THE IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS (JAPANESE EDITION)

電子情報通信学会 論文誌 D 情報・システム

VOL. J104-D NO. 12
DECEMBER 2021

情報・システムソサイエティ

一般社団法人 電子情報通信学会
THE INFORMATION AND SYSTEMS SOCIETY
THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

論　文

キー音を用いたリズムアクションゲームのための譜面自動生成システムの開発*

福永　大輝†　　越智　景子††　　大淵　康成††a)

Development of Automatic Chart Generation System for Rhythm Action Games Using Key-sounds*

Daiki FUKUNAGA†, Keiko OCHI††, and Yasunari OBUCHI††a)

あらまし　「リズムアクションゲーム」とは，楽曲のリズムに合わせた「譜面」と呼ばれる視覚的な記号に沿って操作を行うゲームである．本研究の目的は，キー音を用いたリズムアクションゲームの制作を容易にするために，任意の楽曲から譜面を自動生成するシステムを開発することである．そのために本研究では，楽曲中の音の音響的特徴と再生される時系列の情報から機械学習を行い，譜面を自動生成する手法を提案する．また，譜面作成者の個性の違いを学習に反映させるため，譜面データのクラスタリングを行う．こうした手法により作成したシステムについて，既存データの再現性という観点で評価を行った．更に，本システムにより自動生成した譜面の一例を用いてユーザ評価を行った．

キーワード　リズムアクションゲーム，譜面生成，キー音，機械学習，クラスタリング

1. まえがき

「音楽ゲーム」とは，音楽を主軸に置いたコンピュータゲームのジャンルである．デバイスやプレイヤーに求めるスキルから音楽ゲームを分類する試みがなされているが[1]，楽曲のリズムに合わせた「譜面」と呼ばれる視覚的記号に沿って操作を行い，その正確さによって得点が記録されるものを特に「リズムアクションゲーム」と呼ぶ[2]．これらのゲームでは演奏する記号一つを「ノート」と呼ぶ．譜面はノートの集合である．譜面は音楽の楽譜とは異なり，リズムアクションゲームのゲームデータに相当する．その中には，音楽中のどの時刻にどの音データが再生されるか，またユーザはどの時刻にどの操作（鍵盤やスクラッチ，ダンスパッドなどのコントローラ操作）を求められるかが記述されている．

リズムアクションゲームにおいて，多くの場合プレイヤーが行った操作に応じてサウンドが再生される．プレイヤーが行った操作に応じて再生されるサウンドは，「タップ音」と「キー音」の二つに分類することができる．タップ音とは楽曲とは関係のない効果音であり，本来の楽曲の音とは別に重ねて再生されるものである．また，キー音とは楽曲中の音の一部が切り出されたものである．

キー音方式のゲームでは，楽曲が多数の音オブジェクトの集合として表現される．オブジェクトは，特定の楽器の単一の音若しくは数音からなる短いフレーズに対応しており，それぞれ再生のタイミングが指定される．複数の楽器が同時に鳴っている場面では，再生のタイミングが重なりあうこともある．こうしたオブジェクトのうちの一部がキー音として指定される．主旋律を担当する楽器音がキー音に指定される場合もあれば，副旋律や打楽器音がキー音に指定されることもある．また，全く同じタイミングで再生される複数の音がキー音に指定されることもある．こうした違いは，ゲーム制作者の意図を反映している．

† 東京工科大学大学院バイオ・情報メディア研究科メディアサイエンス専攻，八王子市
Graduate School of Bionics, Computer and Media Science, Tokyo University of Technology, Hachioji-shi, 192–0982 Japan
†† 東京工科大学メディア学部，八王子市
School of Media Science, Tokyo University of Technology, Hachioji-shi, 192–0982 Japan
a) E-mail: obuchiysnr@stf.teu.ac.jp
＊本論文は，システム開発論文である．
DOI:10.14923/transinfj.2020JDP7079

本研究の目的は，任意の楽曲の波形データから，キー音を用いたリズムアクションゲームの譜面を自動生成することである．これまでに，波形データからタップ音方式の譜面を自動生成した研究例 [3], [4] はあるが，プレイヤーの操作によって楽曲自体が変化するキー音方式のゲームで，波形データからの譜面自動生成を試みた例はない．キー音方式の譜面の自動生成を，MIDI データを用いて行った例 [5] や，楽器名のラベルを用いて行った例 [6] はあるが，本研究では，一般的なユーザが自分の好きな曲を使ってゲームをプレイすることができるよう，一切の付加情報を使用せず，波形データのみから譜面の自動生成を試みる．

波形データからキー音方式の譜面を生成するには，以下の三つの過程が必要とされる．

（1） 楽曲から個々の音（オブジェクト）を抽出する
（2） 全オブジェクトをキー音と BGM に分類する
（3） キー音を特定のユーザ操作に対応させる

譜面生成過程（1）においては，曲全体を表す一つの波形データを，多数の細かい波形に分割する．分割された波形は「オブジェクト」と呼ばれる．個々のオブジェクトは，一般的な楽譜の音符に近い単位であるが，必ずしも音符と一致しなくても良く，体感上の演奏の単位と感じられるものであることが望ましい．この過程については，完全な自動化がなされているわけではないが，Be-Music Helper [7] や Mid2BMS [8] など，補助を目的としたツールが多数存在している．また，librosa [9] や Spleeter [10] といった音源分離ツールにより，楽器ごとのトラックが抽出できることも知られている．

楽器音の音源分離は困難なタスクであり，最新の手法を用いても，分離音にひずみが生じてしまうことは避けられない．その結果，分離音に対する後段の分類タスクにおいて，精度が低下することが懸念される．これに対しては，楽器ごとに分かれている学習データをいったん混合し，音源分離アルゴリズムにより分離されたデータを学習データとして用いることで，改善されると期待される．なお，自動生成されたゲームにおいては，分離音の総和がおおよそ原音に一致するような手法を用いれば，プレイヤーが正しくプレイした場合にひずみを感じることはない．ひずみを感じるのはプレイに失敗したときであり，多少不自然な音になったとしても，ゲームとしての不利益はあまり大きくない．

譜面生成過程（2）では，オブジェクトの中からユーザ操作の対象となるキー音を選定する．選定されなかったオブジェクトは「BGM」と呼ばれる．過去に行われたタップ音方式のゲームの自動生成では，これら譜面生成過程（1）（2）の代わりに，ユーザ操作のタイミングを決める処理が行われる．この譜面生成過程（1）（2）の入出力がタップ音方式の場合と大きく異なり，楽曲を構成する個別のオブジェクトごとの音響特徴の関係を譜面生成に活かしていることが，本研究の独自性の大きな部分を占めている．

譜面生成過程（3）では，キー音として選ばれたオブジェクトを，複数存在するユーザ操作の中からどれに対応させるべきかを決定する．この過程は，タップ音方式のゲームにおいても，例えばダンスのステップの向きの決定などで必要となる．本研究では，鍵盤とスクラッチで構成されたゲームコントローラを想定し，各キー音がどの操作に割り当てられるかを決定する．

本研究では，既に譜面生成過程（1）を終えたデータを対象として，譜面生成過程（2）（3）の自動化のシステムを完成させる．ゲーム制作者による手動ゲーム作成では，譜面生成過程（2）では主旋律をキー音として選択すること，譜面生成過程（3）では音高が高い音ほど右に配置するなどの工夫が必要だといわれている [11]．人間がこうした工夫により「面白い」と言われるゲームを作るのに対し，同じような効果が自動化システムにより得られることが目標である．

譜面生成過程（2）（3）を分類問題として定義し直すと，以下の三つのタスクで表すことができる．

（A） キー音・BGM 分類タスク
（B） 鍵盤・スクラッチ分類タスク
（C） 鍵盤配置タスク

これらの分類タスクにより，オブジェクトがどのように分けられていくのかの概要を，図 1 に示す．ひとつひとつのオブジェクトはキー音と BGM に分かれ，更にキー音は鍵盤とスクラッチに分けられる．どのような音がキー音に割り当てられるかに制約はなく，作品によって大きく異なる．図 2 は既存のゲームに含まれる二つの曲において，鍵盤及びスクラッチに割り当てられた代表的な音のスペクトログラムである．楽曲 A においては，メロディーを構成する管楽器（ビン笛）の音が鍵盤に割り当てられる一方，スクラッチには旋律をもたないシンバル音が割り当てられている．それに対して楽曲 B では，シンバルの音が鍵盤に割り当て

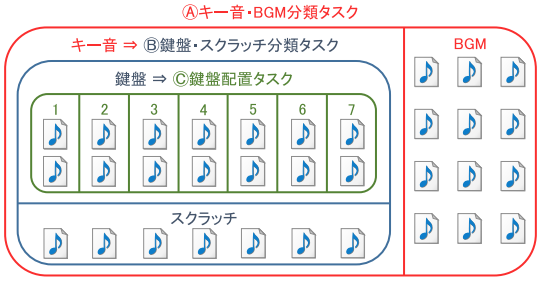


図 1　オブジェクトの分類
Fig. 1　The classification of objects.

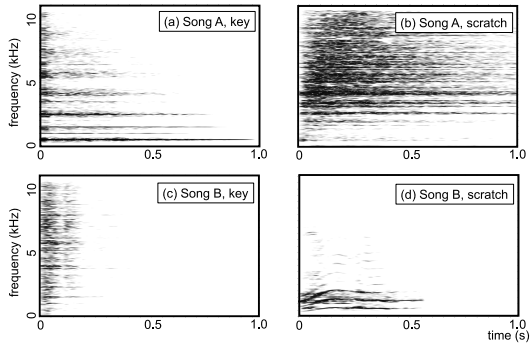


図 2　キー音の例：（上段）楽曲 A の鍵盤 (a) とスクラッチ (b)，（下段）楽曲 B の鍵盤 (c) とスクラッチ (d)
Fig. 2　Examples of key sounds. (top) key (a) and scratch (b) of Song A, (bottom) key (c) kand scratch (d) of Song B.

られる一方，ピッチが揺れるようなボーカルサンプリング音がスクラッチに割り当てられている．

これらのオブジェクトを分類するタスクを，先に述べた三つの過程の中に位置づけると，タスク A は譜面生成過程（2），タスク B，C は譜面生成過程（3）に相当する．具体的には，全てのオブジェクトを対象に，タスク A を行う．次に，キー音と判定されたオブジェクトのみを対象に，タスク B を行う．最後に，鍵盤と判定されたオブジェクトのみを対象に，タスク C を行う．これにより譜面を自動生成するために必要な全ての情報を得ることができる．なお，このような階層的分類モデルの代わりに，七つの鍵盤とスクラッチ及び BGM からなる 9 クラスの単一分類タスクを採用することも可能である．しかし，三つのタスクそれぞれの失敗が意味するところは異なり，ゲーム作成の意図に応じて各タスクの評価基準を調整できることが望ましいと考え，本研究では上述のようなタスク構成を取るものとする．

これまで我々は，オブジェクトの音響的特徴を用いた機械学習をこれらの過程に適用することを提案してきた．タスク A について，分類対象のオブジェクトよりも前に再生されるオブジェクトの情報を用いることで，分類精度が高まることを検証した [12]．本研究の具体的な成果は，学習データに対するクラスタリングの導入 [13] により，文献 [12] に示したタスク A の精度を向上できる可能性を示したことと，この手法をタスク B，C にも応用する仕組みを構築し，キー音方式のゲーム自動生成システムを完成させユーザ評価を通じてリズムアクションゲームの譜面の知見を得たことである．完成したシステムの性能は，自動生成されたゲームが既存のゲームをどの程度再現しているかにより評価した．また，本システムについての知見を得ることを目的として，提案手法により生成したゲームの一例を用いたユーザ評価を行った．

2. 関 連 研 究

本研究で扱ったキー音方式のゲームの自動生成の研究例は少ないが，タップ音方式のゲームでの研究例は幾つか存在する．例えば，Perkasa らは，タップ音を用いたリズムアクションゲームについて，自動的に譜面を生成する手法を提案している [14]．リズムに相当するドラムの音とメロディなどに相当する音高をもつ音を楽曲から検出し，ノートを配置する時刻の推定を行っている．生成した譜面のユーザ評価の結果，難易度の適切さについては高い評価が得られたが，譜面の楽しさについては低い評価であった．

Donahue らは，楽曲の特徴に相当するスペクトル情報とニューラルネットワークを利用して，自動的に譜面を生成する手法を提案している [3]．タップ音のタイミング推定には畳み込みニューラルネットワーク (Convolutional Neural Network: CNN)，ステップ配置には長・短期記憶ユニット (Long Short-Term Memory: LSTM) によるモデルを用いている．

更に，Donahue らの手法をもとに，学習に用いる譜面にクラスタリングを適用する手法が，辻野らによって提案されている [4]．譜面がもつ特徴をもとに学習データのクラスタリングを行い，各クラスタのデータを用いて独立にモデルを生成することで，各クラスタの特徴を反映した譜面が生成されることが確認された．

これらのタップ音方式のゲームの研究例に加え，キー音方式のゲームの研究例も幾つかあるが，いずれも波形データ以外の情報を利用している．例えば，香川らは，MIDI 形式の楽曲をメロディ・ハーモニー・リズム

の 3 層に分解して構造解析することで，譜面自動生成を試みている [5]．また，Lin らは楽曲に付与された楽器別のラベルを利用し，4 層の全結合層をもつモデルによる譜面自動生成を行っている [6]．この研究では，難易度の低い譜面と高い譜面を区別せずに学習データとして使用することによって，分類精度が低下している可能性があることについても示唆されている．

3. 使用データ

PC 用のキー音を用いたリズムアクションゲームである Be-Music Source file: BMS [15] のデータを使用した．BMS のプレイ画面を図 3 に示す．七つの鍵盤と一つの丸いスクラッチのレーンが存在し，全てのノートはいずれかのレーンに配置される．

BMS データは，楽曲中の音が切り出された多数のサウンドオブジェクトのファイルと，同一楽曲に対する 1 個〜数個の譜面ファイルからなる．譜面ファイルには，各オブジェクトの再生タイミングと，各オブジェクトがキー音であるか BGM であるか，更にキー音である場合には対応する鍵盤若しくはスクラッチの情報が書かれている．同一楽曲に対して複数の譜面ファイルが存在するのは，同じ曲であっても難易度の異なるゲームの需要が存在するためである．各譜面に対して，製作者が想定した難易度の数値指標が存在するわけではないが，一般に 1 秒当りのノートの数が多いほど操作が難しくなるといわれており，本研究ではこれを難易度の指標として用いる．

本研究では，熟練したプレイヤーによって BMS データが 399 譜面を含む 85 曲分選定されたパッケージである，GENOSIDE -BMS StarterPackage- [16] を使用した．

BMS では，図 3 に示すような七つの鍵盤と一つのスクラッチをもつ譜面が主流であり，本研究でもその配置を仮定する．ただし，GENOSIDE のデータには，鍵盤の数が 5,10,14 といったものも含まれている(注2)．本研究で扱う，タスク A（キー音・BGM 分類）及びタスク B（鍵盤・スクラッチ分類）の二つのタスクでは，鍵盤の数の違いは関係しないため，これらを含む全てのデータを利用する．一方，タスク C（鍵盤配置）では，学習時と実行時で鍵盤の数が同じである必要があることから，鍵盤数 7 のデータのみを用いる．鍵盤数ごとの譜面の数と，各タスクにおける使用の有無を，表 1 に示す．

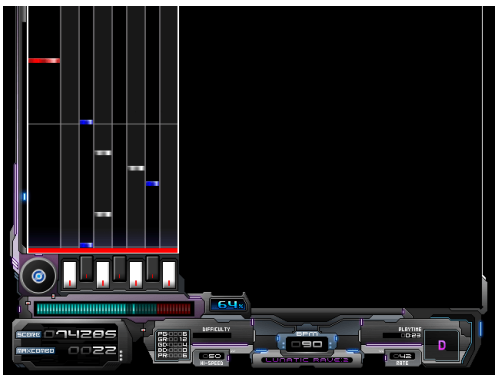


図 3　Lunatic Rave 2(注1) による BMS のプレイ画面
Fig. 3　The play screen of BMS with Lunatic Rave 2.

表 1　データセットに含まれる譜面の種別
Table 1　Chart types included in the dataset.

鍵盤数	譜面数	タスク A	タスク B	タスク C
5	26	✓	✓	
7	304	✓	✓	✓
10	1	✓	✓	
14	68	✓	✓	

4. 提 案 手 法

4.1 想定するシステム

本研究では，任意の楽曲の波形データから，キー音を用いたリズムアクションゲームの譜面を自動生成することを目指している．「まえがき」で説明した譜面を作成するにあたって必要とされる三つの過程と，提案手法に基づいて想定する譜面自動生成システムの対応について図 4 に示す．

譜面作成過程（1）である楽曲からの音の切り出しについては，既存のツールで解決できるとみなし，本研究で直接は扱わない．本研究では，譜面生成過程（1）の出力データを入力として，譜面生成過程（2）に含まれるキー音・BGM 分類タスク，譜面生成過程（3）に含まれる鍵盤・スクラッチ分類タスクと鍵盤配置タスクについて，機械学習を用いた自動化を提案する．

4.2 特 徴 量

譜面を自動生成するにあたり，楽曲がもつ特徴を学習することを目的として，データセットに含まれる全てのサウンドファイルについて，音響特徴量の抽出を行った．

(注1)：http://bmsoffighters.net/lr2/
(注2)：鍵盤数 10,14 は，それぞれ鍵盤数 5,7 のゲームの 2 レーン分を同時にプレイすることを想定している．

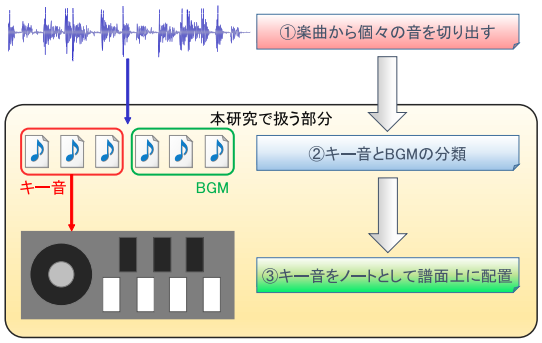


図 4　本研究で想定するシステム
Fig. 4　The system assumed in this work.

音響特徴量の抽出には，openSMILE [17] を使用した．特徴量を抽出するにあたり，全てのサウンドファイルを量子化ビット数 16 bit，サンプリングレート 44.1 kHz，1 チャンネルに変換する前処理を行った．

まず，フレーム長 25 ms，フレーム 10 ms でフレーム化したデータからフレーム特徴量 (Low-Level Descriptor: LLD) を抽出する．次に，各 LLD を全フレーム分集め，それらの統計特徴量（openSMILE では Functional と呼ばれる）を抽出する．これにより，各オブジェクトに対し，LLD の個数と Functional の個数の積で表される数の特徴量が計算される．

使用する音響特徴量は，音楽情報検索の分野でよく使用される音量・ピッチ・スペクトルに関する LLD と，それらの基本的な統計量である最大値・最小値・時間に対する各 LLD の回帰係数などをベースとし，予備実験において高い精度を示した，6 種類の LLD と 10 種類の Functional を使用した．すなわち，各オブジェクトからは 60 個の特徴量が得られる．使用したフレーム特徴量の一覧を表 2 に，統計特徴量の一覧を表 3 に示す．

各オブジェクトについての分類タスクを実行する際には，当該オブジェクト自身の音響特徴量に加えて，時間軸上で近接するオブジェクトの情報も利用することが有効である．本研究では，時間軸に沿った逐次的な分類を行うことを想定して，自分自身以外に過去 P 個のオブジェクトの情報を利用することにする．なお，各オブジェクトはそれぞれ異なる再生時間長をもつが，時間軸上の順番は再生開始時刻を基準として決めた．また，P の値については，基本的には大きいほど良いが，機械学習を行うコンピュータの処理能力の制限もあり，5 と設定した．

表 2　使用したフレーム特徴量 (LLD)
Table 2　The list of used LLDs.

名称	説明
LogEnergy	音量（対数）
HNR	調波成分と非調波成分の比率
F0	基本周波数
SpectralFlux	スペクトルの変化量
SpectralVariance	スペクトルの分散
SpectralSkewness	スペクトルの歪度

表 3　使用した統計特徴量 (Functional)
Table 3　The list of used Functionals.

名称	説明
Max	最大値
Min	最小値
Range	最大値と最小値の差
LinRegC1	線形回帰係数（傾き）
LinRegC2	線形回帰係数（オフセット）
LinRegErrQ	線形回帰 2 乗誤差
Reg-centroid	回帰直線の重心
Skewness	歪度
Kurtosis	尖度
MeanPeakDist	ピーク間の距離の平均

過去のオブジェクトについては，60 個の音響特徴量に加えて，対象となるオブジェクトから何秒離れているかという数値情報と，キー音であるかどうかというバイナリ情報とを利用することができる．したがって，時間軸上で先行する 5 個のオブジェクトからは，各 62 個，合計で 310 個の特徴量が得られる．

最後に，譜面全体から得られる特徴量として，1 秒当りのノートの数を追加する．これは譜面全体の難易度を表し，同一譜面内の全てのオブジェクトが同じ値を共有する．以上により，自分自身から 60 個，先行するオブジェクトから 310 個，譜面全体から 1 個の合計 371 個の特徴量を用いて分類タスクを実行する．

機械学習によるモデル作成及び分類には，機械学習ソフトウェアである Weka [18] を用いた．学習アルゴリズムには，サポートベクターマシン (Support Vector Machine: SVM) の最適化手法の一つである，逐次最小問題最適化法 (Sequential Minimal Optimization: SMO) [19] による多クラス分類を用いた．SVM 自体は 2 クラス分類のアルゴリズムであるが，1 対 1 分類の組合せによる多クラス分類への拡張が Weka に内蔵されている．また，分類対象となる各クラスにデータ数の偏りがある場合には，各クラスが均等に評価されるよう，コストマトリックスを与えて重みづけを行った．

5.　クラスタリング

同じ楽曲，同じ難易度であってもキー音の選択・配

置には多様なパターンがあり，譜面には各タスクの分類基準によってさまざまな個性が存在することが示唆されている [4]．そこで本研究では，機械学習の分類精度向上，また多様な譜面を生成できるようにすることを目的に，この個性をモデリングするために譜面のクラスタリングを行った．

はじめに，データセットに含まれる二つの譜面の類似度を，これらを用いた模擬実験によって数値化する．具体的には，譜面 i を用いてモデルを学習し，譜面 j を用いて評価をした場合の分類精度を A_{ij} とする．各譜面には十分な数のオブジェクトが含まれるため，このような実験が可能である．これにより，二つの譜面の類似度 G_{ij} は以下の式で定義される．

$$G_{ij} = \frac{A_{ij} + A_{ji}}{2} \tag{1}$$

文献 [4] では，譜面から抽出した音響特徴量から PCA で高寄与率成分を取り出し，k-means アルゴリズムでクラスタリングを行っている．しかし，これらの特徴量とゲームの個性の関係は必ずしも明らかではなく，また文献中でも外れ値に対する脆弱性などが述べられている．それに対し，本研究で用いた手法は，分類精度そのものを類似度の指標とすることにより，そうした問題を回避している．

ここで，各譜面の特異度を調べるため，各譜面ごとに他の譜面との類似度の最大値を調べたところ，最小で 71%，最大で 100% となっていた．つまり，どの譜面から見てもある程度似た譜面が存在するということであり，文献 [4] で触れられているような外れ値の除去は，このデータベースに対しては不要であることが分かった．

このようにして定義した類似度に基づき，譜面のクラスタリングを行う．譜面のあらゆるペアに対して類似度が定義できるので，譜面全体は図 5 に示すような完全グラフを成す．ここで各ノードが譜面，各エッジが譜面間の類似度を表す．このグラフに対し，最小カット問題を解くことでクラスタリングが可能となる．

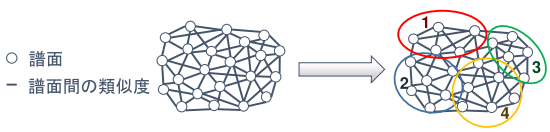


図 5　譜面クラスタリングの概要（実際にはあらゆるノードペア間にエッジが存在する）
Fig. 5　The image of chart clustering. (Actual graph has edges between all node pairs.)

具体的には，まず以下の式でクラスタリングを定義する．s_i が i 番目の譜面を，C_k が k 番目のクラスタを表すとして，譜面 i のクラスタ k への帰属を表す変数 x_{ik} を以下で定義する．

$$x_{ik} = \begin{cases} 1 & (s_i \in C_k) \\ 0 & (\text{else}) \end{cases} \tag{2}$$

この x_{ik} を用いて，任意のクラスタリング結果に対するコスト関数を以下で定義する．

$$C = \sum_{ijk} G_{ij}(1 - x_{ik}x_{jk}) \tag{3}$$

ただし，G_{ij} は譜面 i と譜面 j の類似度である．このとき，コスト関数 C の値を最も小さくするようなクラスタリングを求めるのが，最小カット問題である．類似度の高いペアを結ぶエッジをなるべく切らないようにすることで，結果的に似たような譜面が集まるクラスタを得ることができる．

最小カット問題を解くためのツールは幾つか知られているが，本研究では Metis [20] を用いた．なお，Metis の使用に際しては，類似度 G_{ij} は整数である必要があるため，パーセント表記の小数第一位で四捨五入を行った．

6. 分類の評価

譜面の自動生成を行うにあたり，譜面生成過程（2）（3）における機械学習の有効性を確かめるため，キー音・BGM 分類タスク，鍵盤・スクラッチ分類タスク，鍵盤配置タスクについて機械学習モデルの精度評価をを行った．また，クラスタリングにおけるクラスタの数を変更することで分類精度がどのように変化するか検証を行った．

6.1 実験条件

評価実験は，2 分割のクロスバリデーションで行った．この際，同じ楽曲に対する譜面が学習データとテストデータの両方に含まれることがないようにした．キー音・BGM 分類タスクと鍵盤・スクラッチ分類タスクでは [16] に含まれる全 399 譜面，鍵盤配置タスクでは鍵盤数 7 の 304 譜面を使用した．それぞれのタスクで用いたオブジェクトの数を表 4 に示す．

次に，学習データに対してクラスタ数 1 から 10 ま

表 4　実験に用いたオブジェクトの数
Table 4　Number of objects used in each experiment.

	キー音・BGM 分類タスク	鍵盤・スクラッチ分類タスク	鍵盤配置タスク
1 鍵			62,230
2 鍵			36,787
3 鍵			48,547
4 鍵	431,359	406,650	44,065
5 鍵			48,022
6 鍵			40,867
7 鍵			38,846
スクラッチ		23,576	
BGM	913,670		
合計	1,345,029	430,226	319,364

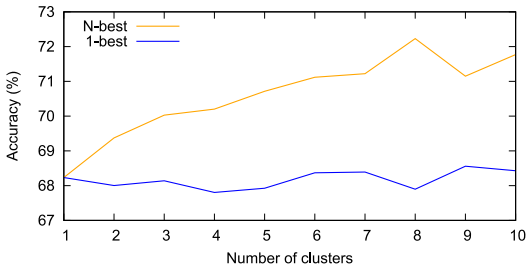


図 6　キー音・BGM 分類タスクの結果
Fig. 6　The result of key-sound / BGM classification task.

でのクラスタリングを行った．クラスタ数 1 がクラスタリングを行わないことに相当する．こうして得られた各クラスタのデータを用いて SVM によるモデル作成を行った．個々のクラスタデータからのモデル作成の詳細については付録 **1.** で述べる．

モデル化に際しては，**4.2** で述べたように，分類対象となるオブジェクトに対して時間的に先行する五つのオブジェクトから得られる特徴量を用いた．したがって，楽曲の開始から五つめまでのオブジェクトからは特徴抽出ができないため，これらは学習及び評価の対象から除外した．

6.2　評価基準

SVM の適用により，テストデータに含まれる各オブジェクトに対し，クラスタの数だけ分類結果が得られる．これらに対して 2 種類の評価を行った．

第 1 の評価は，譜面単位でただ一つのクラスタを自動選択するもので，これにより各オブジェクトに対してただ一つの分類結果が確定する．これを「1-best 評価」と呼ぶ．分類結果の選択基準には，Weka による分類の信頼度を用いる．この評価は，「楽曲が与えられた場合，最適な譜面化というものがただ一つ存在する」という考え方に基づくものである．

第 2 の評価は，各クラスタによって得られた分類結果をそのまま保持し，クラスタ数だけの譜面を作成するもので，「N-best 評価」と呼ぶ．このとき，分類精度の計算にあたっては，元のデータに付与されたラベルとの一致率が最も高い譜面を採用する．この評価は，「楽曲が与えられた場合，作成者の個性により N 種類の譜面が得られるのが自然であり，元データに付与されたラベルはその中の一つに過ぎない」という考え方に基づくものである．そのため，クラスタ数だけの譜面を生成した中に，一つでも元の譜面との一致度が高いものがあれば良いとするわけである．なお，本論文の設定においては，N-best 評価を導入してもチャンスレートの向上は 1.2 ポイント程度に留まる（付録 **2.**）．

なお，分類精度の計算にあたっては，各ラベルのデータ数に大きな偏りがあるため，各タスクにおけるラベルごとの正解率の算術平均を分類精度とした．

6.3　キー音・BGM 分類タスク

キー音・BGM 分類タスクに譜面のクラスタリングを適用した結果を図 6 に示す．横軸がクラスタ数，縦軸が分類精度を表す．2 クラスの分類であるため，チャンスレートは 50% である．

クラスタリングを行わない場合，68.23% の分類精度を得ることができた．クラスタ数を増加させることで，N-best 評価の分類精度が向上することが分かった．クラスタ数 8 のときが最も高く，72.23% となった．一方で，1-best 評価の分類精度には大きな改善は見られなかった．

6.4　鍵盤・スクラッチ分類タスク

鍵盤・スクラッチ分類タスクの結果を図 7 に示す．2 クラスの分類であるため，チャンスレートは 50% である．

クラスタ数が 1 の場合の分類精度は 55.30% となった．1-best 評価では，クラスタ数 3 と 7 でわずかに上昇したが，目立った改善は見られなかった．N-best 評価では，クラスタ数を増加させることによって大きな分類精度の改善が見られた．クラスタ数 9 のときが最も高く，72.30% となった．

6.5　鍵盤配置タスク

鍵盤配置タスクの結果を図 8 に示す．7 クラスの分類であるため，チャンスレートは 14.29% である．

他のタスクとは異なり，クラスタ数が増加するほど 1-best 評価の値は低下し，N-best 評価での精度向上も見られなかった．クラスタリングを行わない場合の分

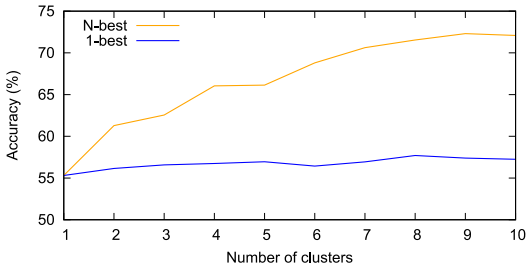


図 7　鍵盤・スクラッチ分類タスクの結果
Fig. 7　The result of keys / scratch classification task.

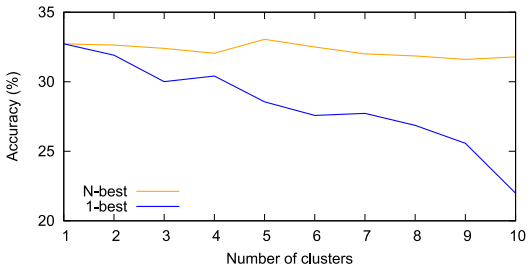


図 8　鍵盤配置タスクの結果
Fig. 8　The result of key assignment task.

類精度も，32.72% と低い値となった．クラスタ数 5 のときが最も高く，33.05% となったが，ごくわずかの改善でしかない．

ここで，異なる観点からの評価を試みる．鍵盤配置については，音高の高低に基づいて行われることが多い [11] ことなどを念頭に，ある鍵盤ノートの位置が，直前の鍵盤ノートの右・左・若しくは同じのいずれであるかで正解判定を行った．結果を図 9 に示す．3 クラスの分類となるので，チャンスレートは 33.33% である．

この場合でも，クラスタ数を増加させることによる分類精度の向上は見られなかった．しかし，クラスタ数 1 の場合の分類精度が 61.40% と，相対的な鍵盤の位置の方向についてはある程度適切に推定ができていることが分かった．

6.6 考　　察

キー音・BGM 分類タスクと鍵盤・スクラッチ分類タスクについては，クラスタ数の増加に伴って N-best 評価の精度が向上していることから，クラスタリングの有効性が確認された．一方，鍵盤配置タスクについては，クラスタリングが有効であるとはいえない結果であったが，相対的な鍵盤の移動方向の推定については，3 クラス分類で 60% 以上の正解率が得られた．

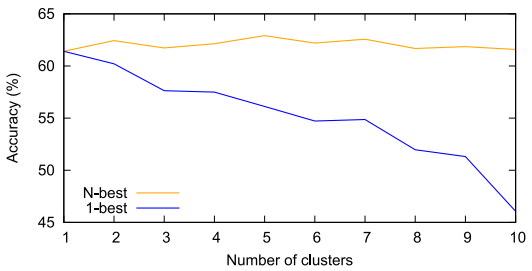


図 9　移動方向に基づく鍵盤配置タスクの結果
Fig. 9　The result of key assignment task based on direction.

このことから，キー音・BGM 分類タスクと鍵盤・スクラッチ分類タスクの分類基準には譜面がもつ個性によって大きな基準の差異が存在するが，鍵盤配置タスクにおける分類基準はどのような譜面であってもおおよそ共通しているといえる．そのため，鍵盤配置タスクでは，クラスタリングを行わずに十分な学習データ量を確保すべきであると考えられる．ただし，クラスタリングに際して，1-best 評価の精度には向上は見られなかった．これは，データセットの作成者の個性を再現するために，分類の信頼度以外の基準が必要なことを示唆している．複数のクラスタが生成した譜面の中から，ユーザが好むであろう譜面を選択しやすくするためには，個々のクラスタがもつ特徴を言語化することも重要であろう．

7. ユーザ評価

本研究で開発した譜面自動作成システムの定量的な評価は，譜面生成過程（2）（3）の各タスクに対して行うことができた．一方，既存の譜面との一致度に留まらず，それ以外の面での主観的な評価を行うことは，今後のシステムの改善において有効であると思われる．ここでは，本研究のシステムによって生成した譜面を実際のユーザに体験してもらった結果について紹介する．

7.1 実 験 条 件

既存の低難易度の譜面が存在する楽曲の中から，BMS Starter Pack [21] というパッケージの 2008 年版に含まれる「Stand by me」を使用し，3 種類の譜面を用意した．

第 1 の譜面は，譜面作成者が手動で作成した譜面で，パッケージに含まれるものをそのまま使用した．これを「人手による譜面」と呼ぶ．第 2 の譜面は，キー音・BGM 分類タスク，鍵盤・スクラッチ分類タスク及び

鍵盤配置タスクを，全て完全なランダム分類で生成したものである．これを「ランダム配置の譜面」と呼ぶ．第 3 の譜面は，本研究のシステムにより生成した譜面で，これを「提案手法による譜面」と呼ぶ．提案手法による譜面の生成にあたっては，クラスタ数を 10 とし，無作為に抽出した一つのクラスタを使用した．入力特徴量に含める「1 秒当りのノートの密度」は人手による譜面に合わせて 1 個としたが，実際にキー音・BGM 推定を行った結果，113 個のノートが得られ，比率にすると 0.96 個/秒となった．なお，入力特徴量の中で，時間的に先行するオブジェクトがキー音であるかどうかについては，推定結果を後続のオブジェクトの推定の入力として用い，再帰的な推定を行っている．また，各楽曲の先頭から五つめまでのオブジェクトについては，先行するオブジェクトを特徴量として使用しないモデルを別途用意し，当該オブジェクトから得られる特徴量のみを用いて推定を行った．なお，付録 **A.1** に記載したクラスタデータの分割・多数決の方式は，再帰的な推定との組合せ方が煩雑になる(注3)ため，ここでは無作為に抽出した一つのデータセットのみを用いる方式により簡略化した．

ランダム配置の譜面の生成にあたっては，1 秒当り約 1 個のノートが得られるよう確率を調整した結果，提案手法による譜面と同じ 113 個のノートが得られた．また，ランダム配置に際しては，7 個の鍵盤がそれぞれ 13.5%，スクラッチが 5.5% の確率で配置を行った．

ユーザ評価の被験者は 30 人（10〜30 代，うち男性 25 人，女性 5 人）の大学生・大学院生である．三つのゲームをランダムな順番でプレイしてもらった後，「譜面の楽しさ」「譜面の難しさ」「音楽と合っているか」の 3 項目で評価してもらった．「音楽と合っているか」という質問については，どのような観点で合っているかを判断するかに回答者の主観が関与し得るが，今回の設問では特段の注釈は設けていない．各項目の選択肢は 5 段階である．なお，本実験では，十分な数のデータによる定量評価を行うことではなく，限られた数のデータから定性的な知見を得ることを目的としていることから，ゲーム間の差が顕著に現れるよう，三つのゲームで同じ選択肢を選ばないように依頼した．また，被験者は各ゲームがどの手法で作られたかを知らされていない．

(注3)：どのデータセットを用いたかによって，時間的に後に続くオブジェクトの推定に用いる入力特徴量が異なってしまう．

7.2 結果と考察

回答の集計にあたっては，被験者の自己申告に基づき，被験者を二つのグループに分割した．リズムアクションゲームのプレイ頻度に対する質問に対し，「経験なし」「数回だけ」と回答した 11 名を低頻度グループ，「たまに遊ぶ」「よく遊ぶ」と回答した 19 名を高頻度グループとする．また，「キー音を知っているか」という質問と，ゲームについての自由記述欄を設けた．なお，以下の集計で，「譜面の楽しさ」の数値が高いほど楽しいことを，「譜面の難しさ」の数値が高いほど難しいことを，「音楽と合っているか」の数値が高いほど合っていることを表している．

ユーザ評価の結果を図 10 に示す．まず，「譜面の楽しさ」については，ランダム配置の譜面の評価が最も低く，人手による譜面と提案手法による譜面の評価が同程度となった．また，低頻度グループでは人手による譜面の評価が最も高かったが，高頻度グループでは提案手法による譜面の評価が高かった．

プレイ頻度によって譜面の評価が分かれたことと関連して，自由記述には，提案手法による譜面について，複雑な配置で歯ごたえがあり楽しいという意見が見られた．一方で人手による譜面については，叩いている音が分かりやすく初心者向けであるという意見が見られた．

次に，「譜面の難しさ」については，いずれのグループでも，人手による譜面が最も易しく，ランダム配置の譜面と提案手法による譜面がほぼ同程度に難しいという結果であった．

参考までに，統計分析ソフト R [22] を用いてアンケート結果の検定を行った．「キー音を知っているか」と「譜面の種類」の 2 要因に着目して，譜面の難しさについての反復測定二元配置分散分析を行った結果，譜面の種類に有意な主効果が認められた ($p < 0.05$)．更に，譜面の難しさについて，対応のある t 検定により多重比較を行った結果（Bonferroni の補正を使用），人手による譜面とランダム配置の譜面，人手による譜面と提案手法による譜面の間において，有意な差が認められた ($p < 0.05$)．あくまでも本実験に用いた譜面に限定していえることだが，他の譜面と比べて，人手による譜面は難易度が低いことが確認できた．

このように，ノートの密度に大きな差がないのにもかかわらず難易度に明らかな差が生まれた理由として，人手による譜面では譜面全体におおよそ均等にノートが存在するが，それ以外の譜面ではノートの分布にば

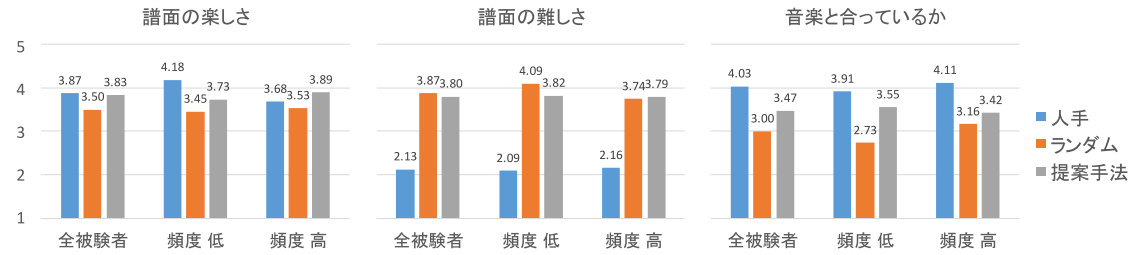


図 10　ユーザ評価の結果
Fig. 10　The result of user evaluation.

らつきが存在することが挙げられる．図 11 は，ユーザ評価に用いた譜面の第 24 小節から第 27 小節に現れるキー音を，対応するスクラッチ (s) 若しくは鍵盤番号の欄に示したものだが，人手の譜面（左）ではサックスによる主旋律（灰色）のみがキー音になっているのに対し，提案手法（右）では，ベースギター（白）やバイオリン（黒）の伴奏音もキー音になっている．また，第 24 小節の 4 拍目や第 26 小節の 4 拍目など，二つのキー音が同時に設定されている例もある．このように，単位時間当りのノート数を難易度の指標として与えても，それ以外の部分で難易度の異なる譜面が生成され，それが上級者にとっては面白さと感じられることもあり得る．実際，自由記述の回答でも，スクラッチの連続や，同時に複数の鍵盤にノートが存在する箇所が難しいという意見があった．こうした現象は新鮮な面白さともいえるが，一方で難易度設定の不安定さと考えることもできるため，学習データに対して正しい難易度のラベルを付与することが，今後重要になると考えられる．

最後に，譜面と音楽が合っているかについては，人手による譜面が最も評価が高く，ランダム配置の譜面が最も評価が低かった．提案手法による譜面の評価はランダム配置の譜面より高く，人手による譜面には及ばないものの，ある程度は音楽に合ったキー音選択ができていることが示唆される．

プレイ頻度別に見ると，高頻度グループでは，提案手法による譜面とランダム配置の譜面の評価の差が小さかった．リズムアクションゲームをよくプレイするユーザほど，人手による譜面を十分に再現できていな譜面に対しての違和感が大きいのではないかと推測される．

また自由記述には，難易度が高いことでゲームクリアに失敗しても，キー音と音楽が合っていると楽しいという意見が見られた．このことから，譜面と音楽が

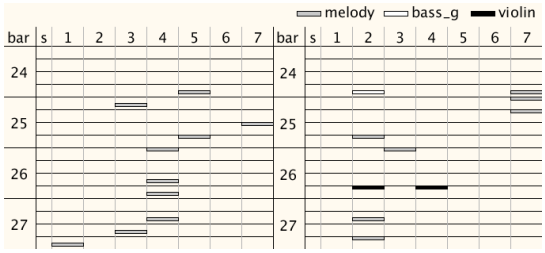


図 11　譜面の例：（左）人手（右）提案手法
Fig. 11　Chart examples: (L) Manual (R) Proposed

合っているかは譜面の楽しさに大きな影響を及ぼしていることが予想される．実際に，全ての被験者の回答における譜面の楽しさと譜面と音楽が合っているかの項目について，人手による譜面で $r = 0.45$ ($p < 0.01$)，ランダム配置の譜面で $r = 0.40$ ($p < 0.05$)，提案手法による譜面で $r = 0.48$ ($p < 0.01$) となり，有意な正の相関が見られた．

8. むすび

キー音を用いたリズムアクションゲームを対象として，楽曲中のオブジェクトから得られる音響特徴量と，譜面から得られる時系列情報を用いて機械学習を行うことで，譜面の自動生成を行うことができた．

譜面のクラスタリングを行うことで，キー音・BGM 分類タスクと鍵盤・スクラッチ分類タスクの N-best 分類精度を向上させることができた．一方で鍵盤配置タスクについては，クラスタリングが有効であるとはいえない結果であったが，相対的な位置については 60% 以上の精度で推定が行えた

これに対し，1-best 分類精度はクラスタリングによっても向上せず，適切なクラスタの選択基準について，更なる検討が必要だと思われる．その際，最適なクラスタを一意に決めるのではなく，ユーザの嗜好にあっ

たクラスタを選べることが重要であり，そのためにどのような情報を提供できるかが，今後の課題である．

楽曲の時間的な構造を譜面に反映させるために，推定対象のオブジェクトに対して時間的に先行するオブジェクトの情報を用いたが，その時間的なスコープの大きさは，扱うツールの処理速度など，実装上の制限に基づいて選ばざるを得なかった．今後更に推定精度を高めるためには，こうした制約を満たしつつ，より大きな時間的スコープを扱えるよう，特徴量選択の仕組みを工夫していくことが求められる．

我々が扱っているデータセットには十分な量のデータが含まれることから，深層学習モデルの適用を行うことも有効であると考えられる．Donahue らの手法を応用し，時系列の学習に適した LSTM などのモデルを導入することで，分類精度を高め，高品質な譜面を生成することが可能であると考えられる．また，本論文の成果を音源分離と組み合わせるため，ひずみを伴うデータを使った学習について更なる検討を行うことも，今後の重要な課題である．

最後に，提案するシステムによって生成した譜面をゲーム化し，ユーザ評価を行った．限られた範囲の評価ではあるが，提案手法による譜面が，ランダム配置の譜面に比べて楽しいものとなっているという結果が得られた．また，提案手法による譜面が，人手による譜面と同じノート密度でありながら，異なる難易度に感じられていることから，難易度をより正確に数値化する指標の必要性が明らかになった．難易度の適切な設定は，初心者・熟練者双方のプレイヤーを満足させるために重要な項目であり，今後更なる検討が求められる．

文　献

[1] 橋本祐輔，橋田光代，片寄晴弘，“音楽音響信号を対象とした指揮演奏システムの開発，”情処学研報，第 2009-EC-12 巻，pp.43–50, 2009.

[2] K. Collins, Game Sound: An Introduction to the History, theory, and practice of video game music and sound design, pp.74–75, The MIT Press, Cambridge, United States of America, 2008.

[3] C. Donahue, Z.C. Lipton, and J. McAuley, “Dance Dance Convolution,” Proc. 34th Int. Conf. Mach. Learn., vol.70, pp.1039–1048, 2017.

[4] 辻野雄大，山西良典，山下洋一，井本桂右，“ダンスゲーム譜面の特性分析とクラスタリングに基づく特徴的な譜面の自動生成，”エンタテインメントコンピューティングシンポジウム 2019 論文集，pp.96–103, 2019.

[5] 香川俊宗，手塚宏史，稲葉真理，“音楽の重要な構成要素の抽出の提案—音楽ゲーム用譜面自動生成のために—，”エンタテインメントコンピューティングシンポジウム 2015 論文集，pp.326–333, 2015.

[6] Z. Lin, M. Riedl, and K. Xiao, “GenerationMania: Learning to semantically choreograph,” Proc. 15th AAAI Conf. Artificial Intelligence and Interactive Digital Entertainment, vol.15, pp.51–57, 2019.

[7] exclusion, “Be-Music Helper (beta 4’)”. https://excln.github.io/bmhelper.html (accessed 2021-6-26)

[8] yuinore, “Mid2BMS BMS Improved Development Environment”. http://mid2bms.web.fc2.com/ (accessed 2021-6-26)

[9] B. McFee, C. Raffel, D. Liang, D.P.W. Ellis, M. McVicar, E. BattenBerg, and O. Nieto, “librosa: Audio and music signal analysis in Python,” Proc. 14th Python in Science Conference, pp.18–24, 2015.

[10] L. Prétet, R. Hennequin, J. Royo-Letelier, and A. Vaglio, “Singing Voice Separation: A Study on training data,” Proc. Int. Conf. Acoustics, Speech Signal Process., pp.506–510, 2019.

[11] Kz, “Obj Tech Lovers | Guidance chapter2”. https://nekokan.dyndns.info/~otlovers/guidance/guidance_2.html (accessed 2021-6-26)

[12] 福永大輝，越智景子，大淵康成，“リズムアクションゲームにおけるキー音の自動推定，”芸術科学会論文誌，vol.18, no.1, pp.10–18, 2019.

[13] D. Fukunaga, K. Ochi, and Y. Obuchi, “Training data clustering for key-sound estimation in rhythm action games,” NICOGRAPH International 2019, pp.70–73, 2019.

[14] D.B. Perkasa and N.U. Maulidevi, “Beatmap generator for osu game using machine learning approach,” 2015 Int. Conf. Electrical Engineering and Informatics, pp.77–81, 2015.

[15] U. Yane, “BM98Data_format_specification”. http://bm98.yaneu.com/bm98/bmsformat.html (accessed 2021-6-26)

[16] lobsak, “GENOSIDE -BMS StarterPackage-”. http://nekokan.dyndns.info/~lobsak/genoside/ (accessed 2021-6-26)

[17] F. Eyben, F. Weninger, F. Groß, and B. Schuller, “Recent developments in openSMILE, the Munich open-source multimedia feature extractor,” Proc. 21st ACM Int. Conf. Multimedia, pp.835–838, 2013.

[18] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, “The WEKA data mining software: An update,” SIGKDD Explorations, vol.11, no.1, pp.10–18, 2009.

[19] J.C. Platt, “Sequential minimal optimization: A fast algorithm for training support vector machines,” Technical Report MSR-TR-98-14, Microsoft Research, 1998.

[20] G. Karypis and V. Kumar, “A fast and high quality multilevel scheme for partitioning irregular graphs,” SIAM J. Scientific Computing, vol.20, no.1, pp.359–392, 1998.

[21] Yamajet, “BMS Starter Pack 2009”. http://www.yamajet.com/bmssp/index.html (accessed 2021-6-26)

[22] 舟尾暢男，The R Tips 第 3 版：データ解析環境 R の基本技・グラフィックス活用集，オーム社，2016.

付　録

1. Weka の高速化のためのデータ再分割

学習データをクラスタリングした後に SVM によるモデル化を行うが，データサイズが大きくなると Weka

の処理時間が著しく長くなるため，データの再分割を行った．各クラスタのデータをランダムに並べ替えた後，キー音・BGM 分類タスクでは 20 分割，鍵盤・スクラッチ分類タスク及び鍵盤配置タスクでは 10 分割した後に Weka を適用した．そのため，各クラスタに対して複数のモデルが存在することになるが，分類タスク実行時には，それらのモデルによる分類結果の重み付き多数決により，当該クラスタの分類結果を決定した．ここでいう重み付き多数決とは，各 SVM モデルが出力する分類結果に付与された信頼度を合算し，合計値が最も大きくなった分類結果を採用するというものである．分類の信頼度は，N クラス分類問題に対して 0 から 2/N の範囲になるよう Weka により正規化され，最大値を取った選択肢が出力される．

2. N-best 実行時のチャンスレートの計算

一つの譜面に含まれるオブジェクトの数を K とし，各オブジェクトの推定が正解確率 p の独立事象だとすると，譜面全体での正解数は二項分布となるが，K が十分に大きければ正規分布で近似できる．正解数を K で割って正解率とすると，その確率密度 $f(x)$ は，平均 $\mu = p$，分散 $\sigma^2 = p(1-p)/K$ の正規分布となる．このとき正解率の累積確率密度関数は，

$$F(x) = \frac{1}{2}(1 + \mathrm{erf}(\frac{x-\mu}{\sqrt{2}\sigma})) \qquad \text{(A·1)}$$

と表される．これを N 回試行して最大値を取ると，その累積確率密度関数は以下で得られる．

$$F_{max}(x) = F(x)^n \qquad \text{(A·2)}$$

確率密度関数はこれを x で微分すれば得られる．

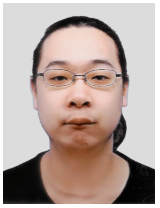
$$f_{max}(x) = nF(x)^{n-1} f(x) \qquad \text{(A·3)}$$

となる．このときの x の期待値は，

$$E[f_{max}(x)] = \int_0^1 xnF(x)^{n-1} f(x)dx \qquad \text{(A·4)}$$

で与えられる．この積分を平易な式で表すことはできないが，数値計算により値を求めることはできる．本論文のキー音・BGM 分類タスクでは，399 の譜面に 1,345,029 個のオブジェクトが含まれているので，1 譜面当りのオブジェクト数 3371 を K の値とし，2 クラス分類のチャンスレートを 0.5，N を 8 として計算すると，期待値は 51.22% となる．また，クラスタリングを行わない場合の正解率 68.23% を p に代入した場合，期待値は 69.37% となる．なお，後者の場合に N-best の正解率が 72.23% 以上になる確率は，式 (A·2) を使って求めることができ，0.00024% となる．

（2020 年 12 月 24 日受付，2021 年 7 月 21 日再受付，
8 月 27 日早期公開）



福永　大輝

2018 東京工科大学大学メディア学部メディア学科卒．2020 東京工科大学大学院バイオ・情報メディア研究科メディアサイエンス専攻修士課程了．メディアコンテンツへの音楽・音響情報処理，音声言語処理分野の応用に興味をもつ．リズムアクションゲームの譜面自動生成に関する研究に従事．

越智　景子　（正員）

2011 東京大学情報理工学系研究科電子情報学専攻博士課程了．国立障害者リハビリテーションセンター研究所流動研究員 (2011～2014)．同客員研究員 (2014～2016)．国立情報学研究所特任研究員 (2015～2016)．同特任助教 (2016～2017)．2017 より東京工科大学メディア学部助教．博士（情報理工学）．音声合成，韻律，音声分析，音声インターフェース，言語訓練の研究に従事．

大淵　康成　（正員）

1990 東京大学大学院理学系研究科物理学専攻修士課程了．1992 同博士課程中退．1992 より 2015 まで（株）日立製作所中央研究所及び基礎研究所勤務．その間，Carnegie Mellon University 客員研究員 (2002～2003)，早稲田大学客員研究員 (2005～2010)，クラリオン（株）(2013～2015)．2015 より東京工科大学メディア学部教授．博士（情報理工学）．