

発話の感情要素と感情強度を考慮した楽曲推薦手法の検討

池田 将† 酒向 慎司†

†名古屋工業大学

1 はじめに

近年の音楽配信サービスでは、視聴履歴などを利用してユーザの好みに合う楽曲を推薦する仕組みが導入されている。一方で、ユーザの現在の状況や心情などに応じた楽曲推薦を行うための研究もなされている。本研究ではユーザの感情を考慮した楽曲推薦手法について取り組む。この研究の狙いは、楽曲推薦をより柔軟に行うほかにも、ユーザの状況や心情に適した楽曲を推薦することで、心理状態を安定させたりすることなどの音楽療法への応用が考えられる。

音楽療法への応用のため、本研究では同質の原理に基づいて楽曲推薦を行う。この原理は音楽療法の基本原理であり、聴取者はその時の感情に同質な印象の音楽を望むというものである。この原理を利用するために、ユーザの感情を推定することが本研究の課題である。

2 提案手法の概要

2.1 Arousal-Valence (AV) 空間

本研究では、感情を表す空間として Russel [1] が提案した Arousal-Valence(AV) 空間を採用する。AV 空間上の座標である AV 値は人の感性や感情を表すことができ、本研究では原点からの角度は感情の意味合い（感情要素）、原点からの距離は感情の強さ（感情強度）を表すものとする。

2.2 提案手法の手順

提案手法の概略図を図 1 に示す。システムは、ユーザの発話として言語情報（テキスト）と音声情報が与えられる。最終的な楽曲推薦を行うための材料として、言語情報から感情要素、音声情報から感情強度を推定する。それぞれの推定方法の手順を以下に示す。

言語情報から感情要素

1. 言語情報から発話テキストを取得する
2. 発話テキストを形態素解析し、コンテンツワードを決定する
3. コンテンツワードから感情要素を推定する

音声情報から感情強度

1. 音声情報から音声データを取得する
2. 感情要素が AV 空間上の 4 つの象限のどちらに向いているかにより感情強度認識モデルを選択する

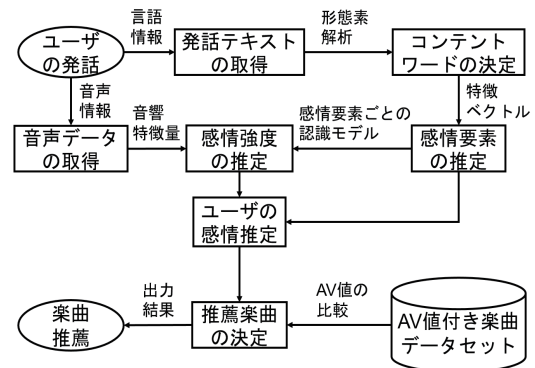


図 1: 提案手法の概略図

3. 音声データから取り出された音響特徴量と認識モデルを使って感情強度を推定する

推定された感情要素と感情強度から AV 値に変換する。推薦楽曲は AV 値付き楽曲データセットを用いて、ユーザの感情を表す AV 値と楽曲の AV 値を比較し、最も値の距離が近い楽曲を推薦楽曲とする。本研究は特に発話から感情状態を推定することに注力する。

3 発話による感情要素と感情強度の推定方法

3.1 言語情報による感情要素推定

感情要素推定手法の例を図 2 に示す。図 2 の 1 でユーザの発話から言語情報として発話テキストを取得し、図 2 の 2 で発話テキストを Mecab によって形態素解析することで、動詞、形容詞、形容動詞、名詞を取り出す。これらを感情を表す言葉としてコンテンツワードと呼ぶ。

図 2 の 3 で Word2Vec によりコンテンツワードを感情語に変換する。ここでは、日本語 Wikipedia の約 100 万文のテキストを用いて学習されたモデル (50 次元) [2] を利用した。このモデルを用いて、コンテンツワードと本間の研究 [3] で作成された AV 値が既知な感情語

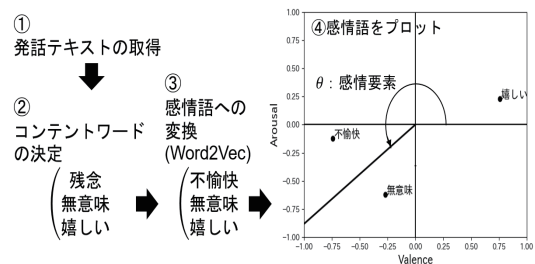


図 2: 感情要素推定手法の例

A study on music recommendation system using emotional elements and emotional intensity of speech
 †Sho IKEDA †Shinji SAKO
 †Nagoya Institute of Technology

表 1: 認識モデルに使用する学習データセット

OGVC 演技音声 (4 話者, 感情強度 1~3)		
象限	感情	発話数
第 1 象限	喜び, 驚き	540 (各強度 180)
第 2 象限	嫌悪, 怒り	480 (各強度 160)
第 3 象限	恐れ, 悲しみ	492 (各強度 164)
第 4 象限	受容, 期待	480 (各強度 160)

との類似度を計算し, コンテントワードと最も類似度が高い感情語を決定する. 図 2 を例にすると, 「残念」というコンテントワードは感情語に存在しないため, 最も類似度が高い「不愉快」という感情語に変換された.

図 2 の 4 で感情語を AV 空間上にプロットする. 図 2 では第 3 象限にプロットされた感情語が最も多いため, 第 3 象限に位置する「不愉快」と「残念」の感情要素の平均をとることでユーザの感情要素を決定する.

3.2 音声情報による感情強度推定

感情要素ごとの感情強度認識モデルを構築する. 学習データセットとして感情評定値付きオンラインゲーム音声チャットコーパス (OGVC)[4] を使用した. データセットの詳細を表 1 に示す. 表 1 の音声データから openSMILE [5] を使用して音響特徴量を抽出する. 本研究では, INTERSPEECH2009 の特徴量セットである MFCC, F0 など 384 次元の特徴ベクトルを抽出する. 抽出した特徴量を用いて, 多項式カーネル関数を使用した SVM による認識モデルを作成した.

4 評価実験

4.1 実験準備

評価データとして, 日本語 MULTEXT 韻律コーパス [6] 内の 10 文章とそれに対応する感情が入った 10 発話を使用した. この評価データに感情要素と感情強度のラベル付けをするため, 20 代男性 3 人に協力してもらった. 感情要素ラベルは, 評価データの 10 文章を見てもらい, それぞれの文章が AV 空間の象限をそれぞれ 2 分割した 8 段階中のどこに位置するのかを評価させた. 感情強度ラベルは, 10 発話分の音声を聞いてもらい, それぞれの音声の感情の強度を 1 (感情が弱い) から 3 (感情が強い) の 3 段階で評価させた.

4.2 実験結果

実験結果を表 2 に示す. 感情要素推定実験の評価条件として, 実験 1-1 (感情要素ラベルが一致), 実験 1-2 (感情要素ラベルに対応する象限が一致), 実験 1-3 (AV 空間を 8 分割したときに隣接する値まで正解とみなす) のそれぞれでシステムの正解率を出した. また, 感情要素ラベルが 3 人中 2 人以上一致している文章を使用したため, サンプル数は 7 となっている. 感情強度推定実験の評価条件として, 実験 2-1 (システムにより感

表 2: 実験結果

感情要素推定		感情強度推定	
評価条件	正解率	評価条件	正解率
実験 1-1	14.3	実験 2-1	33.3
実験 1-2	28.6	実験 2-2	44.4
実験 1-3	57.1	実験 2-3	80.0

情要素推定された象限をもとに認識モデルを選択), 実験 2-2 (感情要素ラベルに対応する象限をもとに認識モデルを選択), 実験 2-3 (実験 2-2 の手法で, サンプル数が異なる) のそれぞれでシステムの正解率を出した. また, 実験 2-1 と 2-2 は感情要素ラベルに対応する象限が 3 人中 2 人以上一致している文章を使用したため, サンプル数は 9 で, 実験 2-3 は感情要素ラベルに対応する象限が全員一致している文章を使用したため, サンプル数は 5 となっている.

4.3 考察

感情要素推定実験では, 実験 1-3 の条件では有用性が見られるが, 実験 1-1 と実験 1-2 の条件では精度が低い. 精度を向上させる方法として, Word2Vec のコーパスをチャット用のコーパスに変更することなどが考えられる. また, 感情強度推定実験では, 実験 2-3 の条件で有益な結果が見られたことから, 正解率が感情強度のあいまいさに依存すると考えられる.

5 むすび

本稿では, ユーザの感情を認識する手法, およびその評価を示した. 今後はシステム全体の精度を上げるとともに, 推定感情から楽曲を推薦するシステムの構築に取り組む.

参考文献

- [1] James Russell, "A circumplex model of affect", *Journal of Personality and Social Psychology*, Vol.39(6), pp.1161-1178, Dec. 1980
- [2] 株式会社白ヤギコーポレーション, "日本語 word2vec モデル"
<http://public.shiroyagi.s3.amazonaws.com/latest-ja-word2vec-gensim-model.zip>, 2017
- [3] 本間喜子, "単語の感情価と覚醒度に基づいた単語刺激の作成", 愛知工業大学研究報告, 第 49 号, Mar. 2014
- [4] 有本泰子, 河津宏美, "音声チャットを利用したオンラインゲーム感情音声コーパス", 日本音響学会 2013 年秋季研究発表会講演論文集, 1-P-46A, pp.385-388, Sep. 2013
- [5] Eyben Florian, et al., "openSMILE", *Proceedings of the international conference on Multimedia-MM*, Vol.10, 2017
- [6] S. Kitazawa, et al., "Japanese MULTEXT: a Prosodic Corpus", *Proc. International Conference on Language Resources and Evaluation*, pp.2167-2170, 2004