

時空間的特徴と帰納推論を用いた動作データの検索

Retrieval of motion data using spatio-temporal features and inductive learning

脇坂 健一† 向井 智彦† 栗山 繁†
Ken-ichi Wakisaka Tomohiko Mukai Shigeru Kuriyama

1 はじめに

近年、映画・ゲームなどの娯楽分野や教育分野で動作データは幅広く利用されている。一般的に、こうしたデータを効率的に管理、検索するために、各動作にインデクスを付ける事が多い。しかし、手作業によるインデクス付けには膨大な作業量が必要である。

そこで、動作データを自動的に分類し、検索のためのインデクスを付与するシステムが提案されている。例えば姿勢の幾何学的特徴量に基づく確率的なテンプレートを用いる検索法 [1] や、機械学習に基づく分類器を用いて動作データに注釈をつける方法 [2] がある。しかし、いずれの手法も検索キーとデータベース要素をテンプレートマッチングする際に、その類似度を数値的に評価する必要があるため、同一カテゴリに属する動作でも見た目が大きく異なる動作を検索することができない。そのため、「上投げ」と「下投げ」のように意味的には類似するにも関わらず、見た目が異なるような動作は同時には扱えない。

そこで本研究では、動作データの論理形式表現と帰納推論を用いた動作データの検索手法を提案する。本手法では、まず動作データの幾何学的特徴量とその時間的関係を意味する時空間的特徴量に基づき、あらかじめ分類された動作データから分類規則を導出する。そして、未知のデータに対して特定のカテゴリに属する部分に意味情報を付与する。これにより動作の意味的な類似性を考慮した検索ができる。

2 アルゴリズム

2.1 提案手法の概要

図1に提案手法の処理の流れを示す。まず事前にカテゴリ分けされた全ての教師データから時空間的特徴量を抽出する。次に、各動作カテゴリに共通の特徴量を解析するために、時空間的特徴量を元に帰納推論し、分類規則を抽出する。そして与えられた動作データに対しても同様に時空間的特徴量を抽出し、分類規則を参照しながら動作に対して意味付与を行う。最後に意味付与された動作区間と検索キーを比較することで、目的の動作セグメントを検索する。各処理の詳細を後の節にそれぞれ示す。

2.2 動作データの時空間的特徴量

まず、分類において本質的でない微小な動作を省き、正確な指標化を行うために、動作データの特徴を表すキーフレームを抽出する。本研究では、先行するキーフレームからの姿勢距離が一定の閾値以上になる姿勢を順次選択するアルゴリズムを用いる。なお、こうして少数のキーフレームに対してのみ特徴量を計算することで、

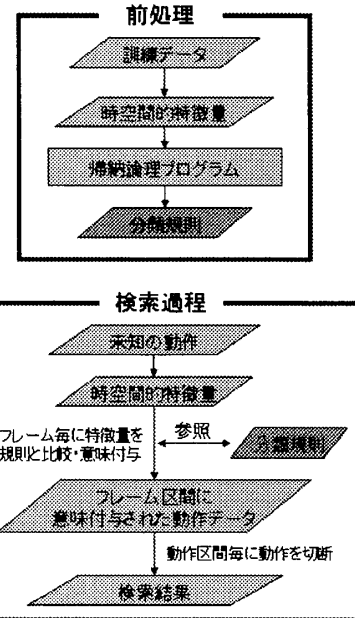


図1 本手法の処理手順

分類規則の導出計算を高速化できる。

次に、各キーフレームにおける姿勢の幾何学特徴 [3] を計算する。これは「手が前に出ている」や「足が上がっている」など、動作中の各姿勢が示す空間的な特徴をオンオフ型の二値で表現するものである。ただし、前実験の結果では、二値の幾何学特徴と帰納推論法を用いた動作データの検索性能は非常に低かった。この原因として、例えば「足が前に出ている」という特徴量が存在した場合、オンオフ型の二値データで表わすと「10cm前に出ている」と「1m前に出ている」という特徴を同じものとして扱ってしまうためではないかと考えられる。動作の特徴として姿勢の外見的特徴を利用することを考えた場合、本研究ではこのような特徴量は別のものとして扱うのが妥当だと考え、特徴量に多段階化を施し、姿勢特徴の程度を表現できるよう拡張したものを特徴量とすることとした。なお、本研究では5段階の多段階化を用いた。さらに、文献 [3] で提案されている特徴量に加えて、「地面から全身が離れている」、「身体が捻られている」などの新たな特徴量を定義する。

また、動作の流れも指標化に利用するために、幾何学特徴の時間関係を表す時間特徴量を加える。すなわち、ある幾何学特徴が出現した後に別の幾何学特徴が新たに出現するまでの経過時間により3種類の時間特徴量を作成する。ここでは、経過時間が [0,0.5) 秒の範囲のものを"s_after", [0.5,1.0) 秒を"m_after", [1.0,2.0) 秒を"l_after"として定義する。そして幾何学特徴の出現

† 豊橋技術科学大学 情報工学系

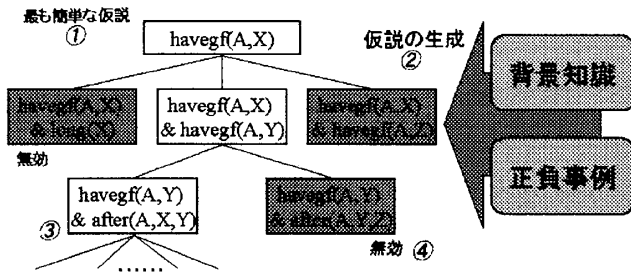


図2 Progolの規則解析手順

時間の長短を表す時間特徴量も新たに加え、 $[0.0, 1.0)$ 秒を”short”, $[1.0, +\infty)$ 秒を”long”として定義する。これら定義した特徴量を動作のキーフレームごとに抽出し、各動作データクリップを動作に含まれる特徴量の集合に変換する。特徴量は論理式の形式で表記されるため、出力されるのは論理式の集合となる。

2.3 帰納論理に基づく分類規則導出

あらかじめ動作別に手作業で分類された各データの特徴量を解析し、自動分類のための規則を導出する。本研究では Progol[4] と呼ばれるシステムを利用する。Progol は帰納論理プログラミング法の代表的なシステムであり、あるカテゴリに共通する特徴を帰納推論する。このとき、全てのデータは論理式の集合によって表現されるため、分類規則は数値的なものではなく、人が見て意味の分かる形式で出力される。解析の際に Progol は正例と反例というものを利用する。すなわち、ある1つのカテゴリの動作を正例とすると、他のカテゴリの動作は全て反例となる。また、分類規則は与えられたデータを正例もしくは反例と判別するために導入され、各カテゴリごとに1つの分類規則を定義する。具体的な解析の手順を図2に示す。Progol は規則導出にトップダウンアプローチを用いており、番号①から④のような流れで解析する。まず、①最も簡単な仮説から検定を開始し、②新たな仮定を仮説に加え、それを背景知識及び事実で仮説検定する。③この仮説が妥当なものであれば、それを元にまた別の組み合わせを検定する。そうでなければ、それを基にした仮説はそこから生成されず、④検査は打ち切れ、別の組み合わせの検定を始める。この操作を論理式の集合から選ぶことのできる全ての組み合わせについて行い、最終的に最も評価の高い組み合わせが規則として出力される。本システムで解析された分類規則を表1に示す。例えば分類名「Throw」の分類規則「 $m_after(A, l_foot_front, l_r_hand_front, 0), havegf(A, 0, hands_touch)$ 」は、「左足が前に出た後、右手が前に出る」かつ「両手が接触している」と読み替えられる。このように、規則を意味的に理解でき、ユーザの手作業によって容易に書き換えられる。

また、帰納論理推論法では規則の導出において背景知識と呼ばれるデータを利用する。これは、解析対象について既知の事柄を表すもので、より効率的かつ妥当性のある解析を行うことができる。さらに、Progol 特有の機能として、反例データを利用しない正例学習が提供されている。つまり、通常の学習ではカテゴリ内だけに共通し、他のカテゴリには含まれない特徴量を排他的に解析

するが、正例学習ではカテゴリ内だけで分類規則導出をする。したがって、通常学習では排他的制約の強い分類規則を導出するが、正例学習では比較的弱い規則を導出する。通常学習による規則だけでは意味付けしきれない部分も拘束の弱い正例学習による規則で意味付けすることが可能となる。本研究ではこの二つの規則を組み合わせ、分類規則として利用する。

一方、従来手法 [1] では、同じ動作カテゴリの特徴量テンプレートを確率的に重ね合わせることで検索キーを生成する意味において、本研究における正例学習規則の単体適用に等しい。そのため、少量の教師データからテンプレートを作成すると、テンプレートに用いられている特徴量が他の動作にも多く含まれている可能性があるため、妥当性のある検索を行うことができない。よって正例学習規則に加え、反例データも導入した通常学習規則を多段階で適用することにより、従来法よりも検索精度が向上すると考えられる。

2.4 分類規則を用いた意味付与

本手法では二段階の処理により動作データに対して意味情報を付与する。まず、与えられた未知の動作に対して通常学習による分類規則を参照し、検索対象の動作中に出現する特徴量を先頭のフレームから順次見てゆく。ある分類名に用いられている特徴量と同じ特徴量が出現している動作フレームに対して、現在参照している分類名で意味付与してゆく。次に、意味付けされていないデータ部分に対して、正例学習による規則を参照し、同様の処理を行う。それぞれの規則はあらかじめ教師用データから帰納推論によって解析されたもの他にも任意の特徴量を用いて手作業で分類規則を記述することが可能である。これにより、それぞれの規則単体で処理を行った時よりも動作に正確に意味情報を付与でき、また、分類規則を手作業で記述することにより直感的な検索が可能となる。

最後に、近傍にある意味付与された同カテゴリの動作を結合する。まず、ある動作区間の直前と直後の動作区間に注目する。その両者のうち離れているフレーム数が小さい動作を結合の対象とする。次に離れているフレーム数と結合の対象となっている動作の長さを比較し、離れているフレーム数のほうが小さければ、1つの動作として結合する。その後、各動作についての開始フレームと終了フレームを調べ、データを切り分ける。

3 実験結果

今回訓練用に利用した動作データは Walk, Run, Sit, Throw, Jump, Lift, Lying を各6動作ずつ、計42動作である。これらを用い、従来手法の代表的な方法である、サポートベクタ法、テンプレートマッチング法の二手法と本手法の比較を行う。

3.1 サポートベクタ法との比較

数値分類の代表的な手法である Support Vector Machine (以下, SVM) を用いて、分類の実験を行った。SVM は N 次元データを $N-1$ 次元の超平面で区切り、与えられた訓練データとのマージンを最大化するような分類器を数値的に学習する。本手法との比較を行うため、訓練用データに含まれる全ての時空間特徴量を用いて、Context-Aware SVM を分類器として利用し、訓練用デー

表1 各分類の分類規則

分類名	通常学習分類規則	正例学習分類規則
Walk	s_ after(A,r_ foot_ bent,0,r_ foot_ front,2) & slow(A)	after(A,r_ foot_ front,3,legs_ cross) s_ after(A,r_ foot_ bent,0,r_ foot_ front,2)& slow(A)
Run	havegf(A,0,r_ foot_ up,0) & s_ after(A,r_ foot_ up,0,r_ foot_ front,0)	s_ after(A,r_ foot_ bent,0,r_ foot_ front,2)& slow(A) s_ after(A,l_ foot_ up,2,l_ hand_ bent,0)
Sit	long(A,r_ foot_ bent,0) & long(A,l_ foot_ bent,0)	havegf(A,r_ foot_ bent,0) & short(A,r_ hand_ bent) havegf(A,body_ bent,0) & long(A,r_ foot_ bent)
Throw	m_ after(A,l_ foot_ front,1,r_ hand_ front,0) & havegf(A,0,hands_ touch)	s_ after(A,l_ hand_ up,2,hands_ touch) m_ after(A,l_ hand_ bent,0,l_ foot_ up,0)
Jump	havegf(A,jumped) & slow(A)	m_ after(A,move_ upward,move_ downward) s_ after(A,move_ upward,move_ downward)
Lift	m_ after(A,r_ hand_ front,3,twist,0) m_ after(A,gradient,0,l_ hand_ front,2) & long(A,r_ hand_ front,0)	s_ after(A,l_ hand_ front,4,r_ hand_ up,0) m_ after(A,gradient,0,l_ hand_ front,2) long(A,r_ foot_ side)
Lying	havegf(A,lying)	havegf(A,lying)

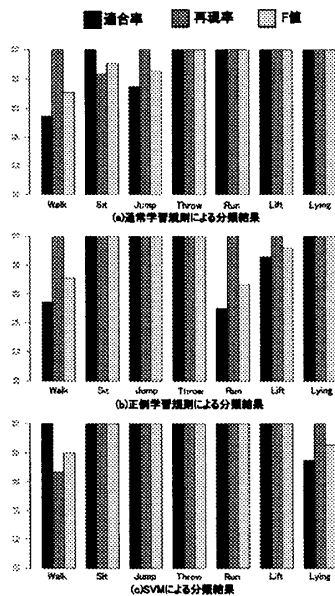


図3 SVM法と本手法のF値の比較

タを7クラスに分ける分類平面を作成した。この分類器を用いて訓練用データを正確に分類できるか実験した。評価値として検索精度の指標となるF値を用いた。これはデータベース中に含まれる目的データを検索した際の適合率、再現率から算出される指標である。適合率は検索結果の中にどれだけ検索に適合したものが含まれているかという正確性を表したものであり、再現率は検索対象のデータ群の中からどれだけデータを検索できているかという網羅性を表したものである。SVMと本手法による検索結果を図3に示す。図3のF値だけを見るとSVMの方が良好な手法のように見える。しかし実験ではまったく外見の異なる「Walk」を「Lying」と意味付けしていた。このように数値的な分類は処理に失敗するとまったく違う意味が割り当てられてしまうという可能性が存在する。また、扱っているデータが数値データという直感的にわかりにくいデータであることと、時系列データの特徴量を与える場合にはあらかじめ手作業によってデータをいくつかの区間に分割しておかなければ

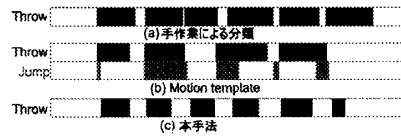


図4 訓練用データ指標化結果-男性-(投げる)

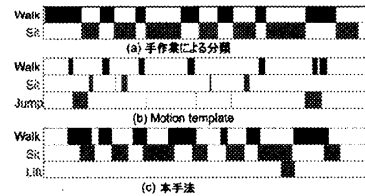


図5 訓練用データ指標化結果-男性-(座る)

ならない。一方本手法は正確な分類に失敗しても意味的に似通った分類ができ、また時系列中の適合区間をテンプレートマッチングに似た方法によって容易に検索できる。

3.2 テンプレートマッチング法との比較

与えられた動作に対して動作データのテンプレートマッチング法 [1] と本手法を用いて検索を行った。結果を図4と図5に示す。図4は分類名「Throw」の訓練用データを含み、図5は分類名「Sit」の訓練用データを含む動作である。それぞれの訓練データには同じ動作カテゴリに属するが、見た目の異なる動作が含まれている。図4を見ると本手法では手作業で意味付けした場合とほぼ同じように処理されている。一方、従来法を用いた場合の処理結果を見ると「Throw」とほぼ同じ部位に重複して分類名「Jump」が割り当てられており、正しく処理されていない。また、図5で従来法は手作業で分類した場合と比較するとほぼ全ての部分において処理に失敗しており、動作の取りこぼしも非常に多いが、本手法はほぼ全ての動作において正確に処理され、動作の取りこぼしも無いように見える。しかし、図5では一部重複して分類名「Lift」が割り当てられている部分がある。これは、着座姿勢と腰を曲げて物を拾う姿勢が特徴量として非常に近いために生じている。

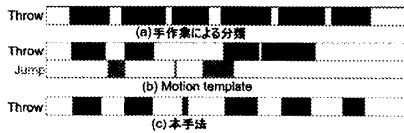


図6 非訓練用データ指標化結果-男性-(投げる)

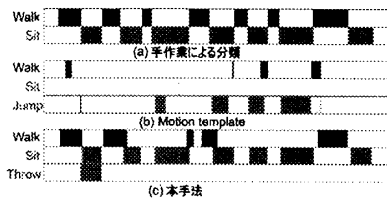


図7 非訓練用データ指標化結果-男性-(座る)

次に、訓練用データを含まないデータに対しての検索結果を以下に示す。図6は分類名「Throw」と同じカテゴリの動作を含むデータであり、図7は分類名「Sit」を含むデータである。図6を見ると本手法は、ほぼ正確に検索できているが一部重複して処理されている部分がある。この原因は前述の通り、姿勢が似通っているために起こる現象である。また、図7を見ると本手法で、一部検索できていない部分がある。これは検索の際に姿勢距離を用いてキーフレームを抽出し、処理の高速化を施しているため、本来その動作に含まれる特徴量をそのフレームで取得できなかったのではないかと考えられる。従来法での処理結果を見ると図6では重複して「Jump」が割り当てられ、図7では着座姿勢そのものが取得できていない。また、訓練に用いたデータと同様に動作の取りこぼしが非常に目立っている。なお、いずれの実験でも本手法が検索処理に要した時間は数秒程度である。

結果のムービーファイルは以下のアドレスを参照されたい。

<http://val.ics.tut.ac.jp/VAL/project/ilpretrieval/>

4 まとめ

本研究では動作の時空間的特徴量と帰納論理推論法を用いた動作の検索手法を提案した。まず、訓練用動作から多段階化した幾何学的特徴及びその時間的關係である時空間的特徴量の集合を作成した。次に、Progolを用いた帰納推論では通常学習、正例学習による二種類の規則を導出し、規則は人間が見てわかりやすい論理式の形式で導出され、容易な編集を可能とした。また、二種類の分類規則を段階的に適用することで、より妥当性のある意味付けも提案した。動作の検索の際には未知の動作データに対してあらかじめ抽出されたカテゴリ名または手作業で入力した分類規則を検索キーとして入力することで、従来法では処理できなかった外見的特徴は異なるが、意味的に類似した動作を効率的に検索、意味付けを行え、取りこぼしの多かった動作も取得することができた。

しかし、Progolは大規模な学習データに対し、計算量が爆発的に増加するので、計算時間が増大するばかりか、計算そのものが破綻し、規則が導出できないことがある。実際、本実験で大規模な訓練データを用いた規則

導出を行ったとき、全ての計算を終わらせるまでに数日間を要した。このため、動作データの全フレームを取り出さず、キーフレームからの特徴量抽出を行い、計算量を可能な限り減少させた。しかし、余りにもデータ数が少なすぎると今度は特徴量の少なさから、妥当性のある規則を導出できない。このトレードオフの關係の中で最適な特徴量集合の数を考慮する必要がある。また、外見が類似しているが意味の違う動作を同じカテゴリに分類してしまうという弱点も存在する。

今後の課題としては、幾何学特徴以外の特徴量を用いた手法や、大規模なモーションデータベースを用いた実験などが挙げられる。

5 謝辞

本研究は、日本学術振興会科学研究費・基盤研究(B)18300068「大規模なヒューマンデータの知的検索と再利用」および同・若手(B)19700090「人体動作データの運動学的スキル解析と可視化」の援助を受けた。ここに感謝の意を表す。

参考文献

- [1] M.Müller, et al., "Motion Templates for Automatic Classification and Retrieval of Motion Capture Data", SCA 2006, pp. 137-146, 2006
- [2] O.Arikan, et al., "Motion Synthesis from Annotations", ACM TOG, 22(3), pp. 402-408, 2003.
- [3] M.Müller et al., "Efficient Content-Based Retrieval of Motion Capture Data", ACM TOG, 24(3), pp. 677-685, 2005.
- [4] 古川康一, 尾崎知伸, 植野 研: "帰納論理プログラミング", 共立出版, 2001