

動作データからのモチーフの発見について

Motif Discovery Algorithm from Motion Data

上原邦昭 田中良季 巻尾晃輔
Kuniaki Uehara Yoshiki Tanaka Kosuke Makio

神戸大学大学院自然科学研究科
Graduate School of Science and Technology, Kobe University

Detection of a previously unknown, frequently occurring pattern has been attracted much attention. Such pattern is called “*motif*”. Many researchers have proposed algorithms for discovering the motif. These algorithms extract the motif that has user-defined length of it. However, if the optimal length of the motif is not known in advance, we cannot use these algorithms for discovering the motif. In this paper, we attempt to dynamically determine the length of the motif using the MDL principle. Moreover, our algorithm can extract the motif from the multi-dimensional time-series by using Principal Component Analysis. Finally, we show experimental results and discuss the efficiency of our motif discovery algorithm.

1. はじめに

人間の身体運動計測は、映像産業での CG データ作成、スポーツ科学でのフォーム解析、手話のジェスチャ認識、医療での病気や障害の診断など、様々な分野で応用が進められている。中でも、人間の動作解析は注目すべき研究分野として、めざましい進歩を遂げている。動作解析には大別して2つのアプローチがある。ひとつは、時系列画像をもとに人間の動作を解析する、画像処理のアプローチである。もうひとつは、体表面にマークを取りつけ、専用のハードウェア（以下モーションキャプチャと呼ぶ）によってマークを追跡して、人間の動作を計測、解析するアプローチである。

画像処理による動作解析では、マークなどの接触型センサを人間に装着させる必要はないが、動画から特徴点を抽出し、事前に作成したモデルと入力データをマッチングさせる必要がある。このモデルについては、人物の対象部分を3次元形状モデル (3D Model) で表現する場合と、画像中の人物の対象部分の見え方そのものから抽出した特徴をモデル (View-based Model) とする場合がある。いずれの場合も、画像からのモデルの特徴が正確に抽出されないと認識できないため、画像から位置情報を抽出する手法が問題となっている。

一方、体表面にマークを取りつけ、モーションキャプチャによってマークを追跡して、動作解析を行うアプローチがある。モーションキャプチャによって得られる動作データには、各身体部位の3次元座標位置情報が含まれているために、特徴点を抽出する必要がなく、画像処理のアプローチと比べて計算コストが小さくてすむという利点がある。また、画像処理では分析できないような細かな動きを分析することができる。このため、モーションキャプチャによるアプローチを用いている。本研究で用いるシステムは、6台のカメラに囲まれた領域内で、身体に18個のマークを付けた演技者(図1(a))が行う動作からデータを取得している。各カメラから演技者の動作をビデオイメージとして記録し、そのイメージからマークの3次元座標を計算することにより、図1(b)に示すような、人間の各部位の軌道をあらわす、3次元時系列データが得られる。

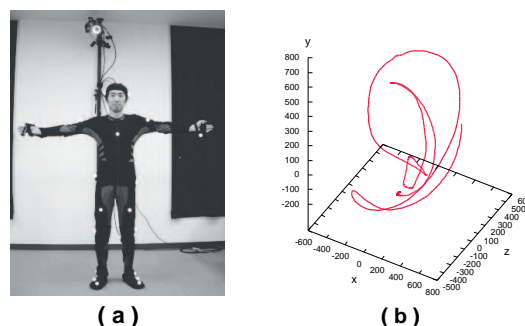


図 1: (a) モーションキャプチャシステムとマークを装着した演技者 (b)3次元時系列データの例

2. 動作データからの相関ルールの発見

体の個々のパーツに注目すると、動作は必ず静止状態から速度を上げ、ある地点で速度が減少し、静止するという過程を持っている。この状態の変化を利用して、動作データを基本動作に分割することを検討した。すなわち、動作の3次元時系列座標データの微分係数を x, y, z 各軸についてそれぞれ求め、微分係数が正から負、負から正、正から零、零から正、負から零、零から負に変化する点を検出し、分割点の候補とした。しかしながら、速度の変化に基づいた検出の結果、分割点の中には身体部位の微妙な揺れによって検出されたものがある。たとえば、時間的に極端に短い、または空間的にあまりに小さな動きを表しているなど、基本動作として意味のない内容を持ったセグメントに分割されるという問題が生じた。このため、Minimal Distance/Percentage Principle (MDPP) によって不要な分割点を削除している [2]。MDPP は、時間的、空間的なしきい値により分割点を評価し、時間的、空間的に意味のあるセグメントを生成するアルゴリズムである。

一方、人間の動作は音声と同様に、同じ人が同じ内容の動作を行なったとしても、動作を行なうのにかかる時間は微妙に異なっている。この違いは動作データの要素数の違いとなって現われる。このため、音声認識の分野で開発された Dynamic Time Warping (DTW) を導入し、動作データの類似度を計算している [1]。DTW は、動的計画法を用いてデータの伸縮を行い、比較的少ない計算量で要素数の異なるデータ同士の距

離を求めるアルゴリズムである。2つのデータ間の距離が小さいほど類似度が高くなるため、同一基本動作とみなすようにしている。

基本動作に基づいて動作データ中の各セグメントにラベル付けすれば、動作データをシンボルマルチストリームに変換することができる。図2にその一例を示す。A, Bで表される記号は基本動作を表しており、各パーツの動作は時系列で横一列に並んでいる。また、 W_a , W_p は、それぞれ依存関係を調べる際のウィンドウを表しており、 Int はその二つの動作の間隔を表している。

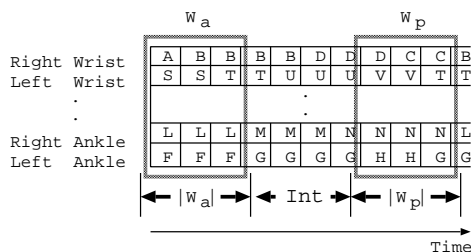


図2: シンボルマルチストリームの一例

このシンボルマルチストリームを用いて、動作データの依存関係を表す相関ルールを発見した。以下では、社交ダンスのひとつとして知られる、ルンバの動作データに対して基本動作の抽出を行い、さらに相関ルールを発見した例を示す。

$$\text{if 'left ankle 166' then 'left ankle 167' (1)}$$

この相関ルールは、“左足を踏み込むとともに右手が多少下がったのち、左足が持ち上がる”という依存関係を示している。これは、移動ともなっており、ルンバにおける特徴的な動きである。すなわち、“左足を踏み込む”という動作は移動する前の予備動作であり、一歩進むために“左足が持ち上がる”という動作が発生している。しかしながら、“右手が多少下がる”という動作は演技者個人の癖であると考えられる。このように、重要な相関ルールにより表される依存関係の中には、動作の特徴だけでなく個人の癖を表す情報が含まれている。

3. 動作データからのモチーフの発見

上記のアプローチでは、速度が0となる場所で動作データを分割するため、基本動作数が膨大になるという問題があった。また、動作データを分割してから、類似な基本動作をクラスター化してラベル付けするという2段階のステップを踏むため、分割誤りが起こると、それが新たな分類誤りを引き起こすという問題があった。本章では、動作データ中に頻出するパターン（モチーフと呼ぶ）を抽出するというアプローチを提案する[3]。モチーフは、与えられた時系データの特徴を表し、頻出する類似の時間的変移を持つ部分を示すパターンであると定義されている[4]。

3.1 動作データの次元縮約

動作データのような多次元時系列データからモチーフを発見する際、いくつかの問題を解決しなければならない。まず第一に、莫大な計算量が必要とされることがあげられる。また、計算アルゴリズムが複雑になる点も問題となる。このため、多次元時系列データを1次元時系列データに縮約することを試

みる。データの次元縮約を行えば、モチーフを発見するためのアルゴリズムが単純になり、計算量を軽減することができる。

次元縮約を行う際には、与えられた動作データの持つ情報の損失を最小にする必要がある。少なくとも、動作データに繰り返し現れる特徴的なパターンの情報を保持しなければ、次元を縮約した意味が損なわれる。以上のことから、統計学で広く用いられている主成分分析 (PCA)[5]に注目した。PCAは、多変量で表せるデータから観測変数の特徴を表す指標を発見する手法である。

PCAでは、入力された長さ n の多次元時系列データ C の次元数に等しい主成分が発見され、与えられたデータに対する寄与率（どの程度データを説明しているかを表す尺度）が大きい順に、第1主成分、第2主成分...と決定される。第1主成分が最も多く与えられたデータの情報を保持していることから、多次元時系列データの次元縮約には第1主成分を用いている。言い換えると、時系列データ C から直接的にモチーフを抽出する代わりに、1次元時系列データ \hat{C} を用いている。

3.2 モチーフの時間幅の決定

モチーフ抽出アルゴリズムを有用なものとするためには、モチーフの時間幅を動的に決定する必要がある。このため判定基準としてMDL原理[6]を用いている。MDL原理は、Rissanenにより提案された確率モデルの最適化原理である。MDL原理では「与えられたデータを、モデル自身の記述長も含めて、最も短く符号化できるような確率モデルが最良のモデルである」と主張している。さらに、「最良のモデル」がデータの特徴（内在的構造）を最もうまく表現していると仮定している。したがって、MDL原理の「最良のモデル」を1次元時系列データのモチーフとみなせば、MDL原理を用いてモチーフの時間幅を決定できることになる。

本手法では、MDL原理を最も満たすパターン（以降、「MDLパターン」と表現する）を抽出し、モチーフの時間幅を決定するために、まず、与えられた1次元時系列データの振る舞いを近似的に表現したシンボル列に変換している。本手法では、図3に示す、PAA (Piecewise Aggregate Approximation) 表現[4]に基づいたシンボル列変換手法を用いている。

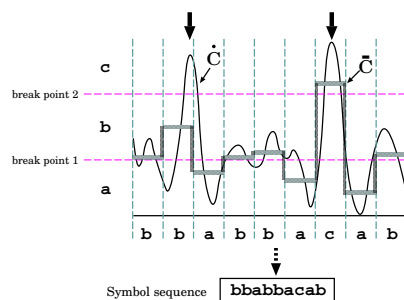


図3: PAA表現に基づくシンボル列変換

PAA表現は、1次元時系列データのある時間幅で区切り、各セグメントの平均値を用いたベクトル表現であり、1次元時系列データ $\hat{C} = \hat{c}_1, \dots, \hat{c}_n$ を w 次元のベクトル空間 $\hat{C} = \hat{c}_1, \dots, \hat{c}_w$ で表現する。また、 w 次元のベクトルをシンボル列に変換するために、波形の高さに応じて分割点 (breakpoint) を決定している。なお、正規化された時系列データは正規分布に従うという事実に基づいて、breakpointは正規確率分布において同じ大きさの領域を与える分割点としている。たとえば、図3の例では、それぞれの確率分布がガウス分布において3分の1に

なるように区切られた2つの閾値をそれぞれ breakpoint としている。そして、最小の breakpoint より順にシンボルが割り当てられる。最終的に、すべての区間にシンボルが割り当てられ、1次元時系列データがシンボル列に変換される。

変換されたシンボル列からパターンを抽出し、MDL原理を満たしているかを評価するためには、シンボル列の記述長を定義しなければならない。ここで、パターン SC で用いられるシンボル数を n_p 、パターン SC に使われるシンボルの種類数を s_p とおくと、パターン SC の記述長は以下のように定義される。

$$DL(SC) = \log_2 n_p + n_p \log_2 s_p \quad (2)$$

圧縮されたシンボル列のシンボル数を \hat{n}_a 、シンボルの種類数を s_a 、パターン SC がシンボル列 \tilde{C} 中に現れる回数を q とおくと、 \tilde{C} 中のパターン SC をひとつのシンボルに変換した場合の記述長 $DL(\tilde{C}|SC)$ は次のように定義される。

$$DL(\tilde{C}|SC) = \log_2 \hat{n}_a + \hat{n}_a \log_2 (s_a + q) \quad (3)$$

右辺第1項は \tilde{C} のシンボル数をエンコードするために必要なビット数、第2項は \hat{n}_a 個のシンボルのラベルをエンコードするために必要なビット数である。

最終的に、シンボル列 \tilde{C} 中のパターン SC に対する MDL 評価関数 $MDL(\tilde{C}|SC)$ は以下のように定義される。

$$\begin{aligned} MDL(\tilde{C}|SC) &= DL(\tilde{C}|SC) + DL(SC) \\ &= \log_2 \hat{n}_a + \hat{n}_a \log_2 (s_a + q) \\ &\quad + \log_2 n_p + n_p \log_2 s_p \end{aligned} \quad (4)$$

抽出されたパターンのうち、最小の MDL 評価値をもつパターン SC をシンボル列 \tilde{C} の MDL パターンとみなして、対応する部分時系列の長さを最適時間幅としている。また、それらの部分時系列をモチーフ候補とみなして抽出している。最終的に、時系列データ間の類似度を用いたクラスタリングをおこなう ADM アルゴリズム [7] を用いて、最も多くの候補を持つパターンをモチーフとして決定している。

図4は、どのようなパターンがモチーフとなるか直感的にわからない、非周期な動作である。このデータから抽出されたモチーフのうち、部分時系列 A, C は他の部分時系列と振る舞いが異なるように見える。しかしながら、それぞれの部分時系列の変位の大きさを正規化して示した図4(b)を見ると、各部分時系列は、ふたつの谷を持つ、同じ振る舞いを表わしていることがわかる。このことから、提案手法では、非周期なデータからモチーフが検出できていることがわかる。

図5は、「歩く」動作時の足の軌跡から抽出されたモチーフを、各座標からの視点で示した図である。抽出されたモチーフを y 座標から見ると、それぞれの部分時系列は同じ振る舞いをもち、かつデータの特徴的なパターンをとらえている。一方、 x 座標、 z 座標におけるモチーフは、1番目と4番目の部分時系列の振る舞いが、2番目と3番目の部分時系列の振る舞いと異なっていることがわかる。この現象は、多次元時系列データを1次元に縮約する際に生じると考えられる。図5の場合、PCA は y 座標を重要な座標とし、 x 座標と z 座標はそれほど重要でないと判断している。上記の特徴を人間の動作という観点からみれば、「歩く」という動作の特徴は、上下の動きを表す y 座標、左右の動きを表す x 座標、前後の動きを表す z 座標の順に現れる。したがって、 y 座標が3次元時系列データに大きな影響をもたらす変数として検出されたことは妥当であると考えられる。

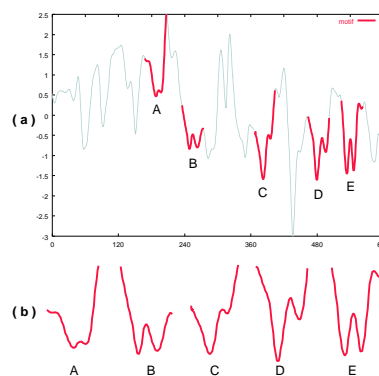


図4: (a)1次元化された3次元データから抽出されたモチーフ (b) 変位の大きさを正規化して並べた図

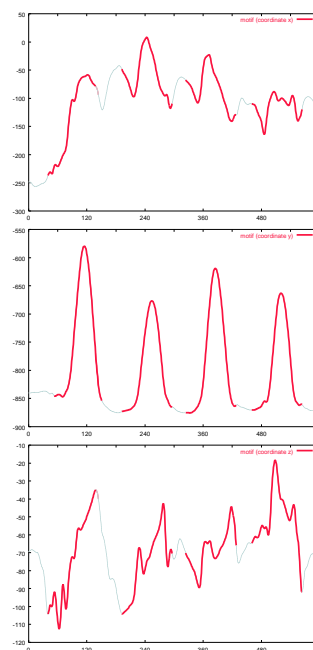


図5: 3次元時系列データから発見されたモチーフ例 (上から x , y , z 座標)

4. 運動スキルの比較

前章までに述べたアルゴリズムは、各個人の動作の特徴を抽出するものであり、運動スキルの異なる複数の被験者間の違いを明確にするには至っていない。このため、動作データを3次元空間中を動く軌跡としてとらえるだけではなく、その動作が行われる速度、あるいは動作中の姿勢を解析することで、運動スキルの比較を行い、高度なスキルを持つ被験者の特徴を得る実験を行っている。

まず、運動スキルの上達のためには運動リズムの習得が不可欠であると考えたため、動作の速度を考慮している。例えば、野球の初心者か、熟練者と同じ軌跡で野球のバッティング動作を行ったとしても、ボールとバットが当たる瞬間に、バットに加わる力のピークを合わせなければ、最大限の飛距離を出すことはできない。このため、動作が行われる軌跡と、その速度との関係を解析することで、効率的な運動が行われる際の特徴を得ることができると考えられる。

また、運動における一連の動作は、多くの場合、体の一部だけではなく、全身を用いて行われるものであり、各パーツの位置関係を適切に保たなければ、求める結果を得ることは困難である。したがって、運動が行われる際の姿勢を解析している。まだ明確な方針を得るには至っていないが、現在までに得られた結果を以下に紹介する。

4.1 動作速度の解析

本実験では、テニスのサーブ動作における右手の動きを、モーションキャプチャシステムを用いて取得したものを解析対象とした。また、被験者としては、テニス経験が十分にある熟練者2名、ほとんど経験のない初心者2名の計4名に実験に参加してもらった。解析の結果を図6に示す。横軸は被験者に対して前方を正とするz座標軸、縦軸は各サンプリング点における速度を表す。

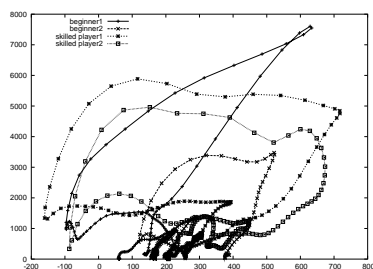


図6: 速度と座標の関係から解析した結果

この図から、初心者の2名のデータと熟練者の2名のデータは異なった形状を示していることが分かる。初心者の2名の速度のピークが、z座標のピークとほぼ一致するのに対し、熟練者2名の速度のピークは、z座標が比較的小さい位置に存在している。これは、初心者はラケットを振り終える頃になってから、ようやく最高速度に達するのに対し、熟練者はラケットを高く構えた直後に最高速度に達し、その速度を維持したまま振り抜いていることを示している。より速い速度でラケットを振ればより強力なサーブが打てること、打点が高くなればサーブのボールは相手のコートに届かないことを考慮すれば、熟練者がこのようなリズムで体を動かすのは理にかなっている。

4.2 姿勢の解析

テニスのサーブ動作を実験の対象として、姿勢の解析を試みた。図7にその結果を示す。グラフは極座標を用いて表されており、同心円を反時計回りに、ラケットを持つ右肘のなす角度、同心円の大きさが、被験者の垂直上方向を正とするy座標軸の大きさを表している。

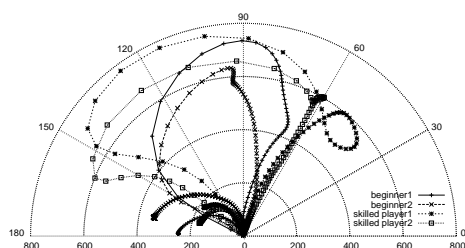


図7: 姿勢の解析結果

ここでも、初心者の2名と熟練者の2名のデータの間には大きな差があることがわかる。初心者の2名は、ラケットを高く構えたとき腕が曲がった状態にあり、振り下ろしながら腕を伸ばす動作をとっている。一方、熟練者の2名はラケットを高く構えたときに腕はほぼまっすぐに伸びており、その状態を維持したまま振り下ろしている。このことから、熟練者は腕を伸ばすことで打点を高くして、サーブを相手のコートに入れやすくしていると考えられることができる。

また、熟練者の軌跡には図の右上に円軌道を描いて交差している点が見られるが、これは、熟練者がサーブの構えをとったときに、腕を十分に曲げて力をためていることを示している。逆に、初心者の軌跡にはこのような「ため」の動作は見られない。このことから、熟練者は腕をバネのようにしなませることで、強いサーブを打つことが可能になっていると分かる。

5. まとめ

本論文では、複数の多次元時系列データからの相関関係の発見手法、および、頻繁に出現するモチーフを抽出するアルゴリズムについて述べた。また、運動学的な見地より、モーションキャプチャシステムによって得られた動作時系列データの解析を行い、初心者と熟練者の間の運動スキルの違いを特定しようとして試みた。今後は、運動スキルの違いが顕著に表れている要素を特定するアルゴリズムを開発していく予定である。

参考文献

- [1] R. Osaki, M. Shimada and K. Uehara, "A Motion Recognition Method by Using Primitive Motions", Proc. of 5th IFIP 2.6 Working Conference on Visual Database Systems, pp.117-128 (2000).
- [2] T. Mori and K. Uehara, "Extraction of Primitive Motion and Discovery of Association Rules from Human Motion," Proc. of 10th IEEE International Workshop on Robot and Human Communication, pp.200-206 (2001).
- [3] Y. Tanaka and K. Uehara, "Discover Motifs in Multi Dimensional Time-Series Using the Principal Component Analysis and the MDL Principle," Proc. of 3rd International Conference on Machine Learning and Data Mining in Pattern Recognition, pp.252-265 (2003).
- [4] J. Lin, E. Keogh, S. Lonardi and P. Patel, "Finding Motifs in Time Series," Proc. of the 2nd Workshop on Temporal Data Mining, pp.53-68 (2002).
- [5] D. B. Heras, J. C. Cabaleiro, V. B. Perez, P. Costas and F. F. Rivera, "Principal Component Analysis on Vector Computers," Proc. of VECPAR, pp.416-428 (1996).
- [6] J. Rissanen, "Stochastic Complexity in Statistical Inquiry," World Scientific Publishing (1989).
- [7] D. Shasha and T. Wang, "New Techniques for Best-Match Retrieval," ACM Trans. Information Systems, Vol.8, No.2, pp.140-158 (1990).