

機械学習を用いた音楽フレーズグルーピング法の獲得

山上信一 † 池田 剛 ‡ 但馬 康宏 ‡ 小谷 善行 ‡

† 東京農工大学情報コミュニケーション工学科 ‡ 東京農工大学大学院工学教育部

1. はじめに

音楽フレーズに対しグルーピングを行う方法はいくつか存在する。グルーピングは客観的な音楽認知の手助けとなり、新しいアプローチによる音楽検索の実現や、その他、自動作曲や作曲支援、自動伴奏システムなどへの応用が期待できるものと考えられる。

しかし既存のグルーピング理論に関して言えば、それは区切れとなりうる可能性を論じるだけで、実際にこれらの理論を用いてグルーピングを行うためには主観的要素を必要とする場合が多い。

そこで我々はポピュラー音楽などにおける歌詞の区切れに着目することにより、楽曲の区切れを推定するための方法を提案する。一般に、歌詞と同時進行するメロディは相互に関係性のあるものであり、歌詞に存在する構文的な区切れが、メロディの区切れに影響していると考えられる。そこでこの歌詞の区切れと各種のグルーピング理論で求められる区切れには共通性があると仮定する。本研究ではこれら二種類の区切れの関係を学習し、学習結果より、与えられた楽曲の区切れを推定する。ここでは音楽学におけるグルーピング理論として GTTM を用いる。

なお実験には RWC 研究用音楽データベースのポピュラーミュージックを使用した。

2. 原理

2.1 学習方法

本研究では、歌詞の区切りと音楽理論の区切れの関係について着目する。これら楽曲における二種類の区切れの発生状況を観測し、それらの対応関係を学習する。区切れを導出するために利用する音楽理論である GTTM は、グルーピング選好規則 (GPR) のうち、GPR2 と GPR3 を用いる。この理論の定義を満たす判定式を作成し、

音符情報を入力した際の判定式の値を区切りの情報とする。学習は GTTM による区切れを入力とし、歌詞の区切れを教師値として行う。数値の多入力となるので、学習はニューラルネットを用いた。

今回歌詞データは RWC 研究用音楽データベースに含まれている歌詞情報に適当な区切れを付加し、その区切れを歌詞の区切れとして実験を行った。

2.2 ニューラルネットの適用

機械学習として、次の図で示すようなニューラルネットを用いる。学習方法はバックプロパゲーション法である。

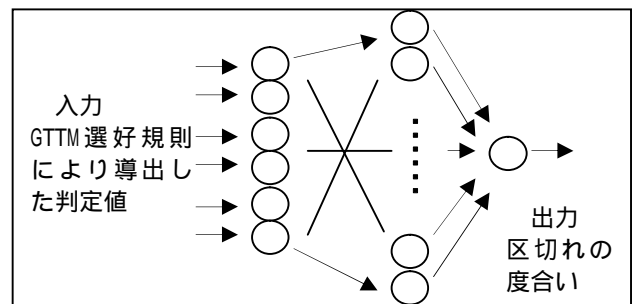


図1 ニューラルネット

6つの入力、は、楽曲中の対象となる音符と次の音符との間で、GTTM グルーピング選好規則が適用されるかを表したものである。これを入力データとする。学習の際の教師値となる歌詞の区切れについては、区切れの有無を二値で表現する。なお、学習効果を確認するために、入力データについて、区切れの判定式の値を、閾値を用いて二値で表したものと、値を直線的に正規化したものの二通りについて比較実験を行う。

3. 実験および結果

3.1 実験方法

実験は、学習用に10曲分、2216音符分のデータと、テスト用のオープンデータとして、4曲分、871音符分のデータを作成し、入力データを変更、ニューラルネット中間層のユニット数の変更など、以下のものを行った。

A Selection of Correct Grouping in Music Phrase by Machine Learning
 Shinichi Yamagami † Tuyoshi Ikeda ‡ Tajima Yasuhiro ‡
 Kotani Yosiyuki ‡
 † Department of Computer, Information and Communication Science, Tokyo University of Agriculture and Technology
 ‡ Graduate school of Information Science, Tokyo University of Agriculture and Technology

(実験 1)二値データ入力による実験

学習データとして二値データをを入力する基本実験である。ニューラルネット中間層のユニットの個数は12とした。

(実験 2)直線的に正規化したデータの入力におけるユニット数変更実験

正規化を行ったデータの入力に対して、中間層のユニット数を、1、3、6、10、12、18、24と変更して学習結果の様子を見た。

(実験 3)二値データと正規化を行ったデータの同時入力による実験

対象となる音符の、GTTMによる二値データと、直線的に正規化を行ったデータを同時に入力し、学習結果を見た。この実験での中間層のユニット数は24である。

(実験 4)前後2つの音符情報を含む正規化を行ったデータ入力による実験

対象となる音符と、その前後2つの音符のGTTMグルーピング選好規則の適用状況を同時に入力し、学習結果を見た。この実験では中間層のユニット数を36とした。

学習回数はすべての実験において5万回とし、初期値の決定により局所最適解におちいる可能性があるため、それぞれの実験において数回の試行を行った。

今回の実験では、出力値が0.5以上であれば区切れ、0.5未満であれば区切れないとし、教師値との正誤を見た。

3.2 実験結果

実験は、それぞれ上に示した条件で学習データを入力し学習させ、テストデータを入力した際の出力と教師値とを比較した。表1に実験の結果を示す。表は各実験において、学習後テストデータを入力した際のニューラルネットの出力を教師値により分類し、0.5の閾値で分けたものの音符数である。(2)の実験についてはユニット数ごとに結果を載せている。

4. 考察

今回の実験では、教師値となる歌詞の区切りに、楽曲に付属する歌詞データのスペース区切れを使用した。GTTMによる区切れとこの歌詞による区切れの個数に開きがあり、学習精度を

表1 実験結果

実験 No/ユニット数	歌詞の区切れなし		歌詞の区切れあり	
	NN 出力 0.5 以上	NN 出力 0.5 未満	NN 出力 0.5 以上	NN 出力 0.5 未満
(1) / 12	173	611	19	68
(2) / 1	23	761	14	73
(2) / 3	11	773	33	54
(2) / 6	25	759	31	56
(2) / 10	7	777	32	55
(2) / 12	2	759	21	66
(2) / 18	11	773	28	59
(2) / 24	17	767	20	67
(3) / 24	14	770	38	49
(4) / 36	2	782	26	61

低減させる原因の一つとなっている。ただし、歌詞の区切れはそのとらえ方によって大きく変わるものであり、区切れの度合いを文節レベル程度にすることにより、適度な精度が得られるものと考えられる。楽曲をどの程度に区切るのかも問題となるが、現在1曲に対し八つほどの区切れを導出しているところをより細かいレベルで区切ろうとするなら、歌詞のデータをより細かく区切ることで対応できる。

今回の実験では、歌詞の区切れのある部分は、実験データでの歌詞の区切れが少ないことからその正答率が25%程度であった。ただし区切れとされない部分については、一番結果の悪い実験でも97%の正解率であり、歌詞によるグルーピングの有効性が示されている。

5. まとめ

以上のように、歌詞の区切れと音楽理論を合わせた新しいグルーピングの手法を示した。実際にニューラルネットを使用して学習を行った結果、GTTMによる区切れの状況から歌詞の区切れが導出できており、区切れの推定が十分可能であると言える。

参考文献

- [1]長嶋洋一他, コンピュータと音楽の世界, 共立出版, 1998
- [2]RWC 研究用音楽データベース, ポピュラー音楽データベース