

MachineDancing: ポーズの変化に関する大局的制約と音楽と動作の局所的制約を同時に考慮したダンス自動生成

深山 覚^{1,a)} 後藤 真孝^{1,b)}

概要: 本稿では音楽に連動した3次元コンピュータグラフィックスによるキャラクターのダンス動作を自動生成する手法 MachineDancing を議論する。音楽に連動したダンス動作を生成するには、時間変化する音楽の特徴に応じてダンス動作を作り分ける必要がある。そこで従来の研究では、拍や小節単位でのダンス動作の素片を準備し、それらを適切に接続する方法が提案されてきた。この方法は音楽と動作の時間的に局所的な制約に着目したダンス自動生成手法であった。この方法には、楽曲中の繰り返しや曲の構造に基づいた時間的に大局的な制約を満たすダンス動作を生成しにくいという課題があった。そこで本稿では、この局所的制約と大局的制約をともに満たすようなダンス動作を生成する方法を検討する。その方法として、曲構造駆動型と概略姿勢駆動型の2種類の方式を提案し、それぞれがどのような定式化でダンス動作を生成できるかについて議論する。曲構造駆動型については、実際に楽曲の構造に基づいたダンス動作を生成することができた。また概略姿勢駆動型については、そのダンス生成が出力合流並列隠れマルコフモデルで定式化されることがわかり、その解法として双対分解に基づいた最適化に基づく方法を検討した。

1. はじめに

本研究では、3次元コンピュータグラフィックスによるキャラクターのアニメーションのために、音楽に連動したダンス動作を自動生成する方法 MachineDancing について議論する。MachineDancing の概要を図1に示す。MachineDancing は楽曲に同期したダンス動作のデータベースをもとに、動作と音楽の関係性についての確率モデルを学習し、任意の楽曲の入力に対してダンスを自動生成する方法である。他の方法としては、数小節単位のダンス動作の断片をあらかじめ準備し、それらを繋ぎ合わせるものがある。この方法はクオリティの高い動作が生成できる利点を持つが、楽曲の内容に応じて適応的に多様な動作を生成するには適していなかった。筆者らがこれまで取り組んできた MachineDancing [1] では、より細かな時間単位で動作を生成し、さらにはデータベースには含まれていない動作や姿勢を生成する方法によって、楽曲に適合する多様なダンス動作の生成を目指している。

ダンス自動生成研究の意義は、ダンスの振り付けという人間の創作活動がどこまで計算機処理によって自動化できるかを科学的に検証することで、改めてダンス振り付けの

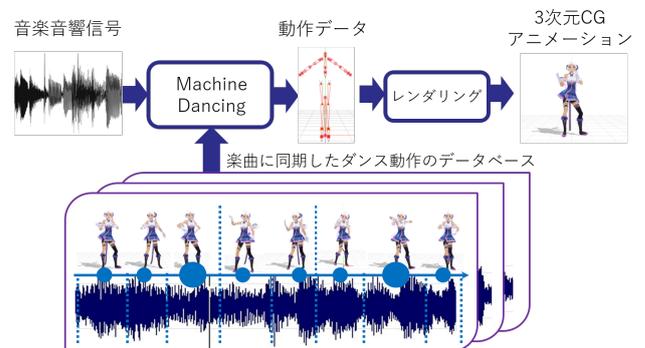


図1 MachineDancing の概要: 音楽と同期したダンス動作のデータから学習し、任意の音楽に対してダンス動作生成を行い、その結果を3次元CGアニメーションで鑑賞できる

ノウハウや人間のダンスに対する理解について知見を深められることにある。またダンスの自動生成エンジンが実現することで、人々が簡単に自身の好きな楽曲に振り付けをし、音楽をダンスとともに多角的に鑑賞できる。さらには、振り付けをする人の有用なアシスタントとしてダンスの自動生成エンジンを使うことで、データから得られるダンスのノウハウを活用しながら振り付けができると期待される。

楽曲に連動したダンス動作を生成するにあたっては、動作を生成する時間単位をどのように決めるのかが問題である。数小節単位の動作素片を準備しておく方法では、ステップを踏むダンスや空間的に大きく移動するダンスをテ

¹ 産業技術総合研究所
305-0058 茨城県つくば市梅園 1-1-1
a) s.fukayama@aist.go.jp
b) m.goto@aist.go.jp

ンプレートとして用いることができる反面、時間変化する音楽の特徴に応じて動作を変化させることができなかった。一方で1拍や1小節単位で動作を生成すると、拍単位での音響特徴量などを反映して動作を変化させることができるが、楽曲構造や小節単位での動作の繰り返しといった時間方向に構成することが不得手であった。このように、動作を生成する時間単位の設定に依存して、局所的制約と大域的制約を同時に満たす動作を生成できない、というのが課題であった。

そこでMachineDancingでは、それらを同時に満たす方法として「曲構造駆動型」と「概略姿勢駆動型」の2種類の方法を提案する。曲構造駆動型では楽曲の構造解析結果に基づく楽曲を通じた大域的制約のもとで、半拍ごとの拍の強さに応じて動作の俊敏さを変えるという局所的制約を満たすようにダンス動作を生成する。概略姿勢駆動型においては、あらかじめ楽曲を通じたダンサーの大まかな姿勢(概略姿勢)を指定しておくことで、それを大域的な制約としつつ、さらに全体を通じて同時に半拍ごとの拍の強度に応じた動作の俊敏さを変える局所的な制約を満たすようにダンス動作を生成する。

曲構造駆動型の技術的課題は、半拍ごとの拍の強度の推定方法、ダンス動作生成のための楽曲構造解析、そしてその構造を反映するダンス生成法の3点である。拍の強度は、楽曲の自動ビート・小節線推定の結果に基づいて、半拍ごとのスペクトルフラックスの総和を三角窓を用いて算出する。楽曲構造解析では、小節単位の拍の強さの系列をパターンと捉え、それらの教師なしクラスタリングを行い、楽曲中の小節を拍の強さの系列のパターンとして分類する。そしてダンス生成では、楽曲中の小節の分類結果に応じて、拍の強さと動作の俊敏さが一致するように半拍単位でダンス動作を隠れマルコフモデルの定式化に基づいて自動生成する。

概略姿勢駆動型では、手動入力や赤外線デプスセンサKinectなどを用いて取得される概略姿勢に類似させながらも、半拍ごとの拍の強さに従って俊敏さが変化する動作を生成する。これをいかに実現するかが技術的課題である。入力された概略姿勢に類似した動作を生成するには、データベースに含まれる姿勢をそのまま活用するのでは類似した姿勢が生成できないため、データベース中の姿勢を上半身と下半身に分けて、その組み合わせによって生成する必要がある。そのため概略姿勢駆動型は、上半身と下半身という2つの動作系列を互いの依存関係を考慮しながら最適化する問題であり、これは時間変化する2つの隠れ変数を持つ隠れマルコフモデルによる定式化に帰着される。そのような隠れマルコフモデルによる問題解法は特定の条件でのみしか知られていないが、本稿では双対分解を用いた新たな解法を議論する。

2. 関連研究

2.1 局所的制約に基づいたダンス生成

先行研究では、音楽に連動するダンスを生成するために、どのような時間的に局所的な制約を用いてダンスを生成すればよいか議論がされてきた。たとえばテンポ [2], [3], 拍のタイミング [4], [5], [6], 音響特徴量の組合せ [7] やピッチ情報 [8] などが用いられて、ダンスが自動生成されている。

2.2 大域的制約を取り入れたダンス生成

大域的な制約条件を取り入れたダンス生成の先行研究では、楽曲の構造を分析フレーム毎の音響特徴量の類似度行列によって解析し、それに基づいてダンスを生成する研究がある [9]。また数拍に渡るやや大域的な制約を用いるものでは、メロディの概形の類似度に基づいて1小節間のダンス動作テンプレートを接続する研究がある [10]。

3. 局所的制約に基づくダンスの自動生成

3.1 ダンス動作のデータ構造

CGによるダンス動作は、キャラクターの骨格に相当するリグの時間的な移動と回転によって表現される。この移動と回転を音楽に連動させて適切に設定することで、CGキャラクターのダンスを生成することができる。本研究では、そのためのデータ形式として、CGキャラクターの動作の制作支援環境であるMikuMikuDance^{*1}で用いられる形式であるvmd形式を用いる。

リグに含まれる(疑似的な)骨格それぞれの動作は、グローバル座標の xyz 方向の位置を表す3成分と、関節の接続点を原点とするローカル座標での回転を表すクォータニオン4成分によって表される。ただしここでのクォータニオンは原点から単位球上の点を指すベクトル間の回転のみを表すために、ノルムが常に1である。これら7成分の値を、今回用いるリグでのすべての骨格(ボーン)について集めたベクトル \mathbf{x} はダンサーの姿勢を表す。本研究では、これを姿勢ベクトルと呼ぶこととする。

CGによるダンス動作は、フレームと呼ばれる短い時間単位における静的な姿勢の連続によって表現できる。本研究での1秒あたりのフレーム数は30フレームである。原理的には、1フレーム毎の姿勢をすべて生成することによってダンス動作を生成できる。しかし身体動作は元来時間・空間的に連続であり、またダンスにおいては半拍単位の姿勢の組合せでダンス動作を説明できることが多いことから、本研究では半拍ごとの姿勢のみ生成し、半拍の時刻以外のフレームについては、レンダリングの際に隣合う半拍の姿勢間の線形補間によって生成される。

^{*1} <http://www.geocities.jp/higuchuu4/>

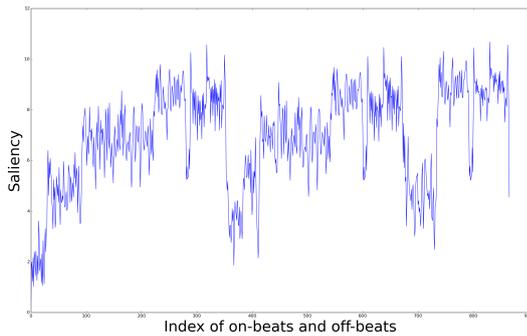


図 2 推定された表拍と裏拍における拍の強さの例 (横軸: 半拍単位の時刻インデックス, 縦軸: 拍の強さ)

3.2 拍の時刻とダンス動作の関係

拍はダンスをする際に重要な音楽要素である。これはダンスの振り付けを例示する時に「ワン・トゥー・スリー」などとカウントをし拍を示すことからわかる。さらに「ワン and トゥー and」といったように、裏拍を明示的にカウントすることもある。したがって、表拍に加えて裏拍もダンスにおいて重要な音楽要素であることがわかる。実際に、ヒップホップダンスで拍にのって動作するときのパターンに「ダウン」と「アップ」があり、表拍と裏拍のどちらで足を曲げるかの差でそれらが区別される。このことも、表拍のみでなく裏拍をも考慮する必要性を示唆している。

音楽音響信号中の拍の時刻はビート推定技術を活用して得ることができる。本研究では再帰的ニューラルネットワークと Dynamic Bayesian Network を併用したビートと小節線の推定手法 [11] を、音楽音響信号処理ライブラリ Madmom [12] を使って実行した。さらに、推定された隣合う 2 つの表拍の時間的に中央に位置する時刻を裏拍の時刻とした。

3.3 拍の強さとダンス動作の関係

表拍と裏拍の時刻とともに重要と考えられるのが、拍の強さである。強さの異なる拍が連なるとリズムが形成される。既存研究 [13] において、音楽のリズムとダンサーの動作のリズム (Motion Rhythm Feature) がよく一致することが指摘されている。また、ダンサーの動作のリズムは、関節の角運動量の総和の時系列が局所的に最小値となる点が連なることで生じるとしている。

拍の強さは、拍の時刻における楽音の大きさというよりも、楽音のアタックの強さであると考えられる。これは同じ音を大きな音量で鳴らし続けてもリズムが生じないことからわかる。そこで音楽音響信号のパワースペクトログラムから音の立ち上がりについての音響特徴量であるスペクトルフラックスを計算し、表拍と裏拍が位置する時刻でのスペクトルフラックスの総和を計算する。計算の際には、総和を計算したい時刻で 1.0 の値を持ち、その両隣の拍で

0.0 の値をとる三角窓を乗じて計算した。ある 1 曲を通じて計算した拍の強さを、横軸を半拍ごとの時刻インデックス、縦軸を拍の強さとして図 2 に示す。曲の構造に応じて値の平均値が変化している様子がわかる。

3.4 音楽に連動するダンス動作の機械学習

ダンス動作を構成する多様な姿勢と、拍の強さと姿勢の関係の双方を、音楽に同期したダンス動作のデータから機械学習することを考える。

はじめに、ダンス動作を構成する姿勢は、拍の時刻での姿勢を収集・分類することで学習できる。音楽に同期したダンス動作のデータベースにおいて、ビート推定結果に基づいた表拍と裏拍の時刻を用いて、ダンス動作中の姿勢ベクトルを半拍ごとに抽出できる。抽出された姿勢ベクトルからダンス生成に用いる姿勢ベクトルを選別するには、手でダンスに用いたい姿勢を選別するか、Mini-batch k-means クラスタリング [14] によって似た姿勢同士のクラスを形成し、そのクラスセントロイドを用いればよい。

次に、拍の強さと姿勢の関係は、ベイズ線形回帰によって確率モデル化できる。拍の強さと関係すると考えられるダンサーの動作のリズムは、関節の角運動量の時系列が局所的に最小値となることで生じるとされているため [13]、拍の強さが大きい箇所では、関節の瞬間的なトルクが大きくなっていると期待される。瞬間的なトルクは、隣接する 3 つの拍での姿勢から計算することができる。そのトルクと拍の強さの値をもとに、ベイズ線形回帰を行うと、拍の強さと隣り合う 3 つ組の姿勢の関係についての確率モデルを学習できる。

3.5 ダンス動作自動生成の定式化

これまでの議論をもとに、局所的な制約に基づいたダンス動作自動生成の定式化を行う。いま表拍と裏拍による半拍刻みの時刻を $t = 1, \dots, T$ とし、時刻 t における姿勢ベクトルを \mathbf{x}_t 、拍の強さを s_t とする。

ここでダンスの振り付けがされやすい音楽の拍の強さの系列は、人間に可能な身体運動によるダンス動作から連想しやすいものと仮定する。なぜなら、前節でみたように、ダンス動作の瞬間的なトルクと拍の強さの間には関係がある上、人間の動作から全く連想されないようなリズムを持つ音楽は、それに対して振り付けをするのは困難であると考えられるからである。

したがってダンスの振り付けがされやすい音楽の拍の強さの系列には、身体運動としてあり得るダンス動作に由来して偏りが生じている。この偏りを確率によってあらわすには、あらゆるダンス動作を想定した上で、そのもとでどのような拍の強さの系列が現れるかを考えればよい。いま、ダンス動作を姿勢ベクトルの系列 $\mathbf{x}_{1:T} = \mathbf{x}_1 \dots \mathbf{x}_T$ として表し、それらダンス動作から連想されるような拍の強

さの系列を $s_{1:T} = s_1 \dots s_T$ と書く。このとき拍の強さの系列についての出現の偏りは、

$$P(s_{1:T}) = \sum_{\mathbf{x}_{1:T}} P(s_{1:T}|\mathbf{x}_{1:T}) P(\mathbf{x}_{1:T}) \quad (1)$$

と書ける。

ダンスの振り付けとは、音楽を聴いたときに、その音楽のリズムを生み出すような、最もふさわしいダンス動作を決めることだと考える。その場合、ダンスを自動生成するには、音楽の拍の強さの系列 $s_{1:T}$ が与えられた下で、最も尤もらしい姿勢ベクトルの系列 $\mathbf{x}_{1:T}^*$ を求めればよい。この姿勢ベクトル系列の尤もらしさは、 $s_{1:T}$ が与えられたときの $\mathbf{x}_{1:T}$ についての確率 $P(\mathbf{x}_{1:T}|s_{1:T})$ として評価できる。よってダンスの自動生成は

$$\mathbf{x}_{1:T}^* = \operatorname{argmax}_{\mathbf{x}_{1:T}} P(\mathbf{x}_{1:T}|s_{1:T}) \quad (2)$$

と定式化できる。ベイズの定理を用いると

$$P(\mathbf{x}_{1:T}|s_{1:T}) = \frac{P(s_{1:T}|\mathbf{x}_{1:T}) P(\mathbf{x}_{1:T})}{P(s_{1:T})} \quad (3)$$

と変形でき、このとき式 (3) 右辺の分母は $\mathbf{x}_{1:T}$ に依らない。よって

$$\mathbf{x}_{1:T}^* = \operatorname{argmax}_{\mathbf{x}_{1:T}} P(s_{1:T}|\mathbf{x}_{1:T}) P(\mathbf{x}_{1:T}) \quad (4)$$

を解けばよい。これは隠れマルコフモデルの復号化問題である。

さて s_t が時刻 t における動作の瞬間的なトルクと関係があること、トルクが両隣を含めた合計3点の姿勢から計算できることに注意すると、式 (3) の分子は、

$$P(\mathbf{x}_1) \prod_{t=2}^T P(s_t|\mathbf{x}_{t+1}, \mathbf{x}_t, \mathbf{x}_{t-1}) P(\mathbf{x}_t|\mathbf{x}_{t-1}) \quad (5)$$

と変形できる。 $P(\mathbf{x}_t|\mathbf{x}_{t-1})$ の項は隣合う姿勢の遷移確率である。

本定式化では、姿勢の系列を Viterbi アルゴリズムを用いて効率的に求めるために、工夫が必要である。3つの姿勢ベクトルの組合せを $P(s_t|\mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{x}_{t+1})$ の項で考慮する必要があり、学習データから抽出された姿勢の総数を N とすると、通常の Viterbi アルゴリズムの計算量が $O(N^2T)$ なのに対して $O(N^6T)$ になってしまうからである。そこで、Viterbi アルゴリズムの前向き計算の際に、各時刻 t の時点での確率値のランキング上位100位までの姿勢だけ用いて計算する枝刈りを行うことで、計算量を削減した。

4. 大局的制約を考慮したダンスの自動生成

4.1 曲構造駆動型

4.1.1 曲構造駆動型の概要

曲構造駆動型では、楽曲の構造に基づいてダンス動作を生成する。楽曲には似たセクションの繰り返しや、曲調が

盛り上がる部分がある。それに応じてダンス動作にも繰り返しや盛り上がりに応じた動作が生成できれば、大局的な制約を反映したダンスにできると考えられる。

楽曲構造と盛り上がりの解析は、拍の強さの系列を小節ごとに区切ったものをクラスタリングすることで行える。本研究では Mini-batch k-means クラスタリング [14] を用い、クラスタ数は10とした。クラスタそれぞれについて、クラスタ平均を拍の強さによる局所的制約として用い、ダンス動作を生成する。最後に、小節ごとに割り当てられたクラスタに従って生成されたダンスを接続する。

4.1.2 曲構造駆動型による生成結果と考察

この方式で生成したダンスを観察し考察を行った。楽曲のサビ (chorus) に対応する箇所では腕や足が大きく動き、サビの直前で一時静かになる部分などでは動きが小さくなる様子が見られた。また、楽曲中の1回目のサビと2回目のサビで同じ動作が使われ、楽曲構成を反映したダンスが生成された。

一方、小節と小節のつなぎ目については、急な姿勢の変化や足が滑る動きなどの動きがみられた。小節同士の接続部分が考慮できていないため、特に空間的移動を多くするダンスが小節内で生成されている場合に、接続部分で身体運動としては不自然な動きがみられる問題がわかった。

4.2 概略姿勢駆動型

4.2.1 概略姿勢駆動型の概要

概略姿勢駆動型では、あらかじめダンサーに取らせた姿勢の概略 (以後、概略姿勢と呼ぶ) を入力し、その概略姿勢に似たダンスを生成するという制約の下で、局所的制約である拍の強さを考慮しながらダンスを生成する方法である。概略姿勢の入力方法としては、特に姿勢を指定したいフレームをキーフレームとして姿勢を手動で指定する方法と、赤外線デプスセンサである Kinect などによる姿勢推定結果や、モーションキャプチャシステムを用いて、あらかじめ音楽に同期して関節の位置・角度情報を収録しておき、それを入力とする方法が考えられる。

4.2.2 概略姿勢駆動型の課題

拍の強さという局所的制約に加えて、概略姿勢に似ているという制約を新たに考慮しながらダンスを生成するには、どのようにしたらよのだろうか。これは式 (4) で示した隠れマルコフモデルによる定式化において、拍の強さについての確率を、拍の強さと概略姿勢についての同時確率にすることで実現できる。具体的には、時刻 t の概略姿勢の姿勢ベクトルを $\hat{\mathbf{x}}_t$ とし、式 (4) を

$$\mathbf{x}_{1:T}^* = \operatorname{argmax}_{\mathbf{x}_{1:T}} P(s_{1:T}, \hat{\mathbf{x}}_{1:T}|\mathbf{x}_{1:T}) P(\mathbf{x}_{1:T}) \quad (6)$$

へと変更し、解を Viterbi アルゴリズムで求めればよい。

この方法では、ダンス動作データベースから抽出できる姿勢の個数が少ない場合、概略姿勢に近い動作を実際に生

成できるか、という課題がある。そこで1つの姿勢ベクトルを上半身と下半身のベクトルへと分割し、元の上半身と下半身の組合せとは異なる組合せによる姿勢も生成できるようにすると、概略姿勢に近い姿勢も生成できるようになると考えられる。

姿勢ベクトル \mathbf{x} を上半身と下半身のそれぞれに関する移動と回転を表す値へと分割したベクトルを \mathbf{x}^U と \mathbf{x}^L とする。このとき式 (6) 右辺は

$$P(s_{1:T}, \hat{\mathbf{x}}_{1:T} | \mathbf{x}_{1:T}^U, \mathbf{x}_{1:T}^L) P(\mathbf{x}_{1:T}^U, \mathbf{x}_{1:T}^L) \quad (7)$$

となる。このような隠れ変数の系列が2つあり、出現確率がそれら2つの隠れ変数を条件とする確率で表現される確率モデルは、出力合流並列隠れマルコフモデルと呼ばれ、多声鍵盤音楽の楽譜追跡や両手部分離に用いられている [15], [16]。このモデルでは尤度最大の姿勢の系列を求める際に、二つの系列 \mathbf{x}^U と \mathbf{x}^L について最適化を行う必要がある。拍の強さは上半身と下半身の動作の瞬間的なトルクの和と関係があるため、上半身と下半身の系列は互いに依存関係にあり、同時に最適化を行う必要がある。

4.2.3 双対分解を用いた概略姿勢駆動型の解法

そこで本研究ではその同時最適化を行うために、双対分解を用いた最適化法を検討する。双対分解を用いた最適化とは、ある関数 $h(x)$ の x についての最大化（もしくは最小化）が困難な場合でも、最適化が容易な2つの関数 $f(x)$ と $g(x)$ の和 $h(x) = f(x) + g(x)$ へと分解可能であれば、もとの最適化問題の双対問題を解くことによって、解が求まる方法である。

双対分解による最適化の概要は以下の通りである。まず変数 y とラグランジュ乗数 u を導入して $h(x)$ を変形した $\hat{h}(x, y)$ をつくる。

$$\hat{h}(x, y, u) = f(x) + g(y) + u(x - y) \quad (8)$$

ここで第3項は変数 x と y が同じ値に近づくような制約項である。すると、以下の二つの問題は双対な関係にある。

- (主問題) $\max_x h(x)$
- (双対問題) $\min_u (\max_{x,y} \hat{h}(x, y, u))$

このとき

$$\hat{h}(x, y, u) = \{f(x) + ux\} + \{g(y) - uy\} \quad (9)$$

のように x と y それぞれのみ依存する形に整理できることから、 x と y についてそれぞれ最適化をして

$$x^* = \arg \max_x \{f(x) + ux\} \quad (10)$$

$$y^* = \arg \max_y \{g(y) - uy\} \quad (11)$$

$$\max_{x,y} \hat{h}(x, y, u) = f(x^*) + g(y^*) + u(x^* - y^*) \quad (12)$$

のように計算できる。

このうえで u について最小化することができれば、元の

最適化問題を解くことができる。 u についての微分は必ずしも計算できるとは限らない。しかし u についての劣微分の一つが $x^* - y^*$ が含まれることがわかる。そこで学習率 α を導入して

$$u := u - \alpha(x^* - y^*) \quad (13)$$

に従って反復的に u を更新することで、 u を求めることができる。

ここで概略姿勢駆動型の問題に立ち戻る。計算を困難にしているのは互いに依存する2つの系列を最適化する必要があるからである。そこで以下のように、上半身の系列と下半身の系列をそれぞれ最適化する問題へと分割する。

- 上半身を概略姿勢にもっとも近い姿勢に決め、下半身の姿勢を拍の強さを制約として最適化
- 下半身を概略姿勢にもっとも近い姿勢に決め、上半身の姿勢を拍の強さを制約として最適化

それぞれの問題については、Viterbi アルゴリズムで最適化する系列が1つだけなので、効率的な最適化が可能である。上記の2つを別々に最適化し、 u を更新する、という2つのステップを繰り返すことで、解を求めることができると考えられる。

4.2.4 シミュレーション実験結果

双対分解を使った解法によって、実際に出力合流並列隠れマルコフモデルの解探索が可能かどうかを、シミュレーションによって検証した。いま、隠れ変数として $z^{(1)}$ と $z^{(2)}$ の2つを考え、時刻 t における変数は時刻 $t-1$ に確率的に依存しており、その依存関係が2つの遷移確率 $p_1(z_t^{(1)} | z_{t-1}^{(1)})$ と $p_2(z_t^{(2)} | z_{t-1}^{(2)})$ によって決まっているとする。隠れ変数から出力され観測できるものとして q_t, r_t, s_t があり、それぞれ出力確率 $p_q(q_t | z_t^{(1)})$, $p_r(r_t | z_t^{(2)})$, $p_s(s_t | z_t^{(1)}, z_t^{(2)})$ によって出力されるとする。概略姿勢駆動型の問題との対応関係としては、 q_t が上半身の概略姿勢、 r_t が下半身の概略姿勢、 s_t が拍の強さに対応する。

本シミュレーションでは簡単のため $z^{(1)}$ と $z^{(2)}$ を離散的な10状態 $z^{(1)}, z^{(2)} \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ とし、遷移確率 p_1, p_2 は 10×10 のランダム行列を各行について正規化したものとした。系列長は50とし、 $q_{1:50}$ と $r_{1:50}$ は各時刻について独立に一様分布からサンプルし、 $s_{1:50}$ は $s_t = q_t + r_t$ となるように決めた。出力確率については、

$$p_q(q_t | z_t^{(1)}) = \mathcal{N}(z_t^{(1)}, 1) \quad (14)$$

$$p_r(r_t | z_t^{(2)}) = \mathcal{N}(z_t^{(2)}, 1) \quad (15)$$

$$p_s(s_t | z_t^{(1)}, z_t^{(2)}) = \mathcal{N}(z_t^{(1)} + z_t^{(2)}, 1) \quad (16)$$

のように設定した。反復回数は100回、学習率は $\alpha = 0.1$ とした。

シミュレーションの結果、 $z^{(1)}$ についてのみ Viterbi アルゴリズムを実行する場合の尤度と、 $z^{(2)}$ についてのみ

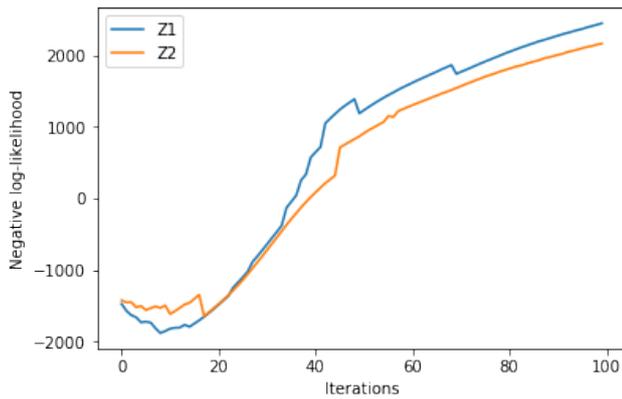


図 3 $z^{(1)}$ を系列として扱った Viterbi アルゴリズムの負の対数尤度と、 $z^{(2)}$ を系列として扱った Viterbi アルゴリズムの負の対数尤度の、反復回数が増えるにあたっての挙動

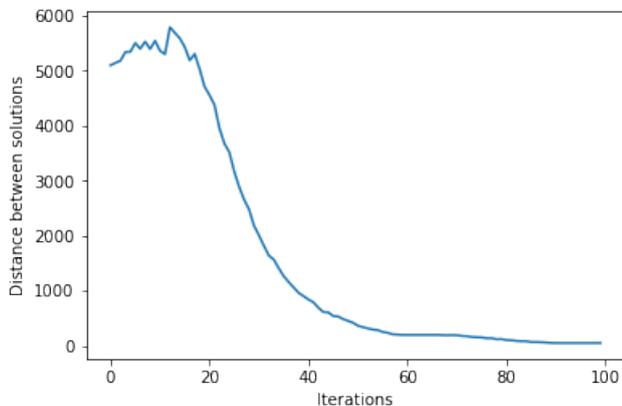


図 4 分解されたそれぞれの最適化での解同士の距離

Viterbi アルゴリズムを実行する場合の尤度は、反復回数が増えるにしたがってともに増加した。図 3 にその様子を示す。また双対分解によって分解されたそれぞれの尤度関数の最適化によって得られる $z^{(1)}, z^{(2)}$ の値の距離（ユークリッド距離）も、反復回数が増えるにしたがって減少し 0 に近い値へと収束した。距離が減少する様子を図 4 に示す。反復の結果、解が近い値へと収束していくことがわかる。今後姿勢ベクトルの値と、ダンス動作のデータベースから計算される確率に基づいて、双対分解を用いたダンスの自動生成に取り組む予定である。

5. おわりに

ポーズの変化に関する大域的制約と音楽と動作の局所的制約を同時に考慮できるダンス自動生成手法 MachineDancing について議論を行った。大域的制約の反映の仕方として、曲構造駆動型と概略姿勢駆動型の 2 種類を提案した。曲構造駆動型については実際に楽曲構造を反映しつつ、半拍単位の拍の強さという局所的制約をも同時に満たすダンスを自動生成することができた。概略姿勢駆動型については、大域的制約と局所的制約を同時に満たす問題が、出力

合流並列隠れマルコフモデルの定式化に帰着されることを示し、双対分解を用いたその解法について議論した。

謝辞 本研究の一部は JST ACCEL (JPMJAC1602) の支援を受けた。

参考文献

- [1] 深山覚, 後藤真孝: MachineDancing: ダンス動作データの自動分析に基づく音楽に連動したダンス生成手法, 情報処理学会研究報告, Vol. 2014, No. 14, pp. 1-7 (2014).
- [2] Chen, K. M., Shen, S. T. and Prior, S. D.: Using music and motion analysis to construct 3D animations and visualisations, *Digital Creativity*, Vol. 19, No. 2 (2008).
- [3] Panagiotakis, C., Holzapfel, A., Michel, D. and Argyros, A. A.: Beat Synchronous Dance Animation based on Visual Analysis of Human Motion and Audio Analysis of Music Tempo, *Proceedings of ISVC 2013*, pp. 118-127 (2013).
- [4] Shiratori, T., Nakazawa, A. and Ikeuchi, K.: Synthesizing dance performance using musical and motion features, *Proceedings of ICRA 2006*, pp. 3654-3659 (2006).
- [5] Kim, J. W., Fouad, H., Sibert, J. L. and Hahn, J. K.: Perceptually motivated automatic dance motion generation for music, *Computer Animation and Virtual Worlds 2009*, Vol. 20, pp. 375-384 (2009).
- [6] Alankus, G., Bayazit, A. A. and Bayazit, O. B.: *Computer Animation and Virtual Worlds*, No. 16, pp. 259-271 (2005).
- [7] Fan, R., Xu, S. and Geng, W.: Example-Based Automatic Music-Driven Conventional Dance Motion Synthesis, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 18, No. 3 (2012).
- [8] Ofii, F., Erzin, E., Yemez, Y. and Tekalp, A. M.: Learn2Dance: Learning Statistical Music-to-Dance Mappings for Choreography Synthesis, *IEEE Transactions on Multimedia*, Vol. 14, No. 3 (2012).
- [9] Lee, M., Lee, L. and Park, J.: Music similarity-based approach to generating dance motion sequence, *Multimedia Tools and Applications*, Vol. 62, No. 3, pp. 895-912 (2013).
- [10] Oore, S. and Akiyama, Y.: Learning to Synthesize Arm Motion to Music By Example, *Proceedings of WSCG 2006*, pp. 201-208 (2006).
- [11] Böck, S., Krebs, F. and Widmer, G.: Joint Beat and Downbeat Tracking with Recurrent Neural Networks, *Proc. ISMIR, ISMIR '16* (2016).
- [12] Böck, S., Korzeniowski, F., Schlüter, J., Krebs, F. and Widmer, G.: madmom: a new Python Audio and Music Signal Processing Library, *Proceedings of the 24th ACM International Conference on Multimedia*, Amsterdam, The Netherlands, pp. 1174-1178 (online), DOI: 10.1145/2964284.2973795 (2016).
- [13] Shiratori, T.: Synthesis of Dance Performance Based on Analyses of Human Motion and Music, *Doctoral Dissertation* (2006).
- [14] Sculley, D.: Web-Scale K-Means Clustering, *WWW2010*, pp. 1-2 (2010).
- [15] 中村栄太, 齋藤康之, 嵯峨山茂樹: 出力合流並列隠れマルコフモデルとその多声鍵盤音楽の楽譜追跡・両手部分離への応用, 情報処理学会研究報告, Vol. 2013, pp. 1-6 (2013).
- [16] 中村栄太, 小野順貴, 嵯峨山茂樹: 出力合流隠れマルコフモデルに基づく多声部音楽のリズム採譜, 情報処理学会研究報告, Vol. 2014, No. 8, pp. 1-7 (2014).