［招待論文：研究論文］

AI は創造性を持ちうるか
生成的敵対ネットワークを拡張したリズム生成モデルを実例に

Can AI Be Creative?
A Proposal of an Extended Framework of Generative Adversarial Networks

徳井 直生
慶應義塾大学大学院政策・メディア研究科准教授
Nao Tokui
Associate Professor, Graduate School of Media and Governance, Keio University
Correspondence to: tokui@sfc.keio.ac.jp

Abstract:　人工知能（AI）の社会実装が進む中で、アートやデザインといった表現領域での AI の活用が模索されている。特に生成的敵対ネットワーク（GAN）に代表される生成モデルを用いることで、まるで人が作ったかのような「それらしい」画像や音楽が AI によって生成できることが、すでに多くの研究・作品例で示されている。一方で、これらの生成モデルはあくまでも学習データに含まれる統計的なパターンを学習し再生産したものともいえ、その表現としての新規性、独創性に疑問を投げかけることも可能だ。

本稿では、こうした現状を考察するとともに、GAN のフレームワークを拡張することで新しい表現、特に音楽表現を創出する手法を提案する。これらを通して、AI が単なる人の創作物の模倣ではない表現の創出に寄与する未来について考察する。

As the social implementation of Artificial Intelligence (AI) advances, the use of AI in the creative domains of art and design is being explored. In particular, many studies and examples of works have already shown that AI can generate “realistic” images and music as if they were created by humans by using Generative Adversarial Networks (GAN) and other generative models. On the other hand, one can argue that what generative models do is simply a reproduction of statistical patterns learned from training data and question their novelty and originality as expressions.

In this paper, we examine the current state of AI and creativity and propose a method for creating novel expressions, especially musical expressions, by extending the GAN framework. Through these, we consider the future in which AI will contribute to creating expressions that are not mere imitations of human creations.

Keywords: AI、人工知能、深層学習、Deep Learning、創造性、音楽
artificial intelligence, deep learning, creativity, music, music generation

1　はじめに― AI、機械学習、深層学習

深層学習（Deep Learning）技術[1]の発展によって、人工知能（Artificial Intelligence、AI）の社会実装が進んでいる。画像認識のようなAIが得意とする単純な認知作業のみならず、一般にAIによって置換されにくい職種として挙げられることの多い、デザイナーや作曲家といった創造性を要する職業領域でもAIを活用しようとする動きが始まっている[1)]。

本稿ではこうした現状に鑑み、音楽生成に関する筆者の研究を例に取り上げながら、AIと創造性について考察する。

まず前提として人工知能、AIとは何かを改めて定義しておくことにしよう。AIの定義は研究者によって異なるが[2]、本稿ではシンプルに「人間の知能を人工物、特にコンピュータによって模倣しようとする試み」とする。AIがある知的なタスクを完璧にこなせるようになると、人はそのタスクを知的なタスクだとみなさなくなる。こうした現象は、「AIのジレンマ」あるいは「AIエフェクト」と呼ばれる。そうしたニュアンスを含む定義として「試み」としている。

また、AIと混同しやすい言葉として機械学習（Machine Learning）があるが、機械学習はAIのサブカテゴリーと言える。機械学習は入力データとそれに対応する答えの組み合わせから、データに内在するルールを導くことを目的とする。一般的なコンピュータのプログラミングが人の定めたルール（アルゴリズム）をもとにデータから答えを導くことを目的とするのと比較するとわかりやすい。

AIには、特に初期のエキスパートシステムに代表されるような人が予め定めたルールの集合によって処理を行うルールベースの考え方も含まれるため、全てのAIが機械学習のカテゴリーに当てはまるとは限らない。

また機械学習における教師あり学習（supervised learning）と教師なし学習（unsupervised learning）の区分も本論において重要である。教師あり学習では、あらかじめ入力に対する答えにあたるデータを合わせて用意し、入力に対して正しい答えを出力するように学習を行うのに対して、教師なし学習の場合は答えにあたるデータなしで、入力データの集合から何らかの有益なルールを導く。教師なし学習の例としては、データを複数のグループ（クラスタ）に

分類するクラスタリングなどが挙げられる[2)]。

2012年前後からその研究が盛り上がり、昨今のAIブームを支える原動力となっている深層学習は、機械学習の一種である。脳の神経細胞の機構に緩やかに基づき、数学的にモデル化した人工ニューラルネットワーク（Artificial Neural Network）の考え方が基盤となっている[1]。ニューラルネットワークの研究自体は半世紀に及ぶ歴史があるが、昨今のインターネットの普及による学習データの爆発的な増大、GPUなどの計算資源が充実したこと、効率的な学習アルゴリズムの発案などが重なり、現在の深層学習の大幅な発展につながっている。

なお、本稿では、特に断らない限りにおいて、単にAIとして言及する場合、深層学習モデルを念頭に置いていること、本来であればAIシステム、AIモデルと書くべきところを、単にAIとする場合がある点に留意されたい。

2　創造性とAI

AIは創造性を持ちうるのか。昨今、様々な場面で話題になるトピックであるが、この議論を進めるためにはまずは、創造性を定義する必要がある。日本創造学会によると、創造とそれに関する論点を以下のように整理している[3]。

人が（創造的人間／発達）
問題を（問題定義／問題意識）
異質な情報群を組み合わせ（情報処理／創造思考）
統合して解決し（解決手順／創造技法）
社会あるいは個人レベルで（創造性教育／天才論）
新しい価値を生むこと（評価法／価値論）

ここでは、創造性の主体を明確に人のみに限定しているが、これは生物の進化のように意図を持たないプロセスを除外するためであると考えられる。

主体的な意図を持たないコンピュータが創造性を持ちうるかは古くから議論されてきた。例えば、世界最初の汎用コンピュータの原型を19世紀中頃に

設計したチャールズ・バベッジの友人で、開発の良きパートナーであったエイダ・ラブレスは、コンピュータは

> 「何か独創するようには作られていない。それは、私たちがどのような実行を命令するか知っていることに限り、何でも実行することができる」

との文章を残している[4]。のちにアラン・チューリングによって、コンピュータが思考するかという命題に対する「ラブレス夫人の反論」として取り上げられ、有名になった一文である。この反論に対してチューリングは、ラブレスの言葉の変種として、コンピュータは（その作り手である）「人を驚かせることができない」を挙げ、即座に反論している。チューリングによると「そもそも機械は、きわめて頻繁に私を驚かせている」とする[4]。

本稿では、チューリングがいう「コンピュータは我々を驚かせる」という立場に立ち、創造性の定義をAIのそれを含むかたちで拡張して捉えることとし、ここでは創造性を「新しく（novel）、意外性のある（surprising）、価値ある（valuable）アイデアを生み出す能力」とするBoden[5]の定義に基づく。コンピュータが我々が想像もしなかった新しいアイデア（例えば音楽）を生成して我々を驚かせたとしても、それが意味のあるアイデアでないと創造的とは言えない。単なるランダムな音符の列は新しく、意外性を持ちうるが、音楽としての価値は低いと言わざるを得ない、したがって創造的とは言えないというわけだ。

Bodenは併せてP-CreativityとH-Creativityの区分を提唱する。P-Creativityは、幼児が砂場で遊ぶ時や家庭で新しいレシピを試すときのような個人的な（Personal）創造性の発露を指す。一方で、H-Creativityは人の歴史上（Historical）に存在しなかったような新しいアイデアを作り出す創造性を指す。我々がピカソやスティーブ・ジョブスといった人物とともに想起する創造性の概念である。

一般に人が、AIは創造性を持ち得ないと言う場合に想定している創造性は、後者のH-Creativityの場合であると考えられる。なぜなら後の節で詳しく述べるように、一般的な教師あり学習のモデルにおいても、新しい音楽や画像

を生成することはできる。しかし、それはあくまでもパーソナルなレベルの新しさや有用性であり、H-Creativity には原理的になり得ない。なぜなら、教師あり学習モデルの場合には、基本的にすでに存在する答え、この場合は人がこれまでに作ってきた音楽や絵画のパターンを学習することになるからだ。

教師あり学習のアルゴリズムにおいては、出力がいかに学習データとして与えられたお手本に近いかが評価となり、できるだけこの差（誤り loss）が小さくなるように学習を進める。レンブラントの絵を学習データとして与えられたモデルにとって、ピカソの絵は明らかに「誤り」とされてしまう。

たとえ教師あり学習だとしても、何らかの学習過程での不具合や AI モデルの誤用が、全く新しい表現に結び付く可能性自体は否定できない。ただし、その場合には AI モデルよりも、そうした不具合や誤用を見逃さなかった人間の側に創造性の所在が寄与されるべきだと考える。「半分壊れたコピー機がたまさか予想外の面白い画像を出力するのを期待する姿勢」とは、AI を題材に取り上げたロンドンでのアート展を評した批評家の言葉[3)]だが、あながち的外れとは言えないだろう。

AI が、H-Cretivity を持ちえるとしたら教師なし学習に可能性がありそうではあるが、それもそれほど容易ではない。なぜなら、囲碁や将棋のようにルールがはっきりしている領域とは大きく異なり、表現の領域ではその良し悪しをコンピュータ上で定量的に評価することは非常に難しいからだ。ゴッホがその死後にようやく評価されたように、未知の新しいスタイルの表現の評価を行うのは人の専門家でも難しい。ましてやコンピュータ上で定式化することができるのだろうか。

こうした疑問を念頭に置きつつ、次節では教師なし学習に基づく AI 生成モデルとして昨今注目を集める生成的敵対ネットワーク（Generative Adversarial Networks、以下 GAN）[6] を取り上げ、表現との関係について考察を続ける。

3　GAN とアート

深層学習の考え方そのものが画像認識タスクに対する取り組みの中で生ま

れてきたこと、また学習に利用できる画像データが音声よりも格段に豊富に存在していることなどの理由から、AI生成モデルの研究・実践は音楽よりも画像生成の領域で先行している。

2018年にはAIが「描いた」とされる絵画が世界的なオークションハウスChristie's で競り落とされ、世界中で大きなニュースとなった[4](図1)。この絵を出品したのは、フランスで機械学習を学ぶ学生、ビジネススクールに通う20代の学生から構成された「アーティスト集団」、The Obviousのメンバーたちで、当初予想された、7000から1万ドルという予想をはるかに超え、結果的に432500ドル(日本円で約4800万円)で落札され、様々な議論を呼んだことは記憶に新しい。

「Edmond de Belamy」と名付けられたこの絵の生成には、GANのアルゴリズムが用いられている[6]。GANは生成器(Generator、以下*G*)と識別器(Discriminator、*D*)という二つの人工ニューラルネットワークをまさに「敵対」させることで学習を行うアルゴリズムである(図2[5])。生成器*G*が学習デー

図1　Portrait of Edmond de Belamy ―「AIが描いた」肖像画

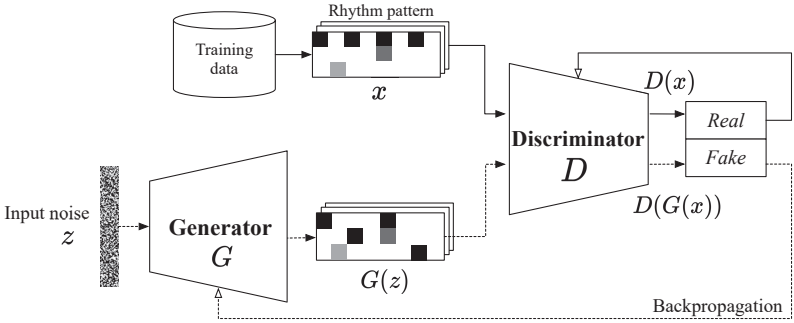

図２　Generative Adversarial Networks（GAN）の概念図

タに含まれるデータのパターンを学習し、ランダムなノイズを入力として学習データに類似するデータを生成するように学習を進めるのに対し、識別器 D のタスクは入力されたデータが学習データに含まれるいわば「本物」のデータなのか、G が生成したいわば「偽物」なのかをより正確に識別できるように学習を進める。この二つのネットワークがお互いを出し抜こうとすることで学習が進み、最終的には学習データにそっくりなデータを生成できるようになる、というのが大まかな枠組みである。

GAN は下記の min-max 関数として定式化される。

$$\min_{G} \max_{D} V(G, D) = \log(D(x)) + \log(1 - D(G(z)))$$

$V(G, D)$：D が最大化、G が最小化しようとする目的関数
$D(x)$：D が判断した、入力データ x が学習データに由来する確率
$G(z)$：ランダムなノイズ z を入力として、G が出力するデータ

GAN の目的は与えられた学習データに内在するパターンを定式化することにあり、入力に対してあらかじめ一対一の関係で与えられた答えを導くわけではないため、一般に教師なし学習の一種と分類される（学習の過程では教師ありの誤差を利用している）。

先述のオークションされた絵画の生成には、パブリックドメインの絵画を

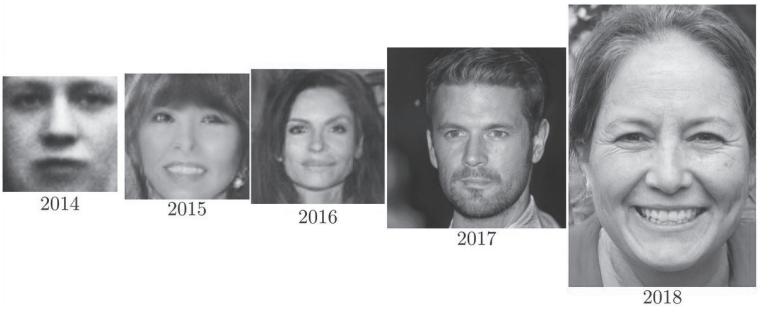

図３　GANで生成される画像の進歩[6][7][8][9][10]

集めた「絵画のWikipedia」、Wikiart[6]が公開しているデータセットを利用し、15000枚の14世紀から20世紀までのヨーロッパの肖像画が学習データとして用いられた。一旦、GANの生成器の学習が終われば、あとは生成器に入力する乱数を変えることで、様々なバリエーションの17世紀抽象画風の画像が無数に生成される。実際にEdmond de Belamyは、Belamy家という架空の貴族の家系の一連の肖像画のうちの一枚として、出品したアートコレクティブによって位置付けされている。

GANのフレームワークが提案されて以降、生成される画像の精度は年々向上し、GANを用いることで本物と見紛うばかりの顔写真（図3）が生成できることが示されている。一方で、GANはあくまでも学習データに含まれるパターンを学習し、汎化されたパターンの中でのサンプリングを行っているに過ぎないとも言える。言い換えると、GANはあくまでも過去の作品のパターンを踏襲し、そのパターンの中での生成を行っているに過ぎず、新しい表現を生み出しているわけではない。

その証拠に簡単な思考実験として、完璧なGANの生成器、識別器を想定するとよくわかる。学習データにあったパターン（上記のオークションされた絵画の場合は、ヨーロッパの肖像画）を丸暗記して再現する生成器と、学習データにあるデータだけを「本物」、それ以外全てを「偽物」とする識別器があれば、学習の損失関数の出力はゼロになる。したがって、本質的にGANのアルゴリズムは新しい表現を生み出すようには定式化されていないと言える。

もう一つ、このオークションが象徴していたのは、AIの自律性に対する誤解である。絵画を出展したアーティスト集団は、GANのアルゴリズムを表す上記の数式を画家の書名の代わりに書き入れ、AI「が」描いたという点を強調する発言を繰り返していた（図1 右下）。しかし、現在のAIシステムは全て特定の目的に特化したものであり、自律的な意図や目的意識を持つことはない（弱いAI）。AIはあくまでも人が設定した条件と与えられた学習データの中で学習し、画像を生成したわけで、AIを使って人が描いたというのが素直な見方であろう。

こうした現状を踏まえて、以降の節では特に音楽領域におけるAIの活用について取り上げるとともに、創造性の高い音楽表現を生み出すための一つのフレームワークを提案する。

4　深層学習と音楽生成

コンピュータで自動的に音楽を生成するという夢は、コンピュータそのものの歴史と同じくらい古くまで遡ることができる[7]。広義での人工知能（AI）を用いた音楽生成も、アルゴリズム作曲 / 自動作曲として古くから試されてきた。一方で昨今AIに注目が集まる大きなきっかけともなった深層学習を音楽生成に用いた例は、画像生成の研究例に比べるとまだ数が少ない[11]。

4.1　表現形式

AIに限らずコンピュータで音楽を扱う場合、その情報をどのように「表現」（Representation）するかが最初の考慮すべき点となる。現在、音楽生成用のAIモデルのために以下のような音楽の表現方法が活用されているが、それぞれに一長一短がある。

（1）音声シグナル Audio Signal

音楽をあつかうのであれば、音声信号を扱うのが一番自然に思えるが、波形として音楽をそのまま深層学習で生成するような研究例はまだそれほど多くはない。その理由としては主に計算量が膨大になり、必要な計算資源もそれに従って爆発的に増えるからである。

2020年4月に発表されたOpenAIのJukeboxは、そうした常識を覆す研究で、ボーカルや伴奏を含むCDクオリティの楽曲を音の波形としてそのまま出力するモデルを提案している[12]。非常に高価なGPUを数百台何週間にもわたって動かし続けた結果であることが論文内で明かされており、そうしたリソースを持たない一般の研究者、ましてやアーティストにとって、こうしたアーキテクチャの実用性は低いと言える。

（2）記号 Symbolic

音声信号の代わりに広く用いられているのが、音楽の内容をシンボリックに表現した、いわば楽譜の情報として音楽を扱う手法である。具体的なフォーマットとしては、MIDI規格を利用する場合が大半である。MIDIでは、音の高さ（pitch）、強さ（velocity）、長さ（duration）の情報として、楽譜内の各音符が扱われる。MIDIには楽譜の情報をほぼ忠実かつ簡潔に表現できるという利点がある一方で、表現された楽譜を実際にどう音に変換するかというところに、曖昧さがある点には留意する必要がある。

4.2　アーキテクチャ

具体的な深層学習を用いた音楽生成のためのアーキテクチャとしては、再帰的ニューラルネットワーク（Recurrent Neural Network、RNN）を音楽の時系列データに応用した研究例が数多く見られる[11]。RNNは時系列データの識別や生成に広く用いられており、自然言語処理システムのベースとなるアーキテクチャである。

例えば、Eckらは、RNNを拡張したLSTM（Long Short-term Memory）を使ってメロディーとコードを出力するシステムを2002年に提案している[13]。メロディー用に13のピッチと12のありうるコードをあわせた25ノードを入力と出力とし、隠れ層としてメロディー用に4つ、コード用に4つ、合わせて8つのLSTMのユニットを用いたアーキテクチャである。メロディー、コードそれぞれの入出力のノードはfully connectedされているうえに、コード用のLSTMのみ、メロディーの出力層ともつながっているのが特徴で、こうすることでコードに基づいたメロディーの生成が可能になると主張する。

その後、同じくRNNをベースに、発音される音符の長さの微妙なタメや強弱なども合わせて学習することで、より自然なピアノ曲を生成できるようなモデルも提案されている[14]。いずれも数小節程度の短いメロディーを生成することに関しては問題ないものの、楽曲全体を通してのマクロな音楽構造の構築を苦手とする傾向が指摘されている。

現在、文書生成の領域では、先行する文章で使われている単語列の構成とともに、そのどこに着目して次の単語を選択すべきかを合わせて学習する、いわゆるAttention機能をRNNに追加するのが一般的である[15]。さらにこのAttentionの機能だけで構成されるTransformerネットワークが提案され大きなブレークスルーを生んでいる[16]。遅れて、音楽生成の領域でもTransformerが応用され、マクロな音楽構造をも以前のモデルに比べて格段に上手く構成できることが確認された[17]。ここまで紹介した研究例は全て教師あり学習のモデルを利用している。

教師なし学習を用いた研究としては、Variational Autoencoder（VAE）を用いて複数のトラック（楽器）を組み合わせた楽曲が生成できることなどが示されている[18]。

また、前節で取り上げたGANを用いた手法も、スケールで条件付けした上でそのスケールに則ったメロディを生成できること[19]や、ポリフォニックな楽曲の生成が可能であること[20]などが示されているが、画像生成のためのGANのように多数の手法が提案されるには至っていない。

5　Rhythm Creative―GANを拡張した新規性の高いリズムを創出するモデル―の提案

前節、前々節でGANの導入とAIの音楽生成領域での活用について概観したところで、本節ではGANのアルゴリズムを拡張し、新奇性の高い表現を生み出すように方向付けした音楽生成モデルを提案する。

音楽を構成する要素の中でも特にリズムの生成を目標とし、対象とするジャンルとして、リズムマシンやシンセサイザーなどの音源から構成されるダンスミュージック（Electronic Dance Music, EDM[8)]）を想定する。

EDMは、1970年代、80年代のディスコやヒップホップなどから、90年代

のハウス、テクノ、90年代後半のジャングル、ドラムンベース（Drum and Bass）、そして今世紀に入ってジュークやトラップなど、多種多様なサブジャンルが毎年のように生まれる音楽ジャンルである。しかもその区分の多くが、リズムパターンによって特徴付けられる。したがって、本論の目的は、AIモデルを使ってこれまでになかったダンスミュージックのリズムパターンを生成できるか、新しいサブジャンルを生み出すことができるかという問いであると言える。

5.1 概要

前述のように創造性とは「新しく、驚きがあり、価値がある」アイデアを創出する能力を指すと定義している。また、囲碁などとは異なり、表現の領域では未知の解（作品）の甲乙を定量的に評価することが難しいこともすでに触れた通りだ。

本節で紹介する手法は、GANのアルゴリズムを拡張することで、「新しさ」や「驚き」を生み出すことを試みる一方で、従来のGANのアルゴリズムの特徴を生かし、従来の表現との類似性をある程度保つことによって、その「価値」や表現としての「それらしさ」をも担保しようとする試みとなっている。

創作活動におけるアーティストの心理を研究する心理学者のColin Martindaleによると、アーティストは常に、自分自身を含めた鑑賞者に知的な興奮を与えるべく、作品の持つ精神的な覚醒力（Arousal Potential）を高める方向で、マンネリを抜け出そうと制作を進めてきたという[21]。一方で、あまりに新規すぎる作品は見る人に受け入れられにくいので、ポテンシャルの変化は小さくしたい。刺激を求める力とマンネリに留まろうとする力が拮抗する中で、刺激を高める方向に向かう力がわずかに強いことによって、アートをはじめとする人間の創作活動は前進してきた。提案手法は、こうしたアーティストの心理を簡易的にモデル化している。

5.2 実験

まず対象として扱うのは、シンボリックな情報、具体的にはMIDIデータとする。

学習データとしては、商用のリズムパターンを集めたMIDIデータ集を利用した[9]。ロックやブルース、カントリーといった多様なジャンルを含む34828個のMIDIファイルのうち、1454がEDMとしてラベル付けされているとともにGeneral MIDIのフォーマットに則ったものであることがわかった。General MIDIは、MIDIのノートナンバーと楽器の対応付けを予め規定したもので、これによってキックドラム、スネアドラムなどに対応するMIDIノートを把握することができる。それ以外のMIDIファイルは、音色との対応付けが規定されていないため、学習データとしては無視することとした。

このように集めたMIDIデータの中に含まれる表1の、9つのジャンルを本実験では扱うこととする（以下、$K = 9$としてジャンル数を表現する）。

5.3　リズムの表現

本実験では、*Kick*、*Snare*、*Hi-hat closed*、*Hi-hat open*、*Cymbal*、*Low Tom*、*High Tom*、*Clap/Cowbell*、*Rim*の9つのドラム音を考慮する。いずれもダンスミュージックで広く使われるドラムマシンに標準で搭載されている音である。生成するリズムは2小節のパターンを考え、最小の時間単位は16分音符とする。したがって、9 × 32のピアノロール表現（マトリクス）で生成されるリズムが表現されることとなる（図5）。

生成されたリズムを定量的に評価するために、Toussaint[22]が提案する以下のswap distanceをリズムの類似度を表現する指標として扱う。二つのリズム

Breakbeats
DnB (drum and bass)
Downtempo
Garage
House
Jungle
Old Skool
Techno
Trance

表1　学習データに含まれるダンスミュージックのジャンル

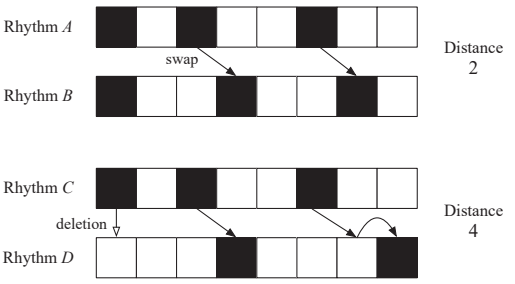

図4　二つのリズムパターンの距離（swap distance）の例[22]

AとBのswap distanceは、リズムAをBに変換するために必要な最小のswap（入れ替え）の回数として算出される（図4）。

図6は、学習データに含まれるリズムパターンの距離のマトリクスを示している。当然同じジャンル内の平均的距離は押し並べて小さい。特に*downtempo*にあたるジャンルはリズムパターンの画一性が高く、*techno*と*old_skool*（オールドスクール・ヒップホップ）は乖離していることがわかる。

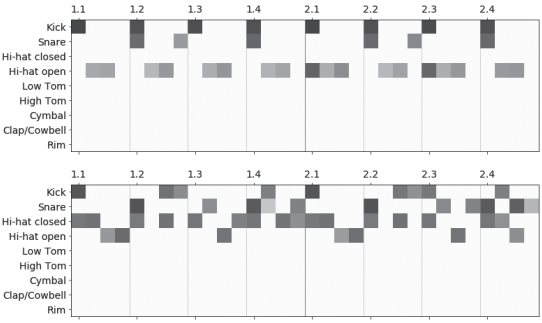

図5　学習データに含まれるリズムパターンの例とそのマトリクス表現
（上：*House*、下：*Breakbeats*）

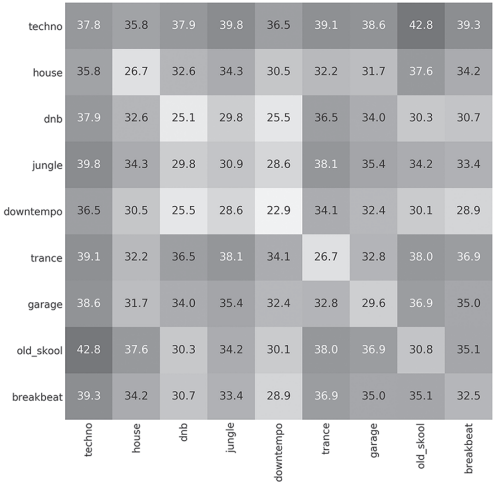

図6　学習データに含まれるリズムパターンの平均距離（ジャンル別）

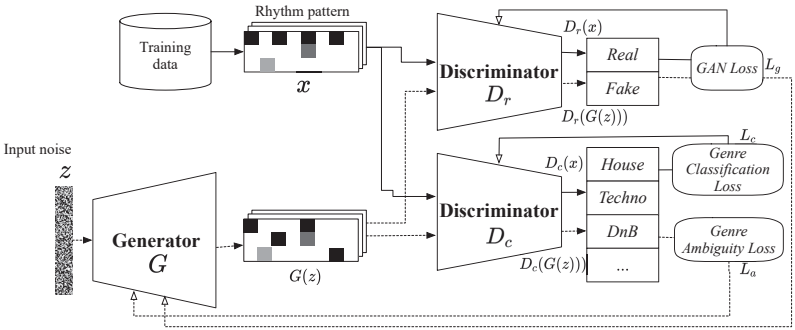

図 7　Genre Ambiguity Loss を含むように拡張した GAN フレームワーク

5.4　提案手法── ジャンルの曖昧さに基づく誤差関数（Genre Ambiguity Loss）

本研究では、元々の D に加えて 2 つ目の D を持つ、拡張した GAN のフレームワークを提案する。ここでは、元の GAN の D を D_r とし、2 番目の D は生成されたリズムのジャンルを分類する識別器とし、D_c と呼ぶことにする（図 7）。

D_r が、生成されたドラムパターンと学習データのドラムパターンを区別するように訓練されるのに加えて、D_c は学習データのリズムパターンに対して、それが K ジャンルのうちのどれにあたるのかを分類するように学習を行う。識別器の D のコスト関数に、ジャンル分類損失 L_c を加えることとする。

一方、G は、D_r と D_c の両方を「混同」させるようにして、D_r が本物だと信じるように訓練するだけでなく、D_c がジャンルを識別できないように訓練する。

生成されたパターンのジャンルの事後確率のエントロピーは、事後確率 $p(c|G(z))$ が等確率（equiprobable）であるときに最大となる。同様に、事後確率が等確率な場合、事後確率と一様分布との交差エントロピーは最小化される。この実験では、事後確率のエントロピーを最大化するのではなく、この交差エントロピーを最小化するように G の学習を促す。すなわち、D_c のジャンル分けの判断がつかない、「どのジャンルでもあり得る」という状態を目指して、G の学習が進むこととなる。

この交差エントロピーを、D_c の不確実性に対応する損失とし、これを

Genre Ambiguity Loss と呼ぶことにする。

このGANの設定を以下、*Creative-GAN* と呼ぶ。

G と D を含む全体のコスト関数は、以下のように再定義できる。

$$\begin{aligned}\min_D \max_G = &\log\left(D_r(x)\right) + \log\left(1 - D_r(G(z))\right) \\ &+ \log\left(D_c(c = \hat{c}|x)\right) \\ &- \sum_{k=1}^{K} \left(\frac{1}{K} \log\left(D_c\left(c_k|G(z)\right)\right) \right. \\ &\left. + \left(1 - \frac{1}{K}\right) \log\left(1 - D_c\left(c_k|G(z)\right)\right) \right)\end{aligned}$$

x と $\hat{c}$ は学習データ *Pdata* に含まれる実在するリズムパターンとそのジャンルである。z はランダムな潜在ベクトルであり、$G(z)$ は G によって生成されるリズムパターンを指す。

平易な言葉で言い直すと、Creative-GANのアルゴリズムでは、ジャンルを識別する識別器によって、どのジャンルとも識別がつかないリズムほど、高く評価される。一方で、元来のGANのアルゴリズムにある識別器はそのまま生かされているため、「真偽」を見極める識別器によって、生成されたリズムのダンスミュージックとしてのリズムらしさの検証は行われる。リズムらしさを担保しつつ追加された識別器を混乱させるように生成器の学習を進めることで、過去のどのジャンルにも属さない[2]新しいスタイルのリズムが生成されることが期待できるという構図になる。

こうした付加的な識別器と誤差関数の構造は、Elgammalらの研究によって最初に提案された[23]。この研究[23]では、様々な時代、スタイルの西洋絵画を名作を大量に集め、GANのアルゴリズムで絵画の生成を試みている。印象派、キュビズム、ロマン派といった絵画のスタイルを識別する識別器を別途追加し、この識別器を混乱させるように生成器の学習を進めた結果、抽象的な絵画表現が生成されることが示されている。

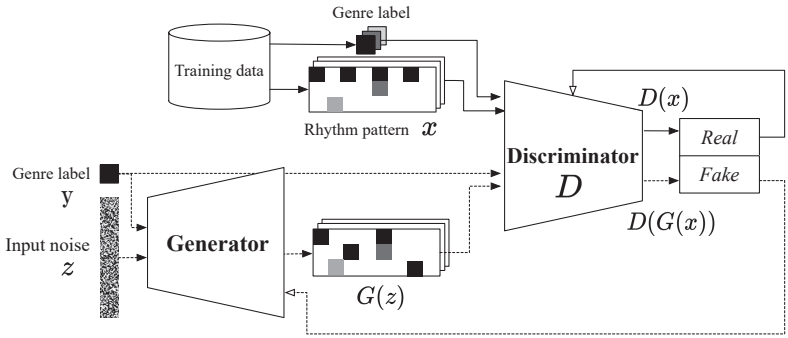

図 8　Overview of Genre-conditioned GAN

5.5　予備実験― ジャンルで条件付けされたリズム生成

まず最初に予備実験として、ジャンルで条件付け (condition) を行った上で指定したジャンルのリズムが生成されるかどうかの実験を行った (図 8)。このモデルでは、識別器と生成器それぞれに、ジャンルで条件付けを行うための入力 y を追加する。y はスカラー［0, K］であり、K は学習データのジャンル数 (この場合は $K = 9$) である。

D では、入力 y は埋め込み層 (embedding layer) を介して、データ表現の項で説明したドラムのオンセット行列と同じ形のベクトルに変換された上で、リズムパターンを表現する行列と連結され、双方向 LSTM の 2 つの層に供給される。G は y を z と同じ大きさのベクトルに埋め込んで、埋め込んだラベルと z を要素ごとに掛け合わせたものを、生成器の入力として使用する。そして、G と D は、リズムパターンとそれぞれに対応するジャンルラベルを用いて、敵対的に学習する。これらのアーキテクチャは、Mirza・Osindero の研究成果 [24] をベースにしている。

5.6　実装

本実験で扱う GAN の識別器はそれぞれ 64 ノードを含む 2 層の双方向 (Bi-directional) LSTM とそれに続く全結合層からなる。全結合層の出力には、シグモイド関数の活性化関数が接続され、その出力が入力されるリズムパターンの真偽 (学習データに含まれるパターンか、それとも生成器の出力か) を識

別する。

一方、生成器 G は、(128, 128, 9) のノードを持つ三層の LSTM レイヤーから構成される。G の出力は、9 × 32 のマトリクスとなる。入力されるランダムな潜在ベクトル z は 100 次元のベクトルとして設定した。活性化関数としては、先述した D の最終層を除いて、LeakyReLU を用い、最適化アルゴリズムとしては Adam を用いている。

実装には TensorFlow バックエンドの Keras を用いた。学習に用いた Python のソースコードと学習データは、Web 上で公開している [10]。

5.7　実験結果― ジャンル条件付け

最初の実験では，ジャンルで条件付けした GAN アーキテクチャが、指定されたジャンルのリズムパターンを生成できることが確認できた。生成されたリズムパターンの例は、Web サイトで試聴することができる [10]。図 9 には、生成されたリズムパターンとそれに対応するジャンルラベルの例を示す。

5.8　実験結果― ジャンル曖昧さロスを含むモデル

続いて、D_c と L_c を含む、2 つ目のモデル、Creative-GAN モデルの実験結果を示す。このモデルの学習は 90 エポック目あたりで収束した。図 10 に、94 番目のエポックで生成されたパターンの例を示す。生成されたリズムはダ

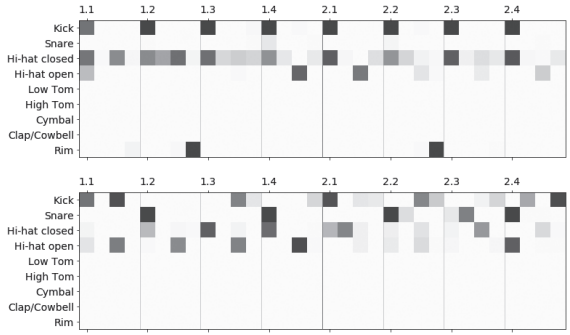

図 9　ジャンルで条件付けした GAN で生成したリズムパターンの例
（上：*House*、下：*Breakbeats*）

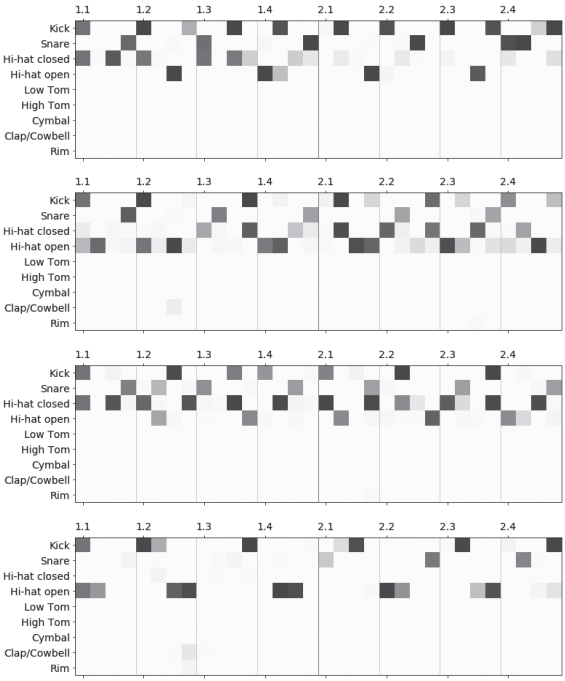

図 10　提案手法 Creative-GAN モデルで生成したリズムパターンの例

ンスミュージックのリズムらしさを担保しつつも、今までにない独特のリズムパターンが生成されることがわかった。生成されたリズムの例は、同じく Web ページ上で試聴できる [10]。

5.9　考察

続いて、ランダムな z の入力ベクトルを用いて 500 個のパターンを生成し、学習データからの距離を計算した。図 11 は、学習データに含まれる全てのジャンルで高い距離を示している。これは、モデルが生成したリズムパターンが、これらのジャンルから乖離していることを意味している。

Creative-GAN で生成されたパターン内の平均距離の値は、学習データの距離よりは低いが、図 6 の値よりは相対的に高い。これは、生成されたリズムパターンに多様性があり、モデルが Mode Collapse (z の値に無関係に $G(z)$

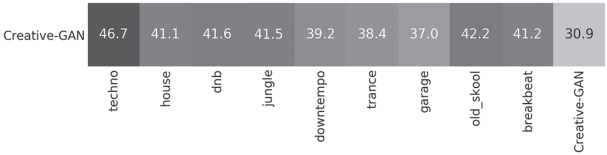

図 11　提案手法で生成したリズムパターンと学習データの距離

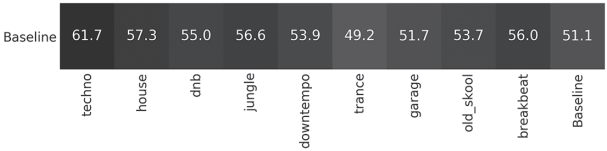

図 12　学習データの統計データに沿って生成したランダムなリズムパターンと学習データの距離

の値が、特定の出力に限定される状況）を回避できたことを意味している。ベースラインとしては、学習データに含まれるリズムパターンの打点の数の平均値と標準偏差からランダムにサンプリングして生成したリズムパターンと比較する（図 12）。図 12 の方が、図 11 よりも高い距離値を示しており、Creative-GAN で生成したリズムの方が学習データに近いことがわかる。前述の Martindale が記したアーティスト心理の反映と比較することができる。

本提案手法によって、学習データに含まれる既存のジャンルに属さない新奇性の高いリズムが生成されることが示された。とはいえ、音楽の定量的な評価には限界があり、今後は試聴者を集めて定性的な評価を行う必要があるだろう。

6　まとめ

本稿では、AI と創造性の関係について、GAN を例にとって考察した。その上で、音楽における AI 活用の現状について簡単にまとめた上で、そのほとんどが「それらしい」作品の再生産を目指す、模倣のための AI であることを示した。

人と技術の歴史を振り返ってみるに、機械による模倣が新しい表現を生み出してきたのは、写真と印象派やキュビズムと言った絵画の関係を見ても明らかである [25]。したがって、AI による過去の作品の再生産、模倣自体に価値

がないわけではない。アート批評家がいう「壊れたコピー機」としてのAIは、アーティストに新しいアイデアを与え、表現の領域を拡張することに寄与する（紙面の関係で本稿では触れられなかったが、筆者はAIの不完全性を生かした作品、パフォーマンスを多く手掛けている）。

一方で後半で紹介したように、教師なし学習のフレームワークを拡張することで、あえて未知の表現領域にAIの学習を方向付けることができることも示した。GANのアルゴリズムを利用することで、音楽らしさを担保しつつ、現存するジャンルのどれにも当てはまらない表現を模倣する仕組みを提案した。実験を通して、創造的なリズム、新しく驚きがあり価値がある出力を得られることが示された。

ルールが明確な囲碁や将棋とは異なり、表現の良し悪しの判断を定式化することは難しい。特にその新奇性の高さを評価することは困難であると言えるだろう。本稿で提案した手法は、この難しい評価をジャンルの「曖昧さ」という比較的定量化しやすい指標に置き換えることで、AI自体に新奇性を模索する仕組みを持たせたという点が重要である。

一方で、このアルゴリズムが示す創造性も人（この研究を行った研究者）が最初に定めた枠組みの範疇から抜け出せてはいないと言える。あらかじめ定めた16分音符単位というグリッドの制限を超えたより細かいリズムや2ステップのようなシャッフルの効いたリズムが生成されることは、システム上ありえない。それでも、アーティストの心理をベースに、単に過去の表現を再生産（P-Creativity）するだけでなく、創造性の高い新しい表現を創出する方向（H-Creativity）に、学習を明示的に方向付ける枠組みは、AIと創造性の未来を考える上で非常に示唆的であると考える。こうした学習のフレームワークを設定したのは筆者であり、創造性の根元は人間にあると言うことはできる。AIに主体的な意図がない以上、AI自体が表現者、アーティストになることも考えられない。しかし、チューリングが言うように、確かにAIの出力は人を驚かすことができる。

創造性を志向するAI。模倣するAI。これらが入り混じり、人とインタラクションすることによって表現領域の拡張を目指す。その先に創造性とAIの未来があると言えるのではないだろうか。

注

1) Machine Learning for Creativity and Design https://neurips2019creativity.github.io/（2020年8月20日アクセス）
2) その他、semi-supervised learning など self-supervised learning などがあるがここでは割愛する。
3) 'I've seen more self-aware ants!' AI: More Than Human review - *The Gurdian* https://www.theguardian.com/artanddesign/2019/may/15/ai-more-than-human-review-barbican-artificial-intelligence（2020年8月20日アクセス）
4) The First Piece of AI-Generated Art to Come to Auction—Christie's. https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx（2020年8月20日アクセス）
5) 図2では、本論の研究に合わせて、生成する対象が音楽のリズムパターン（Rhythm pattern）であるとしている。
6) WikiArt.org - Visual Art Encyclopedia http://www.wikiart.org（2020年8月20日アクセス）
7) First recording of computer-generated music – created by Alan Turing – restored https://www.theguardian.com/science/2016/sep/26/first-recording-computer-generated-music-created-alan-turing-restored-enigma-code（2020年8月20日アクセス）
8) EDMというと日本では、派手なリフやシンセ音を特徴とするダンスミュージックのサブジャンルを指す言葉として定着しているが、一般には電子音を使ったダンスミュージック全般を指す言葉として使われる。
9) Groove Monkee Mega Pack GM https://groovemonkee.com/products/mega-pack（2020年8月20日アクセス）
10) https://cclab.sfc.keio.ac.jp/projects/rhythmcan/（2020年8月20日アクセス）

参考文献

[1] Goodfellow, I. J., Bengio, Y., and Courville A. (2016) *Deep Learning,* MIT Press.
[2] 松尾豊（2015）『人工知能は人間を超えるか：ディープラーニングの先にあるもの』KADOKAWA.
[3] 高橋誠（2002）『創造力事典』日科技連出版社.
[4] Turing, A. M. (1950) "Computing Machinery and Intelligence", *Mind.* 59, pp 433-60. https://doi.org/10.1093/mind/LIX.236.433.
[5] Boden, M. A. (2009) "Computer models of creativity", *AI Magazine*. 30(3), pp. 23-34.
[6] Goodfellow, I. J. et al. (2014) "Generative adversarial nets", *Advances in Neural Information Processing Systems.* 3, pp. 2672-2680.
[7] Radford, A., Metz, L., and Chintala, S. (2016) "Unsupervised representation learning with deep convolutional generative adversarial networks", *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*.
[8] Liu, M.-Y. and Tuzel, O. (2016) "Coupled generative adversarial networks", *Advances in Neural Information Processing Systems*.
[9] Karras, T., Aila, T., Laine, S., and Lehtinen, J. (2017) "Progressive Growing of GANs for Improved Quality, Stability, and Variation", *ICLR*. pp. 1-26.
[10] Karras, T., Laine, S., and Aila, T. (2019) "A style-based generator architecture for generative adversarial networks", *Proceedings of the IEEE Computer Society Conference on*

Computer Vision and Pattern Recognition.
[11] Briot, J.-P., Hadjeres, G., and Pachet, F. (2019) *Deep learning techniques for music generation.* Springer.
[12] Dhariwal, P. et al. (2020) *Jukebox: A Generative Model for Music.*
[13] Eck., D. and Schmidhuber, J. (2002) “A First Look at Music Composition using LSTM Recurrent Neural Networks”, *Idsia.* pp. 1-11.
[14] Oore, S. et al. (2020) “This time with feeling: learning expressive musical performance”, *Neural Computing and Applications*. 32(4), pp. 955-967.
[15] Bahdanau, D., Cho, K. H. and Bengio, Y. (2015) “Neural machine translation by jointly learning to align and translate”, in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.
[16] Vaswani, A. et al. (2017) “Attention is all you need”, *Advances in Neural Information Processing Systems*.
[17] Huang, C.-Z. A. et al. (2019) “Music transformer: Generating music with long-term structure”, *7th International Conference on Learning Representations, ICLR 2019*.
[18] Roberts, A., Engel, J., Raffel, C., Hawthorne, C., and Eck, D. (2018) “A hierarchical latent vector model for learning long-term structure in music”, *35th International Conference on Machine Learning*.
[19] Yang, L.-C., Chou, S.-Y., and Yang, Y. Y.-H. (2017) “Midinet: A convolutional generative adversarial network for symbolic-domain music generation”, P*roceedings of the 18th International Society for Music Information Retrieval Conference, ISMIR 2017*, pp. 324-331.
[20] Dong, H. W. and Yang, Y. H. (2018) “Convolutional generative adversarial networks with binary neurons for polyphonic music generation”, in *Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018*, pp. 190-196.
[21] Martindale, C. (1990) *Clockwork Muse,* Basic Books.
[22] Toussaint, G. (2006) “A Comparison of Rhythmic Dissimilarity Measures”, *Forma*.
[23] Elgammal, A., Liu, B., Elhoseiny, M., and Mazzone, M. (2017) “CAN: Creative Adversarial Networks, Generating “Art” by Learning About Styles and Deviating from Style Norms”, *the eighth International Conference on Computational Creativity (ICCC)*.
[24] Mirza, M. and Osindero, S. (2014) *Conditional Generative Adversarial Nets*.
[25] Hertzmann, A. (2018) “Can Computers Create Art?”, *Arts*.

〔受付日　2020. 9. 24〕