

自己組織化マップを用いた類似楽曲の検索*

伊藤 文也[†] 相場 亮[‡]

芝浦工業大学 システム工学部 電子情報システム学科[§]

1 はじめに

近年インターネットやコンピュータの発展に伴い、MIDI や MP3 が急速に普及してきている。それによりデジタル化された楽曲の検索において効率を良くするための要求が出てきた。従来は作曲家、曲名、ジャンルといった情報から検索がなされていた。iTunesStore[1] 等の大手音楽配信サイトでもアーティスト名、アルバム名、ジャンル等での検索しかできない。しかしこのような従来の方法では楽曲についての情報があまり無い状態からの欲しい楽曲の検索は不可能である。

類似楽曲の検索に関する研究としては、昨年度本研究室の杉本によるグラフマッチングという手法を用いての類似楽曲の検索や、ハミングによる類似楽曲検索 [3] が行われている。

本研究では類似楽曲の検索に自己組織化マップを用いて検索することを提案する。

2 自己組織化マップについて

自己組織化マップ (SOM: Self-Organizing Maps)[4] とは、ニューラルネットワーク [2] の教師なし学習の一種である。まずはニューラルネットワークについて説明する。

2.1 ニューラルネットワーク

ニューラルネットワークとは人間の脳の細胞であるニューロンのしくみをまねた情報処理機構である。ニューロンは、多数の入力を持っており、それぞれに結合荷重と呼ばれる重みがついており、入力と結合荷重の積の総和が閾値を越えたならば 1、越えなければ 0 を出力するという働きをする。そのニューロンが多数接続されたものをニューラルネットワークと呼ぶ。

ニューラルネットワークには学習能力があり、学習能力とは、結合荷重の値を適切な値に随時更新していくことをいう。

ニューラルネットワークの代表的な構成には

1. 階層型ネットワーク
2. 相互結合型ネットワーク

があるが、本研究で取り扱う自己組織化マップは 1. の階層型ネットワークに属する。

2.2 自己組織化マップ

自己組織化マップ (SOM: Self-Organizing Maps) とは、Kohonen によって開発された大脳皮質の神経機能をモデル化したニューラルネットワークで、入力層と競合層からなる。自己組織化マップは非階層型クラスタリングの手法である「K-means」(K 平均法) に似ている。自己組織化マップは厳密にはクラスタリング手法ではないが、明確なクラス境界を与えない以外はクラスタリングと同じ意味を持つ。また二次元格子状にニューロンを配列したネットワークを持ち、入力データによって競合学習モデルによる組織化がなされる。

学習の終了段階では、入力データのうち特徴がよく似たもの同士が近くに配置され、似ていないもの同士が遠くに配置されるような分類フィルタとして機能するようになる。入力データの次元数にかかわらず二次元格子状に配列されるため、有効な次元削減技術の一つとみなせ、その結果入力データのマップにおける位置関係を見るだけで、その類似性を直感的に把握することが可能である。

3 MIDI について

MIDI (Musical Instrument Digital Interface) とは、電子楽器の演奏データを機器間でデジタル転送するための規格である。物理的なインターフェース、通信プロトコル、データ形式、ファイル形式 (Standard MIDI File, SMF) などからなる。SMF にはフォーマット 0 とフォーマット 1 があり、本研究では取り扱いのしやすさからフォーマット 1 を選ぶ。

4 提案手法について

今回は以下の特徴に着目する。

- 似たもの同士が近くに配置されるという自己組織化マップの特徴

MIDI ファイルを入力データとし、そこから「サビ」、「メロディ」、「テンポ」の 3 次元データの特徴量として抽出する。データの抽出方法は後述する。「サビ」、「メロディ」、「テンポ」の 3 次元データを抽出したあと、それを自己組織化マップに入力する。そして自己組織化マップの「似たもの同士が近くに配置される」という特徴を利用して「サビ」、「メロディ」、「テンポ」の似たもの同士が近くに配置された出力結果を、検索結果とする。そうすることにより、楽曲同士の類似度が直感的に把握することができる。これが本研究の提案手法である。

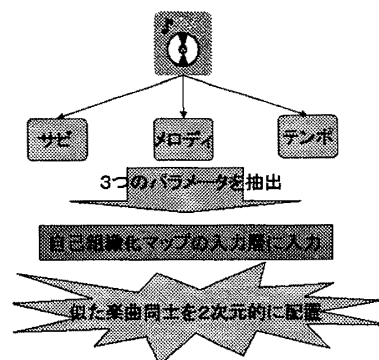


図 1: 提案手法

「サビ」、「メロディ」、「テンポ」の数値化の仕方であるが、SMF のバイナリデータをテキスト表示してくれるソフトがあるので、本研究ではそれを利用し、バイナリデータをテキスト化したあと、必要な部分だけを抽出し、数値化した。

まず「サビ」の区間を秒で指定し、それをチェックに変換する。そこから「NoteOn」と呼ばれるイベントの数値をチェック単位で抽出する。「NoteOn」とは音階 (ドレミファソラシド) に相当するものである。抽出した数値を羅列し

* Searching similar music by Self-Organizing Maps

[†] Fumiya Ito

[‡] Akira Aiba

[§] Department of Electronic Information Systems, College of System Engineering, Shibaura Institute of Technology

て一個の数値として、自己組織化マップの入力に与える。「メロディ」についても同様である。

「テンポ」については、SMFの中でミリ秒で指定されているので、それを抽出し、そのまま自己組織化マップの入力に与える。

5 実験

実験で使う楽曲は「RWC 研究用音楽データベース：ポピュラー音楽&著作権切れ音楽」の全100曲を用いた。その100曲全てから特徴量を抽出し、自己組織化マップに入力した。自己組織化マップの出力結果は図2を参照。また、聞き比べてもらう楽曲は全て、前述した100曲内から選んだ。

被験者10人に対し、アンケートを実施した。まず被験者に聞き比べてもらう楽曲は以下のように指定した。グループ1の「ランダム」は、被験者に自由に楽曲を選んでもらったことを意味する。

	キー楽曲	キー楽曲と比較するための曲
グループ1	キー楽曲ランダム	比較曲ランダム I
		比較曲ランダム II
グループ2	キー楽曲8番	比較曲9番
		比較曲99番
グループ3	キー楽曲36番	比較曲38番
		比較曲96番

表1: 被験者に聞き比べてもらう楽曲一覧

表1の全9曲が、被験者に聞いてもらった楽曲である。2グループ目と3グループ目の楽曲の位置は図2を参照。2グループ目が青文字、3グループ目が赤文字である。

アンケート内容は「サビの流れは似ているか」、「メロディの流れは似ているか」、「曲の速さは似ているか」、「全体的な印象は似ているか」の4項目を、4段階の評価を付けてもらった。「最終的な判断(キー楽曲と聞き比べた楽曲が最終的に似ているかどうかの判断)」の項目では、4段階の評価を付けてもらった。4段階評価のレベルは、1が「まったく似ていない」、2が「どちらかといえば似ていない」、3が「どちらかといえば似ている」、4が「似ている」である。また、3グループそれぞれの、キー楽曲と聞き比べてもらった2曲のうち、どちらがキー楽曲に似ているかの判断をチェックしてもらった。

6 実験結果と考察

結果は以下ようになった。「最終的な判断」で4段階評価を付けたときに、2以下の評価を付けた人は「似てない」に、3以上の評価を付けた人は「似てる」に分類した。その結果が表2である。

	類似度	似てない	似てる
グループ1	キー&比較曲 I	8人	2人
	キー&比較曲 II	6人	4人
グループ2	8&9	9人	1人
	8&99	10人	0人
グループ3	36&38	2人	8人
	36&96	3人	7人

表2: 実験結果(最終的な判断の評価結果)

今回被験者に聞き比べてもらった楽曲は全9曲と少なく、また被験者も10人と少ないが、表2の曲番号36&38と曲番号36&96の最終的な判断の結果を見てみると、似

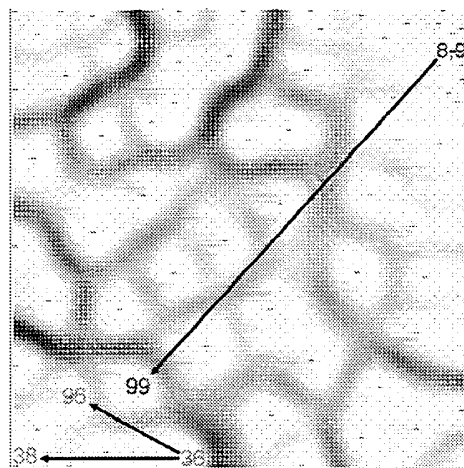


図2: SOMの出力結果

てると評価した人がそれぞれ8割と7割で比較的高い評価を得ている。これらの範囲内で判断すると、類似楽曲の推薦方法として自己組織化マップを用いることは、全体的に見ると有効とは言えない。

今回有効に機能しなかった原因として、図2の出力結果の右上部分のエリアに、曲が不自然に集中したことが原因と思われる。キー&比較曲Iで「似てない」と判断した8人のうち、半分以上がキー楽曲として右上部分のエリアから選んでいたことが分かった。またキー&比較曲IIで「似てない」と判断した6人中4人がキー楽曲として右上部分のエリアから選んでいた。同じ領域内にあるにもかかわらず、聞き比べた結果「似てない」と評価されてしまった原因の1つとして、右上部分のエリアは「その他」として分類されてしまったからではないかと思われる。右上部分の「その他」のエリアを無くするためには、球型の自己組織化マップ[5]を用いることで解決できると思われる。何故なら自己組織化マップの四隅周辺は未学習領域になりやすいため、不完全な学習部分ができてしまうからである。しかし、球型自己組織化マップを用いれば、四隅周辺がなくなるので不完全な学習部分が解消される。

「似てない」と評価された他の原因として、自己組織化マップの入力値そのものが原因であることも考えられる。今回「サビ」、「メロディ」の入力値は音階情報だけだったので、その解決策として音階情報以外の情報も付加して入力することで精度の高い出力結果を得ることが可能と思われる。

また、人によってそれぞれ重視する特徴量が違うので、特徴量ごとに重みを付けられるようにすることで更なる精度の向上が期待できる。

今後の課題としては、球型自己組織化マップの使用、特徴量のデータそのものの改良、特徴量ごとの重み付けが必要である。

参考文献

- [1] iTunes Store
<http://www.apple.com/jp/itunes/store/>
- [2] 熊沢逸夫:「学習とニューラルネットワーク」、森北出版株式会社 1998
- [3] 蔭山哲也、高島洋典:「ハミング歌唱を手掛りとするメロディ検索」、電子情報通信学会論文誌 Vol.77 No.8 Page1543-1551, 1994
- [4] 徳高平蔵、岸田悟、藤村喜久郎:「自己組織化マップの応用」、海文堂出版 1999
- [5] 徳高平蔵、大北正昭、藤村喜久郎:「自己組織化マップとその応用」、シュプリンガー・ジャパン株式会社 2007