

恋ダンスの達人： 深層学習を利用したダンスのキレ採点システム

森 鈴果^{1,a)} 田中 哉汰^{1,b)} 橋田 光代^{2,c)} 片寄 晴弘^{1,d)}

概要：“踊ってみた動画”が流行するなど未経験者でもダンスに触れる機会が増えているが、ダンスの優劣は感性や感覚に依存するため判別が難しい。そこで、我々は踊ってみた動画とコメントでの“キレ”という用語に着目し、深層学習による“キレ”判別器の構成を試みた。楽曲のテンポ情報と OpenPose による骨格情報を入力とするモデル構成により 85 %の精度で“キレ”を判別できることを確認した。本稿では、この識別器を利用して構成したダンスの“キレ”採点システムについて紹介する。

キーワード：ダンス, OpenPose, CNN, LSTM

1. はじめに

SNS の普及により、音楽や映像に触れる機会が多くなったことからダンスを身近に感じる場面が増えている。ドラマやアニメで主人公が踊るシーンをエンディングとした“踊るエンディング”、アーティストの振付を模倣した映像にハッシュタグをつけ、サイトにアップロードする“踊ってみた動画”など、ダンスはステージ上だけでなく動画でも他者に共有することができる。ポカリスエット PR 事務局が 2017 年に実施したダンス意識と経験調査 [1] によると、高校生では 33.7 %、大学生では 41.2 %がダンス動画の撮影経験があり、40 代と比べ約 3 倍となっている。撮影された動画は 78.0 %の人が他者に共有しており、ダンスは新たなコミュニケーションツールとなっている。TikTok などダンス動画が中心となる SNS も存在し、2020 年には 10 代の利用者が 29 %にも及んでいたことから認知度も上がりつつある [2]。このようにダンス動画の利用者は増加傾向にあり、身近なエンタテインメントとなっている。

ダンスと同様に、歌ってみた動画としてエンタテインメント性を持つ歌唱では、カラオケでの採点システムが点数によって評価されるという新たな楽しみ方を創造している。この採点システムが普及したことで、歌を習ったこと

がない人でもビブラートやしゃくりといった歌の表現要素を理解できるようになった。

ダンスにも身近に評価を行うことができるシステムが開発されれば、楽しみ方が範囲が広げられるとともに、“上手さ”の指標となる基礎的要素ではあるが感性や感覚の範疇にあり言語化が難しいテクニックの習得にも寄与できると期待される。本稿では、ダンス動画を入力としてダンスの“上手さ”の要素の一つ“キレ”の採点を行うシステムについて述べる。

以下、2 章ではダンスに関する関連研究を述べる。3 章では本システムのアプローチ、4 章では 3 章のアプローチに基づいた採点システムの実装について述べる。5 章では、本システムの評価について考察を述べる。

2. 関連研究

ダンスに触れる機会が増えていることから様々なダンスに関する研究が行われている。河野ら [3] はダンスの中でもバレエに着目し、バレエ熟練者の運動学的特徴について研究を行なった。バレエ熟練者とバレエ未経験者では関節の動作に有意差が見られ、白鳥の湖による羽ばたき表現の上肢動作において熟練者は肘関節を先行させて動くことが明らかとなった。バレエ熟練者とバレエ未経験者では骨格の動作だけでも差が見られるように、ダンス熟練者とダンス未経験者ではそれぞれに共通する特徴が存在するのではないかと考えた。そこでダンス熟練者とダンス未経験者の特徴を抽出することを検討する。

ダンサーの特徴を抽出する方法に、古市ら [4] の機械学習を用いた骨格情報からダンスの個性抽出を行う研究があ

¹ 関西学院大学
Kwansei Gakuin University

² 福知山公立大学
The University of Fukuchiyama

a) dsi33189@kwansei.ac.jp

b) ddc18531@kwansei.ac.jp

c) hashida-mitsuyo@fukuchiyama.ac.jp

d) katayose@kwansei.ac.jp

る。この研究により特定の振付であってもダンサーによって自身の個性が表れる、その特徴が機械学習によって判別可能であることを明らかにした。さらに、この主観評価実験により、ダンサーの個性は手の形により表れるという結果が見られた。

またダンスジャンルの1つであるブレイクダンスは、自ら創造した動作がパフォーマンスの質として評価される。そこでパフォーマンスにどのくらい個性があるのか数値化するため、平澤ら [5] はブレイクダンスの動作におけるオリジナリティーの程度の評価を行った。この研究ではオリジナリティーについて評価を行うため、ダンスの分類は特定の振付だけでなく、学習データにはない新たなデータにも対応できるよう深層学習が用いられている。

ブレイクダンスにアクロバティックな動作が多く含まれるように、ダンスジャンルごとに動作の特徴が存在する。Tsuchida ら [6] はジャンルごとの特徴を抽出するため、ダンス動画からダンスモーションジャンル分類を行なった。OpenPose[7] により動画から骨格の2次元座標を取得することで、モーションキャプチャや深度カメラを必要とせずにダンサーの動作を認識している。この研究から、3秒程度の僅かな時間のダンスモーションでも、ダンス分類が可能であることが明らかとなった。

このようにダンスに関する研究では、ダンサーの個性やダンスジャンルの特徴を抽出するシステムが考案されている。しかしこれらの研究では、ダンスの優劣に関する分析は行われていない。本研究ではこれらの研究と同様にダンス熟練者、未経験者それぞれに共通する特徴を抽出することでダンスを評価できるのではないかと考えた。

楽曲に基づいてダンスは生成されることから、ダンスモーションとダンス楽曲を組み合わせた分析や生成を行う研究も増えている。楽曲からダンスを生成するアプローチの研究に、Tang ら [8] の楽曲に基づいたダンスモーションを生成する手法がある。楽曲の情報にピッチに関する情報だけでなく拍位置情報も付加したモーション生成を行うことで、高い精度を持つシステムを作成している。この研究では、楽曲情報を入力とし骨格情報を生成することから楽曲情報と骨格情報の特徴量の大きさを合わせる必要があるため、LSTMを用いた楽曲情報の次元圧縮の学習方法について提案された。

このようにダンス楽曲の研究には拍位置を用いる場合が多く、実際ダンスで使われている楽曲には強いビートが入ったものが多い。これらの研究から、ダンスの動作に拍位置情報を加えることでより正確なダンス分析を行えると考える。そこで本研究では、楽曲の拍位置を考慮したダンス採点を行う手法について検討する。

3. アプローチ

本研究では特別なモーションセンサを利用せず、身近な

ダンス動画に対して深層学習ベースで採点を行うアプリケーションの開発を目指す。この構成においてはダンス動画とその評価結果の大量なデータセットが必要となる。本研究では、ダンス動画そのものではなく姿勢推定により抽出された時系列骨格座標データを対象に採点機構を構築する。ダンス動画に対応する評価データについては、別途評価を実施するのではなく、投稿動画に対して寄せられるコメントに着目して確保するものとする。ここでは、時系列情報を扱うものとしてCNNとLSTMを併用した深層学習モデルを構成し、代表的な踊ってみた動画“逃げるは恥だが役に立つ [9]”のエンディングで踊られた“恋ダンス”を対象とした“キレ”採点機構を構成する。

3.1 学習用データセット

機械学習を用いたダンス動画の採点システムを構築するにあたってはダンス動画とその評定の組み合わせによる大量のデータセットの準備が求められる。ダンス動画に対して新たに評価を実施しようとするれば大変煩雑な作業が必要となる。動画投稿サイトYouTube[10]では“コメント”機能が実装されており、ダンスの評価に関するコメントも多く寄せられている。本研究では、ダンス動画に対して寄せられる評価系コメント“キレ”に着目することで、人手での評定作業を実施せずに教師用データに対する“キレ(あり/なし)”のラベルデータを用意する。ダンス動画31件を対象とした予備実験の結果、“キレ”に言及した“コメント”が2件以上あるものに“キレ”ラベルを付与した時の“キレ”判別率が最も高くなった。以降の実験では各ダンス動画に対しこの基準で“キレ”に関するラベルを付与する。

ダンス動画から採点を行うにあたって2次元の動画から人間の動作を測定する必要がある。そこで、画像の入力から人間の骨格を取得できる姿勢推定の手法を用いて人間の動作検出を行う。本研究では、この手法にOpenPose[7]を用いる(図1)。OpenPoseは、画像だけでなく動画に対しても利用可能であり、約30fpsでそれぞれのフレームから骨格を取得できる。この手法を用いて31本のダンス動画から人間の骨格を取得したものを、データセットとする。恋ダンスの踊ってみた動画には、上半身だけ写したものが多く見られることとダンスの個性抽出方法[4]からダンスの個性は手の形により表れやすいと考察されていたことから、左(右)の肩、肘、手首の6箇所の2次元座標に限定して取得する。

3.2 モデルの構成

ダンスは楽曲に基づいて生成されることから、楽曲の情報を考慮したダンス動作を生成する。ダンスの動作は拍のタイミングで動く(止める)ことでメリハリが生まれることから、楽曲の情報の中でも拍位置はダンスのキレにとって重要な要素である。そこでダンスの動作に拍位置情報を足



図 1 OpenPose による骨格座標検出

し合わせることで拍位置を考慮する。本研究では、拍位置情報に表拍と裏拍で異なるピッチを鳴らしたテンポ音を用いる。

また拍位置を考慮する方法に、Tsuchida ら [6] のダンスジャンル分類の研究がある。この研究では、拍位置を考慮しながらダンス動画を分割することで、楽曲の拍位置を考慮する方法を提案している。そこで本研究でも、拍位置に合わせてダンス動画を分割し、その長さ合わせたテンポ音を作成する。ダンスは 8 拍 (ワンエイト) を 1 区切りとしカウントや振付が行われることから、ダンス動画を 8 拍の長さで分割する。恋ダンスの楽曲は 158BPM であるため、3 秒程度の動画を分割したデータとする。31 件のダンス動画をそれぞれ 3 秒程度 (8 拍) の長さで分割することで、894 本のダンス動画によるデータセットを取得する。

以上のデータセットをシステムのパラメータを決定するためのトレーニングデータ (731 本)、初期値やパラメータの更新回数を決定するためのバリデーションデータ (82 本)、システムがどのくらいの性能なのか判断するためのテストデータ (81 本) にデータの分割を行う。尚、トレーニングデータとバリデーションデータに含まれるダンサーはテストデータに含まれていない。

平澤ら [5] のブレイクダンスのオリジナリティーの程度の評価や Tsuchida ら [6] はダンスジャンル分類のようなダンスに関する研究では、深層学習を用いることで高精度なモデルを実現可能としている。その理由として、深層学習の手法には時系列に特化した手法が多く存在することが挙げられる。そこで本研究では、時系列に特化した深層学習の手法である CNN と LSTM を用いてデータセットの学習を行う。

CNN は、時間軸方向に畳み込みを行うことで、時系列データの特徴抽出を行うことのできる手法である。ダンスを学習するにあたって特定の骨格だけを学習するのではなく、周囲の骨格を考慮しながら学習を行う必要がある。そこで、CNN を用いて周囲の骨格について合わせて畳み込み演算を行う。よって時間軸方向だけに畳み込み演算を行う 1 次元畳み込みではなく、時間軸方向と合わせて周囲の骨格についても畳み込み演算を行う 2 次元畳み込みの手法を採用する。さらに、より抽象的な特徴を得るため CNN の層を重ねることでダンスのキレを抽出する。

LSTM は、前フレームの状態を現在の状態にも影響させることで、時系列データ全体を通した学習を行う手法である。ダンス全体の流れを学習することで、時系列データ全体でどのような変化があるか抽出することができる。

しかし、CNN 単体では時系列全体を見渡すことのできない、LSTM 単体では計算コストが大きいといったように、これらを単体で用いる学習方法には欠点がある。そこで本研究では、これらを組み合わせることでダンスの時系列データから学習する方法を採用する。CNN の後に LSTM を組み合わせることで、CNN により次元を圧縮しながら特徴抽出された出力を、LSTM によって全体を通した学習を行うことで、動作のアクセントやポイントを抽出しながら時系列全体を把握した出力が得られるのではないかと考える。以上の深層学習を利用して 8 拍 (ワンエイト) のダンス動画とテンポ音から、ダンスのキレの特徴抽出を行う過程を図 2 に示す。

4. 実装

アプローチに基づき、ダンス動画から得た左 (右) の肩、肘、手首の 6 箇所における骨格の 2 次元座標で作成されたデータセットとテンポ音による音源データから、キレのある (ない) ダンスの特徴量を抽出する採点機構を構成する。しかし、これらのデータには欠損やノイズ、エラー値がある。これらに対応するため深層学習を行う前にデータの前処理を行う。

前処理を行なったデータから深層学習の CNN と LSTM を用いて、キレのある (ない) ダンスの特徴量を抽出する。始めに LSTM から構成された Tempo_Extractor でテンポ音の次元圧縮を行い、CNN と LSTM から構成された Dance_Extractor によってダンスのキレの特徴量を得る。これらのデータの前処理からモデルを実装するまでを以下で述べる。

4.1 骨格座標によるデータ作成

OpenPose から取得したダンスの骨格座標に、撮影したカメラと人間の位置関係等のダンス以外の情報を含まないようにするため、人間の位置や大きさの正規化を行う必要がある。そこで、骨格の首と腰の位置座標を基準に全体の骨格を線型写像する方法を採用する。

また骨格座標には、ダンス動画の中で OpenPose で判定できなかった、または動画に収まらなかった骨格が多く見られた。楽曲に基づいたダンスモーションを生成する手法 [8] では線形補間を利用したデータの再生成を行っていたが、ダンス動作は直線的な変動にはならない。そこで本研究では、3 次式により滑らかな線で補間を行う 3 次スプライン補間を用いて判定できなかった骨格を再生成する。

骨格座標からダンスの動作に関する情報を得るにあたって正規化と補間を行なった 6 箇所における骨格の 2 次元座

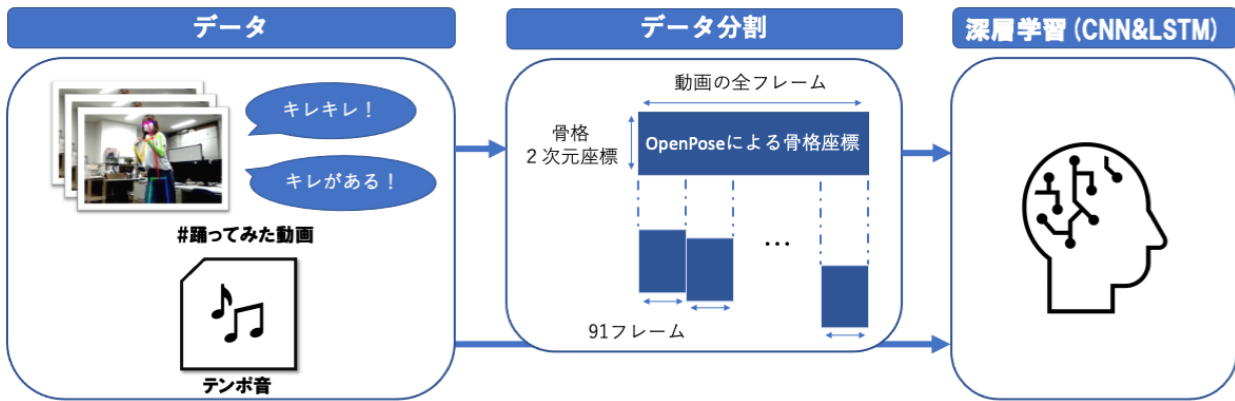


図 2 システムの概要

表 1 速度 (加速度)

骨格	(v_x, v_y)	(a_x, a_y)
左肩	v_i^1, v_i^2	a_i^1, a_i^2
左肘	v_i^3, v_i^4	a_i^3, a_i^4
左手首	v_i^5, v_i^6	a_i^5, a_i^6
右手首	v_i^7, v_i^8	a_i^7, a_i^8
右肘	v_i^9, v_i^{10}	a_i^9, a_i^{10}
右肩	v_i^{11}, v_i^{12}	a_i^{11}, a_i^{12}

標に対し、それぞれを 1 フレームごとに以下の計算を行う。動画の i 番目のフレームにおける骨格の位置座標を $x_{(i)}$ で示す。この位置座標に対し、フレーム間差分の速度と加速度を求めるとして速度 (加速度) 情報を定義する。速度を $v_{(i)}$ 、加速度を $a_{(i)}$ と表し、式 (1), (2) によって算出する。算出した結果を表 1 のように定義する。

$$v_{(i)} = x_{(i)} - x_{(i-1)} \quad (1)$$

$$a_{(i)} = v_{(i)} - v_{(i-1)} \quad (2)$$

さらに 2 次元畳み込みの際、それぞれの骨格が関わりながら学習が行えるよう速度 (加速度) の特徴ベクトルを構成する。関連性の高い骨格が 2 次元畳み込みのフィルターに含まれるように、表 1 による速度 (加速度) 情報の 12 次元の特徴ベクトルを式 (3), (4) のように定義する。尚、OpenPose から 30fps で 3 秒程度のダンス動画を分析するため、 $x_{(i)}$ は 91 フレームの長さを持つデータである。この $x_{(i)}$ に対し式 (1), (2) の計算を行うため、速度情報と加速度情報のフレーム数 (i) はそれぞれ 90 フレーム、89 フレームである。

$$\text{速度情報 } V_i = \langle v_i^1, v_i^2, \dots, v_i^{12} \rangle \quad (3)$$

$$\text{加速度情報 } A_i = \langle a_i^1, a_i^2, \dots, a_i^{12} \rangle \quad (4)$$

また、速度 (加速度) 情報は深層学習のモデルへの入力のため標準化する。 X' を変換後のデータ、 X を変換前のデータで示す。 Q_1, Q_2, Q_3 はそれぞれ X の第 1, 第 2, 第 3

四分位点とし、算出を以下の式 (5) で行う。データの外れ値を考慮するためデータの四分位点を基準にしている。

$$X' = \frac{X - Q_2}{Q_3 - Q_1} \quad (5)$$

4.2 テンポ音によるデータ作成

テンポ音による音源データから拍位置に対する特徴ベクトルを作成するため、オンセットを検出する。音の開始点であるオンセットはサンプリング周波数のフレーム間差分が正の数でありながら尚且つ大きいと考えられる。この条件から求めたフレーム間差分を、フレーム全体で正規化しオンセット検出を行う。手法には、音楽分析のための Python のライブラリである Librosa[11] を用いる。オンセット検出したものを、拍位置情報の 1 次元ベクトルとして式 (6) のように定義する。尚、8 拍の長さを持つテンポ音は 44.1kHz で分析を行うため、フレーム数 (i) は 131 フレームである。

$$\text{テンポ情報 } T_i = \langle t_i \rangle \quad (6)$$

4.3 テンポ情報からの特徴抽出

拍位置情報 T_i と速度 (加速度) 情報 $V_i(A_i)$ は異なるフレーム数を持つため、 T_i の特徴ベクトルを $V_i(A_i)$ と同じフレーム数まで圧縮する。楽曲に基づいたダンスモーションを生成する手法 [8] で LSTM を用いて楽曲情報の圧縮を行っていたことから、本研究では圧縮するための手法に LSTM を用いる。この手法を用いて、速度を入力する場合は 90 フレーム、加速度を入力する場合は 89 フレームまで次元圧縮する。また、速度 (加速度) 情報 $V_i(A_i)$ の 12 次元の特徴ベクトル全てに圧縮した拍位置情報を用いる必要がある。そこで、 T_i の 1 次元ベクトルを 12 次元へ要素の複製を行う。この次元圧縮から複製までの処理を Tempo_Extractor と定義する。 $Tempo_Extractor(T_{(i)})$ の出力に $V_i(A_i)$ を足し合わせることで、拍位置情報を考慮し

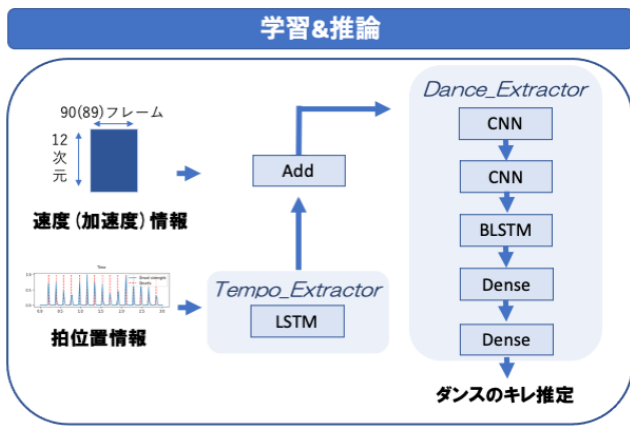


図 3 モデルの概要

た速度(加速度)情報とする。

4.4 テンポを考慮したダンスのキレの特徴抽出

拍位置を考慮した速度(加速度)情報から Dance_Extractor を用いてキレの特徴抽出を行う。Dance_Extractor は拍位置情報を考慮した速度(加速度)情報を入力とし、拍位置でのダンス動作に着目しながらキレの特徴量を求める。この学習モデルを図 3 のように構成する。

1 層目の CNN では、速度(加速度)情報の xy 座標の 2 次元ベクトルについて時間軸方向と同時に畳み込みを行い、どの方向に動かしたかではなく骨格にどの程度の動きがあったかを考える。同様に 2 層目の CNN では、右(左)の骨格の速度(加速度)変化の関連性について考慮するため、肘、手首の 3 次元ベクトルを時間軸方向と同時に畳み込みを行う。

さらに CNN が出力したデータに LSTM を用いて全体のダンス動作の連続性について学習を行う。この学習には、前フレームだけでなく周囲のフレームをバランス良く見ることができる Bidirectional LSTM (BLSTM) を用いる。BLSTM では、LSTM に入力するデータを前から順に並べたデータと逆順に並べたデータ 2 つの出力を連結することで学習を行う。

以上のモデルを組み合わせることで、拍位置情報を考慮したキレの特徴抽出を行う。過学習抑制のため CNN の出力層に Dropout、全結合層の活性化関数 ReLU (Rectified Linear Unit) の前にバッチ正規化を用いる。

4.5 モデルの流れ

構成したモデルは 2 次元のベクトルを出力する。損失関数にはクロスエントロピーを用い、バリデーションロスが最小になったモデルを訓練済みモデルとする。このモデル全体の流れを以下の式 (7), (8), (9) に示す。

$$\mathbf{T}'_{(i)} = \text{Tempo_Extractor}(\mathbf{T}_{(i)}) \quad (7)$$

$$\mathbf{V}'_{(i)} = \text{Dance_Extractor}(\text{Add}(\mathbf{T}'_{(i)}, \mathbf{V}_{(i)})) \quad (8)$$

$$\mathbf{A}'_{(i)} = \text{Dance_Extractor}(\text{Add}(\mathbf{T}'_{(i)}, \mathbf{A}_{(i)})) \quad (9)$$

5. 採点システムの評価

実装したモデルに予め分割していたテストデータを入力し推論を行うことで、モデルの性能を評価する。この評価から実際にキレの特徴量が抽出できているか検討を行う。さらに、CNN と BLSTM が結果にどのように影響するか考察する。

5.1 精度検証

トレーニングデータを用いてパラメータの重みの学習を行い、バリデーションデータを用いてバッチサイズ (146)、エポック数 (200)、学習率 (5e-3) を定めた。訓練済みのモデルに対しテストデータを用いてキレの予測を行い、出力した結果をコメントによって定めた正解ラベルとの適合を行うことで正解率を求める。

正解率を表 2、表 3 に示す。拍位置ガイドありと書かれた行は $\text{Tempo_Extractor}(\mathbf{T}_{(i)})$ を用いて $\mathbf{V}'_{(i)}$ ($\mathbf{A}'_{(i)}$) を求めたもの、拍位置ガイドなしと書かれたものは $\text{Dance_Extractor}(\mathbf{V}_{(i)})$ ($\text{Dance_Extractor}(\mathbf{A}_{(i)})$) のみを用いたものである。また、列に並ぶ CNN+CNN と BLSTM は、Dance_Extractor からそれぞれ BLSTM、2 層の CNN を省くことで CNN と BLSTM が目的の特徴抽出を行なっているか確認するものである。そのため、列の CNN+CNN と LSTM は 2 層の CNN と 2 層の全結合層、1 層の BLSTM と 2 層の全結合層から構成されている。このときの Dropout、バッチ正規化はバリデーションロスが最も最小となるよう適応させた。

分類精度は、拍位置を考慮し 2 層の CNN と BLSTM を用いて構成したモデルが精度 85 % と最も高かった。さらにダンスの動作には速度データを用いる方が分類精度が高かった。

より高い精度が得られた、速度を入力とし 2 層の CNN と BLSTM を用いて作成したモデルの出力による ROC 曲線を、拍位置を考慮した場合を図 4、拍位置を考慮しない場合を図 5 に示す。縦軸にキレがあると予測し実際のコメントによってキレがあると定義されていた割合、横軸にキレがあると予測し実際にはキレがないと定義されていた割合を表している。このときの ROC 曲線の下部分の面積である AUC (Area under an ROC curve) の値は、それぞれ拍位置を考慮した場合で 0.9、考慮しない場合で 0.8 を超えた。

5.2 考察

表 2、表 3 の正解率を比較すると 2 層の CNN と BLSTM

表 2 正解率 (V'_i)

モデル	CNN+CNN	BLSTM	CNN+CNN+BLSTM
拍位置ガイドあり	64 %	51 %	85 %
拍位置ガイドなし	64 %	55 %	81 %

表 3 正解率 (A'_i)

モデル	CNN+CNN	BLSTM	CNN+CNN+BLSTM
拍位置ガイドあり	61 %	62 %	77 %
拍位置ガイドなし	64 %	59 %	75 %

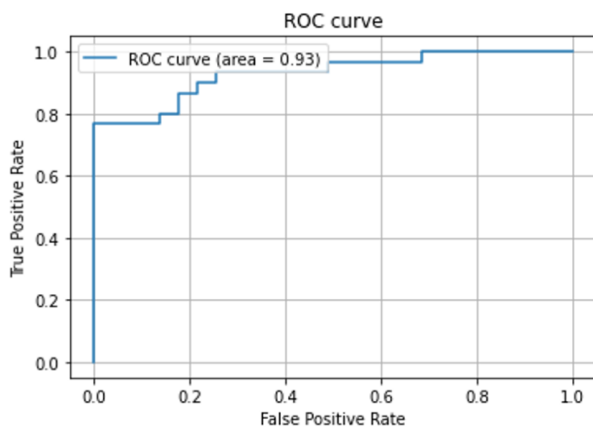


図 4 拍位置を考慮した速度による骨格情報の分類精度評価結果 (ROC 曲線)

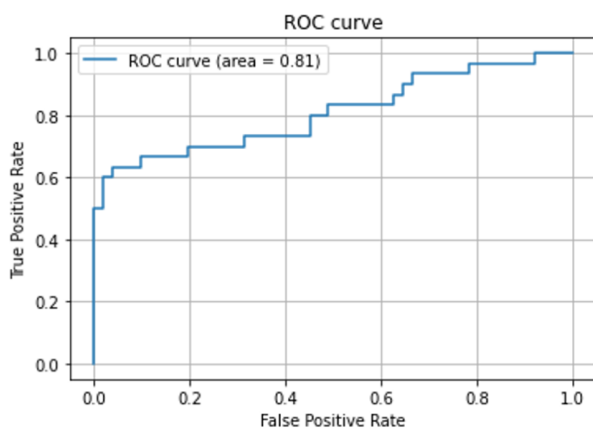


図 5 拍位置を考慮しない速度による骨格情報の分類精度評価結果 (ROC 曲線)

を用いることでキレの分類が行えることがわかる。1 層の BLSTM と全結合層を持つモデルによって学習を行なったところ 51 % まで精度が下がったことから、ダンスのような全体の流れだけでなく動作のポイントによるアクセント等を必要とする時系列データには、CNN が有効であることがわかる。2 つのモデルの比較からダンスの分析には全体の流れを把握するだけでなく、動きのポイントも抽出する必要があることが明らかとなった。また 2 層の CNN と全結合層を持つモデルで 64 % の正解率であったことから、CNN のみのモデルでは、十分な性能を持つことができないと言える。BLSTM の層から全体のダンス動作を把握す

ることで、様々な振付に対応できることがわかる。

さらに、拍位置を考慮した速度 (加速度) 情報を用いることで 2 層の CNN と BLSTM を用いたモデルの正解率が向上したことから、本研究の目的であった拍位置を考慮したキレの特徴量が Tempo_Extractor によって抽出可能であると言える。また速度情報を入力とし 2 層の CNN と BLSTM を用いて構成したモデルの結果について考察を行うと、拍位置を考慮しない場合で 0.8 程度であった AUC の値が、拍位置を考慮した場合 0.9 を超えたことからキレを分類するには拍位置を考慮することで十分なモデルとなることが明らかとなった。

一般的にスポーツ分野での姿勢推定を用いた分析では加速度情報を用いるものが多いが、ダンスのキレへの分析では速度情報を用いた方が分類精度は高かった。この結果からキレの特徴量は速度情報によって顕著に表れるのではないかと考える。また、誤判定してしまったダンス動画は xy 平面ではなく z 方向の奥行きで骨格の動きが見られたものが多く、OpenPose によって測ることができなかったことが原因だと考える。元々動きの小さい振付も同様に誤判定が見られた。よって実装したモデルは、動きの大きさや速さからキレの分析を行っていると考えられる。

6. まとめ

本研究では、姿勢推定によって取得したダンス動作からダンスのキレの特徴量を抽出可能であるか検討を行った。考慮できる時刻を制限し短期的な時系列を扱う 2 層の CNN と、長期的に記憶を保持することで時系列全体を見渡すことができる BLSTM を組み合わせることで、キレの分類が可能であった。また、楽曲の拍位置を学習に取り入れることでさらに高い精度でキレを分類できた。

しかし、本システムは恋ダンスのキレに限定した採点システムである。我々は、様々な楽曲に対応させたキレ採点を行うことで、人間が何を見てダンスの“キレ”を判別しているのか探ることができると考えている。ところが BPM158 であることを条件に様々なダンス楽曲を基とした採点システムを実装した際、十分な学習結果が得られなかった。この原因として、ピッチ変化による楽曲の雰囲気が考慮されていないことが挙げられる。

また、恋ダンスによってキレという評定用語はダンスを評価する上で必要な要素であった。しかし、バレエのように動作にキレを全く必要としないダンスも存在する。このようにシステムを構成する際、どの要素を必要とするかはダンスのジャンルによって異なる。そこでフィギュアスケートなど審美性スポーツでの採点を利用することで、ダンスジャンルにとって最適な要素を採点するシステムを実装できるのではないかと考える。

参考文献

- [1] 大塚製菓株式会社ポカリスエット:ダンス意識と経験調査, 入手先 (<https://kyodonewsprwire.jp/release/201711288452>) (2020.12.10).
- [2] Glossom 株式会社: スマートフォンでの情報収集に関する定点調査, 入手先 (<https://prtnews.jp/main/html/rd/p/000000007.000043425.html>) (2020.12.29).
- [3] 河野由, 小笠原一生, 水村 (久埜) 真由美: 白鳥の羽ばたきの表現を伴う上肢動作におけるバレエ熟練者の運動学的特徴, 日本バイオメカニズム学会誌, Vol. 41, No. 4, pp. 50-54 (2017).
- [4] 古市冨佳, 阿部和樹, 中村総史ほか: ヒップホップダンスにおける骨格情報を用いた個性抽出の検討, 研究報告エンタテインメントコンピューティング (EC), Vol. 2018, No. 23, pp. 1-9 (2018).
- [5] 平澤直之, 清水大地: 深層学習を利用したブレイクダンスにおける新しい価値の創出・評価システムの開発, 人工知能学会全国大会論文集 一般社団法人人工知能学会, pp. 2C1J1203-2C1J1203. 一般社団法人人工知能学会 (2019).
- [6] Shuhei Tsuchida, Satoru Fukayama, Masahiro Hamasaki, and Masataka Goto: *AIST Dance Video Database: Multi-Genre, Multi-Dancer, and Multi-Camera Database for Dance Information Processing*, In ISMIR, pp. 501-510 (2019).
- [7] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. : *OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields*, CoRR, Vol. abs/1812.08008 (2018).
- [8] Taoran Tang, Jia Jia, and Hanyang Mao: *Dance with melody: An LSTM-autoencoder approach to music-oriented dance synthesis*, In Proceedings of the 26th ACM international conference on Multimedia, pp. 1598-1606 (2018).
- [9] 逃げるは恥だが役に立つ: TBS テレビ 2016年10月11日 - 12月20日放送, 入手先 (https://www.tbs.co.jp/NIGEHAJI_tbs/) (2020.07.28).
- [10] YouTube: 入手先 (<https://www.youtube.com>) (2020.07.28).
- [11] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto: *librosa: Audio and music signal analysis in python*, In Proceedings of the 14th python in science conference, Vol. 8, pp. 18-25 (2015).