

# 入力文書の印象と感情に基づく楽曲提供の一手法

理学専攻 情報科学コース 菅野沙也 (指導教員：伊藤貴之)

## 1 はじめに

我々は小説などの文書の印象を様々な形で表現する。特にレビューを書くなどの形で別の文章を用いたり、イラスト化するなどの視覚的手法を用いたりすることが多い。しかし、これらの手段には作品のネタバレやストーリーへの先入観を生む恐れがある。そこで文書の印象を表現する別の手段として、「音楽」で文書の印象を表現することを提案する。音楽にはイメージ誘導効果・行動誘導効果・感情誘導効果があり、これは音楽への興味の有無に関わらず、すべての人に効果があると行動心理学の観点からも語られている。さらに、文学と音楽という異なる種類の芸術をリンクさせることで、新たな感動や刺激が得られることも期待できる。

## 2 入力文書と生成楽曲の条件

本研究を行うにあたり、入力文書と生成楽曲にはいくつかの条件を設定した。

入力文書の条件は以下の通りである。

- ユーザによる改行を用いた場面分け

生成楽曲の条件は以下の通りである。

- コード・リズム進行の合成による楽曲生成
- 1 場面に対して 1 つのコード・リズム進行の適用
- ユーザそれぞれの感性に合わせた楽曲の生成

## 3 提案手法

本手法は大きく分けて 4 つの処理段階で構成される。

- 事前調査：感性語と音楽特徴量の選出
- 辞書作成：単語に対して各感性語の度合いを数値で与える辞書の作成
- ユーザごとの感性データ生成：ユーザの感性を楽曲に反映させるためのデータ生成
- 対話処理：ユーザーの文書入力に対する楽曲生成

各処理段階の詳細について以下に論述する。

### 3.1 事前調査

ここでは、文書解析の基準となる感性語対と、コード進行やリズムの進行の選出基準となる音楽特徴量を選出する。我々は感性語対と音楽特徴量の選出のためにユーザーアンケートを実施した。被験者にはコードとリズムの音源を聴いてもらい、それぞれ候補となる感性語対の度合いを 5 段階評価で回答してもらった。

この結果をもとに、どの感性語対とも相関の絶対値が 0.4 以下となる音楽特徴量、またはどの音楽特徴量とも相関の絶対値が 0.4 以下となる感性語対を対象から除外した。その結果選出された感性語対と音楽特徴量を表 1 に示す。

表 1: 選出したコード進行およびリズム進行の感性語と音楽的特徴

選出したコード進行の感性語	選出したコード進行の音楽特徴量
<ul style="list-style-type: none"><li>・ 明るい-暗い</li><li>・ 楽しい-切ない</li><li>・ 軽快-しっとり</li><li>・ 元気な-落ち着いた</li></ul>	<ul style="list-style-type: none"><li>・ 音程の平均的な高さ</li><li>・ 音の平均的な広がり</li><li>・ 音の平均的な厚み</li><li>・ 不協和音度</li><li>・ メジャー/7th/m7 の割合</li></ul>
選出したリズム進行の感性語	選出したリズム進行の音楽特徴量
<ul style="list-style-type: none"><li>・ 速い-遅い</li><li>・ 軽い-重い</li><li>・ 派手-地味</li></ul>	<ul style="list-style-type: none"><li>・ ドラムセットの各部位の割合</li><li>・ すべての音符数</li><li>・ 16 分音符の割合</li><li>・ 3 連符の割合</li></ul>

### 3.2 辞書作成

本手法では、入力文書とコード・リズム進行を比較できる形にするために数値化する必要がある。そこで、まず一般的な日本語の単語について、3.1 節で選んだ 7 つの感性語対の度合いを数値化する辞書を作成する。本研究では以下の 2 種類の辞書を試作して一方ずつを実験に用いた。

#### 3.2.1 感性極性辞書

感性極性辞書とは、日本語を構成する各内容語に対し、選出した各感性語対のいずれの感性に近いかを数値化する辞書である。具体的には感性語対の一方を +1 に、もう一方を -1 に対応させ、[-1,+1] の範囲の実数値を各内容語に対し算出する。本研究では感情極性抽出手法 [1] を適用し、表 1 に示した各感性語対を種単語として内容語の印象を数値化している。

#### 3.2.2 word2vec による辞書

word2vec を用いて単語と各感性語のコサイン類似度を算出する辞書を作成した。word2vec とは入力データから単語同士の関係を学習し各単語を多次元ベクトルとして保存する手法である。入力データには、一般的な日本語の単語を取録した辞書データ、童話のテキスト、各感性語の類似語データを用いた。

### 3.3 ユーザごとの感性データ生成

続いて、コード進行およびリズム進行の音楽的特徴量とそれに対する各ユーザーの印象の関係を学習し、ユーザごとの感性データを生成する。まず、各ユーザーにコード進行およびリズム進行のサンプル音源を聴いてもらい、3.1 節で選出した感性語への適合度を回答してもらう。以後、この適合度を印象値と称する。ま

## コード進行

印象値  $f_i = a_1 \times [\text{maj の割合}] + a_2 \times [\text{M7 の割合}]$   
 $+ a_3 \times [\text{m7 の割合}] + a_4 \times [\text{音高の平均}]$   
 $+ a_5 \times [\text{音幅の平均}] + a_6 \times [\text{音厚の平均}]$   
 $+ a_7 \times [\text{不協和音度}]$

## リズム進行

印象値  $f_i = b_1 \times [\text{全ての音符数}] + b_2 \times [\text{16分音符の割合}]$   
 $+ b_3 \times [\text{3連符の割合}] + b_4 \times [\text{金物の割合}]$   
 $+ b_5 \times [\text{バスドラムの割合}]$   
 $+ b_6 \times [\text{タムの割合}] + b_7 \times [\text{スネアの割合}]$

図 1: 感性データ生成のための式

た、印象値を推測するための式を音楽的特徴量の線形結合と仮定し、重回帰分析を用いて図 1 に表す式の各係数を算出する。ここで算出された各係数がコード進行とリズム進行の印象値となる。これより、ユーザーごとの印象の違いを考慮した楽曲生成が可能になる。

## 3.4 対話処理

### 3.4.1 文書解析

まずユーザーが入力した文書に対して文書解析を適用する。オープンソース形態素解析エンジン MeCab を用いて入力文書を形態素解析し、辞書を参照して各単語を数値化する。続いて tf-idf 法を用いて各単語に重み付けを行う。最後に式 1 を用いて場面  $s$  の感性語対  $i$  に対する印象値  $g_{s_i}$  を算出する。

$$g_{s_i} = \frac{\sum_{k=1}^{T_s} tfidf_{s_k} \cdot v_{i_k}}{\sum_{k=1}^{T_s} tfidf_{s_k}} \quad (1)$$

- $T_s$ : 場面  $s$  に登場する単語の個数
- $tfidf_{s_k}$ : 場面  $s$  における  $k$  番目の単語の tf-idf 値
- $v_{i_k}$ :  $k$  番目の単語の感性語対  $i$  に対する印象値

### 3.4.2 素材選出

次に楽曲の素材となるコードとリズムを選出する。我々は 3.4.1 項の手法を用いて文書中の  $N_{scene}$  個の各場面の印象値を算出し、これを  $N_{scene}$  個の頂点と  $N_{scene} - 1$  個の稜線から構成される軌跡として扱う。この軌跡と類似する軌跡を描く値を持つコード・リズム進行を、各場面の素材として選出する。具体的には以下の式の値が最小となるコード・リズム進行を選出する。ここで  $f_i$  は  $i$  番目の場面でのコードまたはリズムの印象値、 $g_i$  は文書の印象値、 $f'_{i,i+1}$  や  $g'_{i,i+1}$  はそれらの隣接場面間の差分である。

$$\min \left( \alpha \sum_i^{N_{scene}} \|f_i - g_i\| + (1 - \alpha) \sum_i^{N_{scene}-1} \left( 1 - \frac{f'_{i,i+1} \cdot g'_{i,i+1}}{|f'_{i,i+1}| |g'_{i,i+1}|} \right) \right) \quad (2)$$

軌跡類似度の算出により、各場面の印象だけでなく、全体の流れや印象の変化を考慮した楽曲を生成することができると考えた。

### 3.4.3 楽曲生成

選出したコード進行とリズム進行を合成することで楽曲を生成する。これらの MIDI ファイルはそれぞれ単一のトラックで構成されることを前提とし、この各々

のトラックをユーザに提供する楽曲の MIDI ファイルの 1, 2 番目のトラックにコピーする。

## 4 実行結果

本手法において、以下の 3 つの場面を持つ文書を入力し、楽曲を生成した。文書解析に用いる辞書は感性極性辞書を採用し、感性データは任意のユーザのものを用いた。

**場面 1** あなたが微笑みながら歌う横顔はとても美しかった

**場面 2** わたしはその横顔を見てあなたに恋をした

**場面 3** しかし失恋した

入力文書の各場面の印象値と選出されたコード進行およびリズム進行の印象値を図 2 に示す。

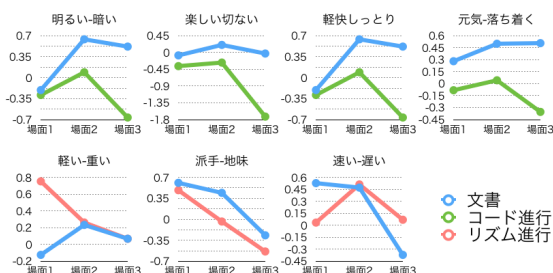


図 2: 実行結果

まずコード進行に関して述べる。場面 1 ではほぼすべての感性語対で座標距離の近いコード進行が選出されていることがわかる。場面 2 と場面 3 では座標距離は遠いものの前後の場面からのベクトルの変化が似ている部分が見られる。

続いて、リズム進行について述べる。「派手-地味」は場面 1 と場面 3 で座標距離が近いリズム進行が選出されており、全体のベクトルの軌跡も類似していると考えられる。一方で「軽い-重い」「速い-遅い」は場面 1 の座標距離が遠く、場面 1 から場面 2 へのベクトル遷移も大きく異なる。これはリズム進行選出のアルゴリズムよりもむしろユーザの感性データ生成に原因があると考えられる。文章の印象値に距離の近い印象値をもつリズム進行が存在しなかったためこのような結果になってしまった可能性があり、この点について改良が必要である。

## 5 まとめと今後の課題

本研究では、文書の印象値と音楽特徴量の比較により、入力文書の印象や感情に基づいた楽曲を提供することで文書の印象表現を促す手法を提案した。今後の課題として、どのような辞書が本手法においては有効なのか明らかにする検証、重要度の高い単語に関するユーザによる印象定義の実装などが挙げられる。

## 参考文献

- [1] 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, Vol. 47, No. 2, pp. 627-637(2006).