

SVMを用いたリズム感のモデル化とドラムパターンの自動生成

浜田鉄平[†] 但馬康宏[‡] 小谷善行[†]

[†] 東京農工大学工学府情報工学専攻

[‡] 東京農工大学共生科学技術研究院先端情報科学部門

あらすじ 入力された旋律譜に対して、人間が聴くうえで“良い”と感じられるドラム譜を自動生成することを目的に、本研究では人間のリズムに対する評価能力をサポートベクターマシンによりモデル化した。さらに、得られたモデルを遺伝的アルゴリズムに組み込むことで、旋律譜に適応したドラム譜の自動生成を行った。アンケート結果によれば、自動生成されたドラム譜はランダムに生成したドラム譜よりも良いと感じられ、また人手で生成したドラムパターンに近い良さをもっていることがわかった。本研究では、モデル生成の際に不特定多数のユーザーから収集した単純な事例データのみを用いていることから、専門家による事例データを用いることなく良いドラム譜を生成することが可能であることが示された。

Modeling of Rhythmic Sense by SVM and Automatic Generation of Drum Pattern

Teppei Hamada[†] Yasuhiro Tajima[‡] Yoshiyuki Kotani[†]

[†]Department of Computer and Information Sciences, Tokyo Univ. of Agriculture and Technology

[‡]Division of Advanced Information Technology and Computer Sciences, Tokyo Univ. of Agriculture and Technology

Abstract The authors constructed a model of the rhythmic sense of human by Support Vector Machine in order to generate a drum track from an input melody track automatically. In addition we generated drum track automatically by Genetic Algorithm using the model. According to the result of questionnaires, automatically-generated drum track is “better” than randomly-generated track and “good” as human-generated track. This paper shows a “good” drum track can be generated without professional training data set because we use only rather simple data set made by amateur.

1 はじめに

自動作曲に関する研究は人工知能分野の研究として近年注目されており、隠れマルコフモデルを用いた旋律の自動生成手法 [1] などが発表されている。また音楽におけるリズムを扱った研究として、アドリブ生成のためのリズム学習手法 [2] などが発表されているが、ドラム譜の生成を目的とした自動作曲手法はいまだ数少ない。そこで本研究では、人間が聴く上で“良い”と感じられるドラム譜の自動生成を試みる。まずはじめにドラム譜の“良さ”を定義するために、人間のリズムに対する評価能力をサポートベクターマシン [4] によりモデル化する。こうして得られたモデルを遺伝的アルゴリズムに取り入れることで、最も良いドラム譜を自動生成することができる。

従来自動作曲手法では、[1] のように既存の楽曲を教師データに用いる場合や、[3] のように楽典や経験に基づき評価関数を設計する場合がある。しかしながら、人間が良いと感じるドラム譜を生成するためには、どのようなドラム譜が良いドラム譜なのかを実際の人間の評価を基に定義する必要があると筆者は考える。これをヒューリスティックではな

い事例ベースの手法で行うためには、評価値の付与されたドラム譜を学習データとして用いる必要があるが、そのようなドラム譜は一般的ではない。そこで本研究では、まずはじめに評価値の付与されたドラム譜を効率よく収集する手法について考える。これは現在普及しているインターネットとPCを用いた新しい取り組みである。

本研究では、音楽に対する知識や経験が無くても、人間は音楽に対して良さの判断を主観的に行うことが可能であると仮定する。本稿で議論する良さとは個人における主観的な評価であるが、我々は異なる個人の間で良さについての共有の感性を有していると考え、それについての考察も行う。このため、本研究で扱うシステムの被験者はすべて、音楽に対する知識や経験の有無を確認していない一般ユーザーである。このような環境であっても、人手で生成したドラム譜に近い良さをもったドラム譜が生成できることを示す。

2 研究にあたっての定義

本研究では楽曲の数値表現の簡単化および計算規模の縮小のため、以下のような定義を行う。

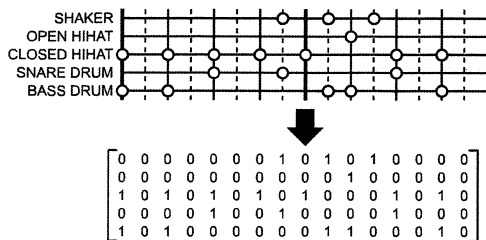


図 1: ドラムパターン

2.1 ドラムパターンとBGM

ドラムパターンとは、四分の四拍子における16分音符を基本単位とした1小節長のドラム譜であるとする。また扱う打楽器の種類は、バスドラム、スネアドラム、クローズドハイハット、オープンハイハット、シェイカーの五種類とする。この定義を用いることで、ドラムパターンは行列を用いて容易に数値表現することが可能となる。図1はあるドラムパターンとそれに対応する行列の例である。

BGMとは、メロディー譜及びベース譜のみで構成される楽曲のことを指し、四分の四拍子で8小節の長さをもつものとする。BGM内の音符は音高と音価の2つの情報をもっているため、ドラムパターンと同様の数値表現するために音高行列と音価行列の2つの行列を用いる。

さらに、ドラムパターンをBGMの小節の数だけ繰り返したものをドラム譜とし、ドラム譜を付与したBGMをミックスと呼称する。

3 リズム判定モデル

本研究では、ドラム譜の良さの定義として人間のリズム感をモデル化する。これがリズム判定モデルである。人間と同様の評価が可能なモデルを得るためには、ミックスに対して人間がつけた評価値のデータを学習データとして用いる必要がある。しかしながら、人間にとって楽曲の良さを絶対的な点数で表現することは容易ではない。またそのような絶対的な評価は基準が曖昧であるため、個人の感覚の違いによって評価が大きく変わってしまうことが考えられる。そこで本研究では、2つのミックスのうちどちらがより良く感じられるかという相対的な評価、すなわち勝敗データに着目する。これは前述の絶対的な評価に比べて判断基準が明確であるため、より信頼性の高い事例データであるといえる。このような勝敗データを基に生成されるリズム判定モデルは、入力された2つのミックスのうちどちらが優

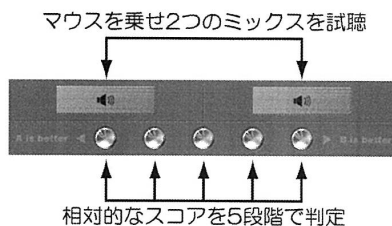


図 2: くちばシステム

れているかを判定する能力をもったものとなる。本研究では、非線形分離が可能な識別手法として注目されているサポートベクターマシン [2] を用いてリズム判定モデルを生成する。

3.1 くちばシステム

前述の勝敗データは、楽譜のように過去に生成された事例データではないため、独自に収集する必要がある。そこで本研究では、インターネットを通じて不特定多数の一般ユーザーから勝敗データを収集するアプリケーションであるくちばシステムを独自に設計した。図2はくちばシステムのインターフェースの一部であり、ユーザーは視覚的な操作により提示された2つのミックスを聴いて、どちらのミックスがより良く聴こえるかという相対的なスコアを次に挙げる5段階で入力する。

- ミックス1が良い
- ミックス1が少し良い
- どちらも言えない
- ミックス2が少し良い
- ミックス2が良い

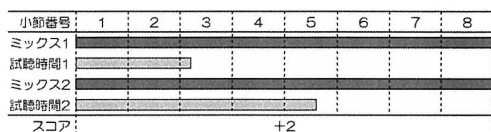
3.1.1 ミックスの生成

ユーザーに提示する2つのミックスは、同じBGMに対して異なるドラムパターンから構成されるドラム譜を付与したものである。本研究では、ある程度リズムとしての体裁をもちつつ多様性のあるドラム譜を提示するために、その構成元となるドラムパターンを以下に述べる方法で生成する。

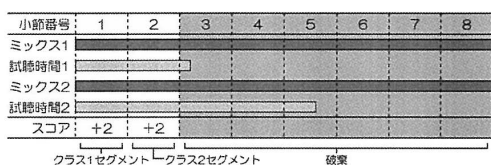
一つのドラムパターンを生成する際には、5つの楽器パート（行列の各行に相当）それぞれについて、ランダムかプリセットかを確率的に決定する。本研究では、確率0.2でランダム、確率0.8でプリセットとなるように設定した。図3は実際に生成されるドラムパターンのイメージである。ただし、ランダムに楽器パートを生成する場合には、密度（あるタイミングにおいてノートが発生する確率）を0から



図 3: KS で提示されるドラムパターン



(a)



(b)

図 4: 勝敗データのクラス分割

1 の範囲であらかじめ決定する。

提示するミックスに用いる BGM は、あらかじめシステムに内蔵された 20 曲の BGMの中からランダムに 1 曲選択する。

今回は本研究の web サイトにてくちばシステムを公開し、オープンな環境にて勝敗データの収集を行った。約 90 日の間に得られた勝敗データのスコアの分布を表 1 に示す。

表 1: 収集した勝敗データのスコア分布

	-2	-1	0	+1	+2
件数	3405	5078	2850	4657	3620

3.2 勝敗データのクラス分割

くちばシステムを用いて収集された勝敗データは、すべてがミックスの先頭から末尾まで聴いたうえで判定されているわけではないため、実際に試聴された部分のみを事例データとして扱うために、次に述べるような勝敗データのクラス分割を行う。

図 4(a) はある勝敗データにおける 2 つのミックス、それぞれのミックスが実際に試聴された時間、及び判定されたスコアを示している。ミックス 1 は 3 小節目の途中まで試聴されており、ミックス 2 は 5 小節目の途中まで試聴されていることがわかる。こ

表 2: クラス分割後のセグメント数

	CL1	CL2	CL3	CL4
件数	13171	6791	4210	2780

	CL5	CL6	CL7	CL8
件数	2053	1742	1574	1074

の場合、実際に試聴され判断されたのはミックスの 2 小節目までと見なすことができ、それ以降のデータは信頼性が低いと考えられるため破棄する。さらに勝敗データにおけるスコアは、各小節それぞれに対する評価の平均値であると仮定し、各小節に対して勝敗データにおけるスコアと同じ値を部分スコアとして付与する。こうすることで、図 4(a) の勝敗データから、図 4(b) のように 1 小節の長さで分割された短い勝敗データが 2 つ得られることとなる。この短い勝敗データのことをセグメントと呼称する。

古典的な楽曲は、動機が 2 小節で構成され、動機が連結されて楽節が構成されるといった、小節単位の楽曲構造をもっている場合が多く、くちばシステムに内蔵されている BGM もこのような形式に沿う楽曲である。このことから、前述のセグメントは、小節ごとに音楽的な意味合いが異なると考えられるため、本研究では勝敗データの中に含まれるセグメントを小節の番号でクラス分割する。図 4 の例では、クラス 1 とクラス 2 のセグメントがそれぞれ 1 つずつ得られることとなる。

このようにしてすべての勝敗データをクラス分割した場合、各クラスごとのセグメントの数は表 2 のとおりになる。これによれば、収集された勝敗データのうち大半はミックスの 1 小節目までを聴いた上でスコア判定されているということがわかる。

3.3 SVM を用いたリズム判定モデル

本研究で生成するリズム判定モデルは、入力されたミックスのうちどちらが良いかを判定するものである。本研究では、このような 2 つの値を識別する手法として、非線形な分離が可能なサポートベクターマシン [4] (以下 SVM) を用いることとする。ただし前節で説明したセグメントは 5 段階のスコアが付与されているため、今回は -1 と -2 のスコアをすべて -1 に、+1 と +2 のスコアをすべて +1 に変換したものを学習データとして用いる。また 0 のスコアについては、ノイズ除去の目的で学習データには用いないこととする。なお、今回 SVM ツールはライブラリが公開されている LIBSVM[5] を用いる。

学習データは、2つのドラムパターンとセグメント内のメロディー譜及びベース譜の情報を特徴ベクトルに変換し用いることとする。特徴ベクトルの詳細を表3に示す。また、セグメントはミックスの小節番号でクラス分割されているため、各クラスごとにSVMモデルを生成する。ここでリズム判定モデルの出力は、すべてのクラスのSVMモデルの出力を平均したものとする。

4 GAに基づくドラムパターン生成

遺伝的アルゴリズム（以下GA）は、自動作曲の研究に応用される場合も多く、本研究の先行研究にあたる[4]においても対話型GAが用いられている。

GAを用いる場合、個体の環境適応度を静的評価する適応度評価関数を設計する必要があるが、本研究では前述のリズム判定モデルを適応度評価関数に置き換えた手法を試みる。

4.1 ドラムパターン生成アルゴリズム

個体集合のうち、環境適応度の高い個体を次世代に残すというのが基本的なGAの考え方であるが、本手法では個体集合から2つの個体を適当に選択し、より環境に適している方の個体を次世代に残すという操作を繰り返す。具体的なアルゴリズムを図5に示す。ここで個体とはドラムパターンを指している。また今回世代における個体数 N は100としている。図中の収束度 S は以下の式で計算する。

$$S = \frac{1}{40} \sum_{i=0}^4 \sum_{j=0}^{15} \left| \frac{M}{2} - G_{ij} \right|$$

$$G_{ij} = \sum_{n=0}^{N-1} T_{nij}$$

T_{nij} : n 番目のドラムパターンの ij 要素

5 評価実験

リズム判定モデルの評価を行うために、リズム判定モデル内部の各SVMモデルについて、セグメントの分類精度を測定した。これはリズム判定モデルの判定がどの程度人間の判定と一致するかを調査するものである。また、節4.1で述べた手法で自動生成されたドラムパターンの良さを評価するために、webアンケートによりユーザーの主観的な評価を収集した。

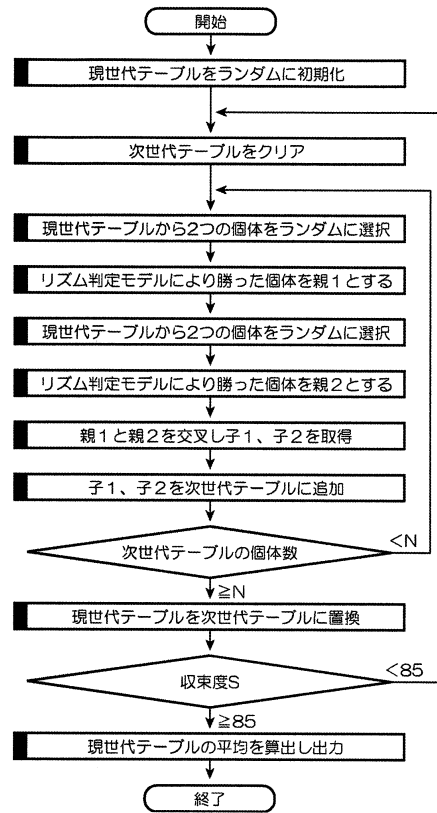


図5: ドラムパターン生成アルゴリズム

5.1 リズム判定モデルの精度

リズム判定モデルの生成に用いた学習データを、再びテストデータとして用いたクローズドテストでの分類精度を表4に示す。すべてのクラスにおいて95%以上の精度が出ていることがわかる。また、勝敗データを学習用とテスト用に分割した5分割交差検定による精度を表5に示す。分割精度の平均値は、すべてのクラスにわたり65%程度の精度で安定していることがわかる。

5.2 アンケートによる評価

本研究で自動生成したドラム譜の良さを測定するために、他の方法で生成したドラム譜との相対的な良さをアンケートにより収集した。比較に用いたドラムパターンは以下の2種類である。

- 人手で生成
- ランダムに生成

アンケートの方法は、生成方法の異なるドラム譜を同じBGMに付与した2つのミックスについて、

表 3: クローズドテスト

	CL1	CL2	CL3	CL4	CL5	CL6	CL7	CL8
精度 (%)	96.8	97.8	98.5	99.1	99.5	99.5	99.5	99.8

表 4: 5 分割交差検定

	CL1	CL2	CL3	CL4	CL5	CL6	CL7	CL8
精度 1 (%)	65.6	62.7	63.9	65.4	64.6	65.5	65.0	65.7
精度 2 (%)	67.5	64.3	65.4	63.4	61.4	64.4	65.4	62.9
精度 3 (%)	66.4	65.4	65.0	62.7	64.9	65.1	67.7	65.1
精度 4 (%)	64.0	64.3	66.7	64.1	63.4	69.0	64.2	60.0
精度 5 (%)	68.2	68.9	66.1	67.5	65.5	65.8	60.7	63.4
平均 (%)	66.3	65.1	65.4	64.6	63.9	65.9	64.6	63.4

表 5: アンケート集計結果

	勝	引	負	合計
自-人	96	41	130	267
自-ラ	193	29	49	271
人-ラ	196	20	37	253

どちらがより良く感じられるかを回答する形式である。提示される 2 つのミックスの生成方法はアンケート被験者には提示しない。アンケートに用いる BGM はくちばシステムに内蔵されているものとは異なる 10 曲を用意した。それぞれの BGM について、自動生成したドラム譜 2 種類、人手で生成したドラム譜 2 種類、ランダムに生成したドラム譜 5 種類を用意し、合計 90 種類のミックスの中から生成手法の異なる 2 つのミックスをランダムに選択し、アンケート被験者に提示する。なお人手によるドラムパターンは、すべて作曲暦 10 年程度のアマチュアコンポーザーにより製作されたものである。

アンケートの各生成手法間における勝敗の集計結果は表 5 の通りである。自動生成したドラム譜は人手で生成したものに対して 36 % 程度の勝率、ランダムに生成したものに対しては 71 % 程度の勝率であった。また人手で生成したドラム譜はランダムに生成したものに対して 77 % 程度の勝率であった。次にアンケート結果から算出した各ミックスのレーティング値¹を図 6 に示す。値の平均値は、自動生成したドラム譜は約 1615、人手で生成したものは約 1712、ランダムに生成したものは約 1368 となっている。

図 7 には実際に生成されたドラムパターンを示す。これによれば、バスドラムは 1 拍目と 2 拍目に

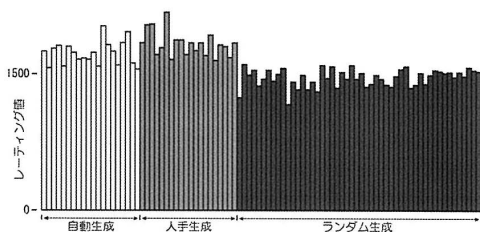


図 6: 各ミックスのレーティング値

鳴らされる場合が多く、スネアドラムとオープンハイハットは 2 拍目と 4 拍目に鳴らされる場合が多い。またシェイカーについては BGM ごとに鳴らされるタイミングが変化しており、密度が他の打楽器に比べて高いことがわかる。

6 考察

クローズドテストの結果において、すべてのクラスにわたり 95%以上の精度が出ていることから、リズム判定モデルは問題なく動作していることがわかる。5 分割交差検定においては、各分割精度の平均値がすべてのクラスにわたり 65%程度で安定しており、2 択問題のベースラインは 50%であることから、人間のリズムに対する評価能力を有意にモデル化することができたと言える。音楽における“良さ”とは定義が非常に難しい問題ではあるが、今回得られた結果から考えられることは、人間のリズムに対する評価能力にはある程度の共通部分が存在し、それを SVM を用いたリズム判定モデルとして数理的に表現することができたということである。

生成手法の異なるドラム譜同士の比較アンケートでは、自動生成されたドラム譜は、ランダムに生成したドラム譜に対して 71%程度の勝率となっており、本研究の手法を用いることでランダムに生成し

¹集計における個の強さを示す値であり、個と個の対戦成績から算出される。

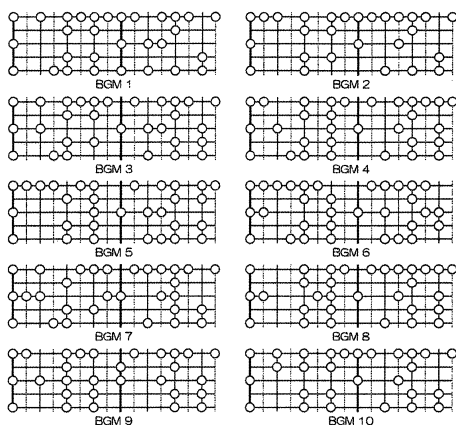


図 7: 自動生成されたドラムパターン

たドラム譜よりも良いと感じられるものが自動生成できたと言える。また、自動生成されたドラム譜と人手で生成したドラム譜の結果を見てみると、自動生成されたものは 36% 程度の勝率となっており、人手で生成したドラム譜の方がより良いものであると言えるが、人手で生成したドラム譜とランダムに生成したドラム譜の結果を見てみると、人手で生成したものは 77% 程度の勝率であることから、本研究の手法で自動生成されたドラム譜は人手で生成したドラム譜に近い良さをもっていると言える。このことは図 6 に示すレーティングの値の各生成手法ごとの平均値からもうかがえる。

図 7 に示す自動生成されたドラムパターンの傾向として、1 拍目に必ずバスドラムが鳴らされていることから、これは良い（聴感上不自然でない）ドラムパターンの条件と考えられる。このことはスネアドラムの 2 拍目と 4 拍目についても同様である。またシェイカーは他の打楽器に比べ密度が高くなっており、これは音質的に目立たない打楽器であるからだと考えられる。スネアドラム等の中心的な打楽器の場合、密度が高いと不快に感じられるという意見もユーザーから回答されている。このような音楽そのものに対する解析も本研究で得られた結果から行うことができる。

7 まとめと展望

入力された楽曲に対して、人間が“良い”と感じるドラム譜を自動生成することを目的に、我々はまずドラム譜の“良さ”の定義として、人間のリズムに対する評価能力を SVM によりモデル化した。これがリズム判定モデルである。リズム判定モデルの

生成に用いた事例データは、2 つのドラム譜の相対的な良さを示す単純なデータであり、それを独自に設計したアプリケーションにより一般ユーザーから収集した。さらにリズム判定モデルを遺伝的アルゴリズムに組み込むことで、入力された楽曲という環境においてもっとも良いと考えられるドラム譜を自動生成した。一般ユーザーを対象としたアンケート調査の結果によれば、自動生成されたドラム譜はランダムに生成したドラム譜よりも良いと感じられ、また人手で生成したドラム譜に近い良さをもっていることがわかった。

今回の用いた勝敗データは、音楽に対する知識や経験を確認していない一般ユーザーから収集した素朴なものであるが、このような事例データであっても有意な良さをもったドラム譜を自動生成できることがわかった。

リズム判定モデルの評価は、人間の評価と 65% 程度一致していることから、人間は異なる個人の間で良さについての共有の感性を有していると考えられ、本研究でその共通部分を数理的にモデル化できた。

創造的な作業を行う上で、“良さ”を定義することは必須である。本研究で試みた人間の評価に基づき良さを定義する手法は単純かつ本質的であり、これは他の創造的な作業にも応用することができる。本研究が人工知能研究の発展に貢献できれば幸いである。

参考文献

- [1] 田村 理遊, 池田 剛, 但馬 康弘, 小谷 善行, “音高と音価の隠れマルコフモデルを用いた自動副旋律生成”, 情報処理学会研究報告, 2007-MUS-69, 2007.
- [2] 薄 浩之, 乾 伸雄, 野瀬 隆, 小谷 善行, “ニューラルネットワークを用いたアドリブ生成のためのリズム学習”, 情報処理学会研究報告, 98-MUS-26, 1998.
- [3] Yaser M.A. Khalifa, Badar K Khan, Jasmin Begovic, Airrion Widdom and Andrew Maxymillian Wheeler, “Evolutionary Music Composer integrating Formal Grammar”, *Genetic and Evolutionary Computation Conference*, 2007.
- [4] V.Vapnik, “The Nature of Statistical Learning Theory”, *Springer*, 1995.
- [5] Chih-Chung Chang and Chih-Jen Lin, “LIB-SVM : a library for support vector machines”, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [6] Tokui N and Iba H, “Music Composition with Interactive Evolutionary Computation”, in *Proceedings of Generative Art*, Milan, Italy, 2000.