 5.7: ユーザ 7 のネガティブ強度の平均値の推移

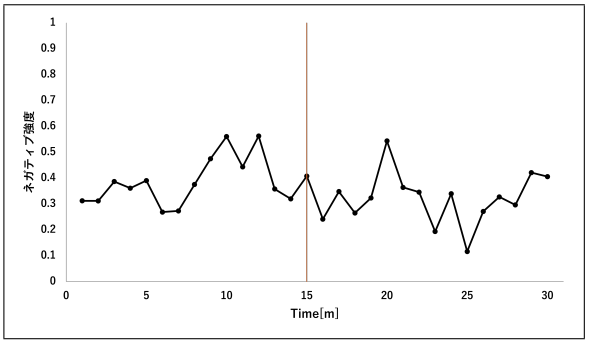

図 5.8: ユーザ 8 のネガティブ強度の平均値の推移

第6章

おわりに

本研究ではトピックモデルと感情推定を活用し，ユーザのネガティブ感情を軽減させるための音楽推薦システムを開発した．具体的には，SpotifyAPI を用いて感情に影響を与える特徴量である，曲調，アコースティック性，エネルギー，テンポを取得した．また，GeniusAPI を用いて楽曲の歌詞を取得した．そして，歌詞に対して BERT を用いた感情分析を行い，分析結果を歌楽曲の特徴量として追加した．次に，楽曲を任意のトピックへクラスタリングを行うため LDA を拡張したガイド付き LDA を用いた．分類するトピックは楽曲でよく用いられる 10 つのトピックとし，それに対応した予約語を主観で選定することで実現した．数値実験では実際にシステムを複数人に使用してもらい，推薦前後のネガティブ強度の平均値を測定することでシステムの有効性を検証した．その結果，楽曲推薦前後においてネガティブ強度が有意に減少していたことが示された．

今後の課題として以下のことがあげられる．1 点目は，楽曲の英語の歌詞への対応である．本研究では歌詞の感情値を分析するために BERT を用いた．しかし本研究で用いた BERT のモデルは日本語の感情付きラベルをファインチューニングさせて感情分析に適応させている．そのため，英語の歌詞を省略し日本語の歌詞の部分のみを分析対象としている．しかし，これでは歌詞の正確な感情値が反映されていない．ガイド付き LDA においても同様に英語の歌詞を省略し，日本語の歌詞のみを分析し，歌詞のトピックを分析している．そのため，英語の歌詞を考慮し分析する方法を模索する必要があると考える．

2 点目はガイド付き LDA の予約語の選出方法である．本研究では歌詞を任意のトピックへ分類するために各トピックに対して，トピックに関連する予約語を主観で決定している．そのため，予約語によっては歌詞のトピックを適切に分類できない可能性がある．そのため，予約語がトピック推定結果に与える影響についての詳細な検証が必要である．

3 点目は表情からネガティブ強度を推定する方法である．本研究では表情からネガティブ強度を分析する手法として，顔認証技術を用い，7 つの感情スコアを推定した．そして，各感情のスコアに対して，単語感情極性辞書を用いてネガティブ強度を推定した．しかし，表情からネガティブ強度を推定する手法はされていなかったため，本研究では表情の重みが単語感情極性辞書に従うという仮定に基づき感情からネガティブ強度を推定した．そのため，単語感情極性辞書を基にしたネガティブ強度の推定が，表情の感情と実際の心理状態を適切に反映しているかの検証が不十分であるという課題がある．今後は，表情からネガティブ強度を推定すると同時に，心拍数や脳波データなどを測定することで，この仮定が妥当であるか検証する必要があると考える．また，本研究では 8 名を対象にシステムを使用してもらい有効性を示したが，今後はより多くのユーザに対して長期間の実験を行い，システムの有効性を検証する必要がある．

謝辞

本研究を遂行するにあたり，多大なご指導と終始懇切丁寧なご鞭撻を賜った富山県立大学情報工学部データサイエンス学科システム数理講座の António Oliveira Nzinga René 講師，奥原浩之教授に深甚な謝意を表します．最後になりましたが，多大な協力をしていただいた研究室の同輩諸氏に感謝致します．

2025 年 2 月

水上 和秀

参考文献

[1] ヒューマンアカデミー，“ネガティブな感情が及ぼす影響とは？コントロール法を知って上手に対処しよう”，2025 年 2 月 2 日閲覧，
https://haa.athuman.com/media/psychology/mental-care/1347/

[2] ヘッドミント草加，“癒しの音楽とその効果について深掘り！”，2025 年 2 月 2 日閲覧，
https://headmint-soka.jp/column/37f7d2a0-2ed7-447a-88dc-89574f1a3863

[3] 公益財団法人長寿科学振興財団，“音楽療法”，2025 年 2 月 2 日閲覧，
https://www.tyojyu.or.jp/net/byouki/ninchishou/music.html

[4] 一般社団法人フラット音楽企画，“音楽療法の効果”，2025 年 2 月 2 日閲覧，
https://flatmusic.or.jp/dispatch/kouka.php

[5] 日本精神神経学会 田町三田こころみクリニック，“うつや不安を和らげる音楽療法の効果とは？”，2025 年 2 月 2 日閲覧，
https://cocoromi-mental.jp/cocoromi-ms/other/selfcare/music/

[6] 日本音楽療法学会，“音楽療法士とは”，2025 年 2 月 2 日閲覧，
https://www.jmta.jp/music_therapist/

[7] Gold，B. P.Frank，M. J.Bogert，B.& Brattico，E. (2013). “Pleasurable music affects reinforcement learning according to the listener”，Frontiers in Psychology，4，541.

[8] Tervaniemi，M.Makkonen，T.& Nie，P. (2021). “Psychological and physiological signatures of music listening in different listening environments—An exploratory study”，Brain Sciences，11(5)，593.

[9] CLRN，“Does music make you happy?”，2025 年 2 月 2 日閲覧，
https://www.clrn.org/does-music-make-you-happy/

[10] Bernardi，L.Porta，C.& Sleight，P. (2006). Cardiovascular，cerebrovascular，and respiratory changes induced by different types of music in musicians and non-musicians”The importance of silence. Heart，92(4)，445–452.

[11] Parncutt，R. (2013)，“Major-minor tonality，Schenkerian prolongation，and emotion: A commentary on Huron and Davis”Empirical Musicology Review，7(3-4)，118–137.

[12] Carraturo，G.Pando-Naude，V.Costa，M.Vuust，P.Bonetti，L.& Brattico，E. (2024).“The major-minor mode dichotomy in music perception”Physics of Life Reviews，52，80–106

[13] Daws，L. (2019). “Emotional responses to acoustic music: A study of instrumental and vocal stimuli”Oxford Academic.

[14] Harmony & Healing, “How music can reduce stress and improve mental health”, 2025年2月2日閲覧,
https://www.harmonyandhealing.org/how-music-can-reduce-stress/

[15] Juslin, Patrik N.“Handbook of Music and Emotion: Theory, Research, Applications ”https://doi.org/10.1093/acprof:oso/9780199230143.001.0001,

[16] Matsumoto, J. (2002).“音楽の気分誘導効果に関する実証的研究——人はなぜ悲しい音楽を聴くのか”, 教育心理学研究, 50(1), 23–32.

[17] Altschuler, I. M. (1954), “The past, present, and future of musical therapy”, In Music Therapy (pp. 24–35). Philosophical Library.

[18] 神嶌 敏弘, “推薦システムのアルゴリズム”, 人工知能学会誌, Vol.22 No.6, pp.826-837

[19] Baeldung, “How Does the Amazon Recommendation System Work?”, 2025年2月2日閲覧,
https://www.baeldung.com/cs/amazon-recommendation-system

[20] Coursera, “Coursera: Learn new skills with online courses from top universities”, 2025年2月2日閲覧,
https://www.coursera.org

[21] Quantum Zeitgeist.“YouTube Music boosts user satisfaction with AI-powered transformers”, 2025年2月2日閲覧,
https://quantumzeitgeist.com/youtube-music-boosts-user-satisfaction-with-ai-powered-transformers/

[22] Artist Push, “How does Spotify algorithm work for artists?”, 2025年2月2日閲覧,
https://artistpush.me/blogs/news/how-does-spotify-algorithm-work-for-artists

[23] Zhou, R.Khemmarat, S.& Gao, L. (2010). “The impact of YouTube recommendation system on video views”, In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement* (IMC ’10) (pp. 404–410) Association for Computing Machinery

[24] He, X.Liu, Q.& Jung, S. (2024), “The Impact of Recommendation System on User Satisfaction: A Moderated Mediation Approach”, Journal of Theoretical and Applied Electronic Commerce Research, 19(1), 448-466.

[25] Thematic, “Sentiment analysis”, 2025年2月2日閲覧,
https://getthematic.com/sentiment-analysis

[26] Cotra, “感情分析とは?初心者にもわかりやすく解説!顧客対応の活用例も紹介”, 2025年2月2日閲覧,
https://www.transcosmos-cotra.jp/sentiment-analysis

[27] 高村大也，乾孝司，奥村学，“スピンモデルによる単語の感情極性抽出”，情報処理学会論文誌ジャーナル，Vol.47 No.02，pp.627-637，2016

[28] 東北大学 乾・鈴木研究室 “日本語評価極性辞書”，2025 年 2 月 2 日閲覧，
http: //www. cl. ecei. tohoku. ac. jp/index. php? Open%20 Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary

[29] G. H. Mohmad Dar and R. Delhibabu, “Speech Databases, Speech Features, and Classifiers in Speech Emotion Recognition: A Review”，in IEEE Access，vol. 12，pp. 151122-151152，2024,

[30] Ekman，P. (n.d.)，“Facial Action Coding System (FACS)”，2025 年 2 月 2 日閲覧，
https://www.paulekman.com

[31] 満倉靖恵 (2020)“脳波によるリアルタイム感性計測とその応用 ——実社会における感情・感性を用いる試みの広がり——”，IEICE Fundamentals Review，13(3)，180-186.

[32] CAC Holdings Corporation，“CAC の表情・感情認識 AI をベネッセが英語オンラインレッスンの講師の指導品質向上に活用〜 オンラインレッスン中の講師の表情やジェスチャーをリアルタイム解析 〜 ”，2025 年 2 月 2 日閲覧，
https://www.cac.co.jp/news/topics_220317/

[33] Amazon Web Services，“Detect sentiment from customer reviews using Amazon Comprehend. AWS Machine Learning Blog”，2025 年 2 月 2 日閲覧，
https://aws.amazon.com/jp/blogs/machine-learning/detect-sentiment-from-customer-reviews-using-amazon-comprehend/

[34] 日立製作所，“社内外に散在するお客さまの生の声や感情を可視化して製品開発やグローバルマーケティングに生かす”，2025 年 2 月 2 日閲覧，
https://social-innovation.hitachi/ja-jp/case_studies/honda_motor/

[35] Devlin，J.Chang，M.-W.Lee，K.& Toutanova，K. (2019)，“BERT: Pre-training of deep bidirectional transformers for language understanding”，
https://arxiv.org/pdf/1810.04805

[36] AI 総合研究所，“Transformer とは？モデルの概要や BERT との違いをわかりやすく解説”，2025 年 2 月 2 日閲覧，
https://www.ai-souken.com/article/transformer-overview#position-wise-feed-forward-network%E5%B1%A4

[37] 長澤 尚武，萩原 将文，“文脈を考慮した単語の感情推定”，日本感性工学会論文誌，Vol. 23 (2023)，No. 2 pp. 87-96

[38] Y. Taigman，M. Yang，M. Ranzato and L. Wolf，“DeepFace: Closing the Gap to Human-Level Performance in Face Verification”，in 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)，Columbus，OH，USA，2014，pp. 1701-1708,

[39] Unite AI, “DeepFace for Advanced Facial Recognition”, 2025 年 2 月 2 日閲覧, https://www.unite.ai/deepface-for-advanced-facial-recognition/,

[40] イメージングソリューション, “アフィン変換（平行移動, 拡大縮小, 回転, スキュー行列）”, 2025 年 2 月 2 日閲覧, https://imagingsolution.net/imaging/affine-transformation/,

[41] Sefik Ilkin Serengil, “Face Recognition with Facebook DeepFace in Keras”, 2025 年 2 月 2 日閲覧, https://sefiks.com/2020/02/17/face-recognition-with-facebook-deepface-in-keras/,

[42] 楽しみながら学ぶ機械学習 / 自然言語処理入門, “畳み込み演算と転置畳み込み演算を理解する”, 2025 年 2 月 2 日閲覧, https://data-analytics.fun/2021/11/23/understanding-convolution/,

[43] David M. Blei, Andrew Y. Ng and Michael I. Jordan “Latent Dirichlet Allocation”, 2003

[44] GMO リクルート “トピックモデルとは？基本概念とその活用方法”, 2025 年 2 月 2 日閲覧, https://recruit.gmo.jp/engineer/jisedai/blog/topic-model/

[45] Maxine, “Latent Dirichlet Allocation. Bookdown”, 2025 年 2 月 2 日閲覧, https://bookdown.org/Maxine/tidy-text-mining/latent-dirichlet-allocation.html

[46] Li, F, “How we changed unsupervised LDA to semi-supervised GuidedLDA”, 2025 年 2 月 2 日閲覧, https://www.freecodecamp.org/news/how-we-changed-unsupervised-lda-to-semi-supervised-guidedlda-e36a95f3a164

[47] Spotify, “Spotify Web API documentation”, 2025 年 2 月 2 日閲覧, https://developer.spotify.com/documentation/web-api

[48] Spotify , “Spotify.”, 2025 年 2 月 2 日閲覧, https://open.spotify.com/intl-ja

[49] Genius, “Genius について”, 2025 年 2 月 2 日閲覧, https://genius.com/Genius-japan-genius-about-genius-annotated

[50] Jitera, “Web スクレイピング完全ガイド：データ収集の基本から法的枠組みまで”, 2025 年 2 月 2 日閲覧, https://jitera.com/ja/insights/48994

[51] TIS Inc. “TIS, AI を活用した新サービスを提供開始”, 2025 年 2 月 2 日閲覧, https://www.tis.co.jp/news/2018/tis_news/20180410_1.html

[52] Chakki-Works, “chABSA dataset.”, 2025 年 2 月 2 日閲覧,
https://github.com/chakki-works/chABSA-dataset/tree/master

[53] cl-tohoku, “BERT pre-trained models for Japanese”, 2025 年 2 月 2 日閲覧,
https://github.com/cl-tohoku/bert-japanese

[54] ChocottoPro, “教師なし学習と教師あり学習の違いとは？わかりやすく解説”, 2025 年 2 月 2 日閲覧,
https://chocottopro.com/?p=658