原著論文

模倣学習の機序解明に向けた意図推定のモデル化：倒立振子課題の分析†

鳥居 拓馬*・日髙 昇平*

他者の行為を観察して自らの行為を学習するには他者の目的を推定する能力が求められる．ある心理実験では，ヒトの18か月児が大人の行為を観察し，その未知なる目的を推論できることを示した．本研究では，この心理実験に着想をえて，どうすれば未完遂の行為の観察からその本来の目的や計画を推定できるかを数値実験で検討する．具体的には，異なる目的（課題）に最適化された２種類の単振子が一見よく似た軌道（動作）を生成するような実験設計を行い，どんな特徴量を用いれば軌道を生成した未知なる制御器（意図）を識別できるかを検討した．本研究の分析では，力学系の不変量「フラクタル次元」は，観察対象に関する事前知識をほとんど要さずとも，意図の識別に十分な情報をもつことを示した．

キーワード：模倣学習，意図推定，力学系，倒立振子課題

1. 序論：行為の模倣

他者の行為を観察して自らの行為を学習することは社会性の基礎をなす能力である [1]．本稿ではある目的に向けた動作を「行為」と呼ぶ [2]．行為を学ぶとは，その動き（動作）を学ぶだけでなく，その目的や計画を推定することを要する．目的や計画を推定することで，表面的な動作の類似を超えた行為の模倣やある種の文化伝達が可能となる [1]．社会において，他者の目的や計画を推定する能力は，他者を助ける（他者の行為を代行する）ときにも必要となる [1, 3]．したがって，他者の目的や計画を推定する能力は社会的なロボット等を構築する上でも基礎と考えられる [4].

人の場合には，限定的な知識・情報でも，目的の推定が可能であることを示す経験的知見がある．Warneken & Tomasello [3] の心理実験では，ヒトの18ヶ月児は大人の行為の結果（目的を反映する）を見ていなくてもその途中の動作から大人の目的を推定できることを示した．とくに，大人が目的を失敗した場合でも，失敗した動作から大人の目的を推定し，その大人の目的を代わりに達成（補完）できることを示した．

社会的なロボットに関しては，大分して２つのアプローチがある．一方は，Human-Robot Interaction（HRI）の分野で，どのようなロボットの動作・表情を用いれば人間との相互作用を円滑にできるかを問題にする．例えば，人間に移動方向を伝える仕草 [5] や喜怒哀楽等の感情を伝える仕草 [6] など，人間に対して目的や意図を顕示する表現がロボット実装と評価実験で調べられている．他方は，認知発達ロボットの分野 [7, 8, 9, 4] で，ミラーニューロン等の神経科学的な知見をロボットに実装し，そこから他者の動作等の社会学習が可能か [7, 9] が研究されている．これらのアプローチは実社会の複雑さの中で実際に動くロボットを作る上では必要な研究といえる．その一方で，どうして人は他者の行為を観察することから目的や意図を読み取れるのか，また，どうして人間は他者の行為を観察することから自身の行為を学べるのかという「社会学習の機序」を問うレベルでは，研究者自身の直観や異分野の知見に基づき研究者自身の手でロボットに作り込むことで根本的な問題が回避されることが多い．その背景には，社会学習や模倣学習の機序に関して現在でも十分な知見がえられていないことが挙げられる．

他者の行為の観察からその背後にある目的や計画を推定するメカニズムの解明に向けて，本研究では，Warneken & Tomasello [3] の心理実験に着想をえて，どうすれば未完遂の行為の観察からその本来の目的や計画を推定できるかを数値実験で検討する．社会学習や模倣学習の状況では，教師に当たる実演者と生徒に当たる観察者（あるいは模倣者）が関わる．行為の模倣を考える上で，基本的な問題は少なくとも２つあると考えられる．

第一に，通常，観察者（子供）は実演者（大人）とは異なる身体をもつほか，他者の内部状態は観察できないため，観察可能な実演者の動作は観察者が自分自身の動作を決定する全変数を指定するにはデータとして不十分である．このように，他者の動作から他者の計画や目的を推定する問題は不良設定問題となる [10, 11].

第二に，通常，ある目的を達成する複数の動作がありえ，また，類似した動作が複数の異なる目的に対してありえる．この潜在的な多対多対応ゆえに，観察者はある目的に関連する動作とそうでない動作を分類する必要がある．この分類は，目的達成時の動作や結果が観察できる場合にはその最終状態によって分類しうる．しかし，本研究のように目的達成時の動作や結果が観察できない（失敗動作しか観察できない）状況では，未知なる目的に沿う行為を推測した上で分類するという難しさがある．

以上の２点を踏まえ，本稿では単振子を単純な身体のモデルとみなして倒立振子課題を行うエージェントの運動制御を“意図”として推定可能か検討する．倒立振子課題では，目的（課題）は単振子を逆立ち姿勢まで振り上げてその姿勢を維持することである．その制御器はある時点の角度と角速度を系の内部状態として受けとり制御命令を決定する関数である．制御器は

---

† Modeling intention inference as a basis of imitation learning: analysis of pendulum swing-up task

Takuma TORII, Shohei HIDAKA

* 北陸先端科学技術大学院大学

Japan Advanced Institute of Science and Technology

目的（課題）に向けた動作（軌道）を逐次決定する計画（制御）[1]であり，その意味で「意図」と呼べる．倒立振子課題を達成する制御器を実演者としたとき，観察者の立場から振子の動作を観測してえられるのは例えば振子頭部の座標データである．そこで目的推定のメカニズム解明に向けた第一段階は，どうすれば，目的に関する事前知識なしに，観察された動作のみからそれを生みだした「意図」の違いを見抜けるかを解明することにある．

2. 意図識別の数値実験

2.1. 実験設計

本稿では目的推定に関する Warneken & Tomasello [3] の心理実験を理想化した数値実験を行う．本節ではその心理実験の概要を述べながら，本稿で用いる用語を定義する．

目的推定の心理実験 [3] では，観察者（子供）と実演者（大人）が参与し，実演者の行為を観察した観察者の反応が調査された．実演者の行為は２種類の条件により異なる．一方の条件では，実演者はある目的の下で行為を行うが，その動作によってその目的を達成できない（実際には役者は「ふり」をする）．本稿ではこれを失敗条件と呼ぶ．他方の条件では，実演者は別の目的の下で行為を行い，その動作によってその目的を達成できる．本稿ではこれを成功条件と呼ぶ．心理実験では，２つの条件間で実演者の動作は異なる目的にも拘わらず一見類似するように行われ，観察者（18か月児）はこの表面的な動作の類似性に騙されず本来の目的を推定することを期待された．

この心理実験を数値実験で扱うべく，失敗条件の要点を整理する．実演者はある課題（目的）のためにある制御（意図）を学習したとする[2]．いまこの制御はその課題を達成できるとしよう．そのとき，この制御はその課題には適しているが，必ずしも他の課題には適さない．もし実演者の想定しない処で課題やその制約が変更された場合，その制御は変更された別の課題には適さない．いわば，課題（目的）と制御（意図）が不整合のとき，その課題に最適な制御といえない点で「目的を達成できない」すなわち「失敗」となる．他方で，課題（目的）と制御（意図）が整合的のとき，その課題に最適な制御といえる点で「目的を達成できる」すなわち「成功」となる．こうした課題（目的）と制御（意図）の整合・不整合を計算機上で作りだすことができれば，心理実験に着想をえた数値実験が可能だと考えられる．

本稿では， [3] の心理実験の論理構造と同型な構造として，一見よく似た振り上げ動作の集合がデータとして与えられたとき，その動作を生成した複数の異なる制御器（意図）を識別できるかを検討する．具体的には，それぞれの課題（目的）を異なる制御（意図）で達成しようとする２体のエージェント（実演者）を用意する．各エージェントはそれぞれの課題の下で最適な制御を学習し，その制御を用いて学習時と同じ課題あるいは異なる課題の下で行為を実演する．学習時と実演時で課題が変更された場合が失敗条件に相当する．言い換えると，一方のエージェントは，その制御の学習時と同じ課題で実演を行い，目的を達成する（成功条件）．他方のエージェントは，その制御の学習時と異なる課題で実演を行い，目的を達成できない（失敗条件）．後に述べるように，学習時と実演時の２つの課題・制約を工夫することで，制御は異なるが類似した動作を行うよう工夫できる．このとき，もしオブザーバ（観察者；心理実験では子供に相当する第３のエージェント）が一見類似した動作であってもその制御の違いを推定できれば，オブザーバが意図の違い（制御の違い）を識別できたと考える．

2.2. 倒立振子課題

本稿では，運動制御の単純な課題として倒立振子課題（単振子の逆立ち課題） [12, 13] を扱う．倒立振子は長さ $l = 1.0$ の棒の先端に質量 $m = 1.0$ をもち（図 1 左上），その状態は角度と角速度の対 $(\theta, \dot{\theta})$ で定まる．倒立振子の運動方程式は次式である [13].

$$m\,l^2\,\ddot{\theta} - 9.8\,m\,l\sin\theta = -b\,\dot{\theta} + u$$

ここで，摩擦 $b = 0.01$，制御トルク $u \in [-5, 5]$ である．倒立状態は $\theta = 0$ で与えられる．倒立振子課題の制御器は振子の状態 $(\theta, \dot{\theta})$ から制御トルク u への関数である．

倒立振子課題では，ある制御器の“良さ”は評価関数 $\sum_t \cos\theta_t$ が用いられる [13]．この評価関数は「素早く倒立姿勢まで振り上げてその姿勢を維持すること」がより高い評価を与えると記述する．

古典力学では，振子の系の状態を記述する特徴量として力学的エネルギーが知られている．力学的エネルギー E は運動エネルギー K と位置エネルギー U の和で表される．

$$E(\theta, \dot{\theta}) = \frac{1}{2}\,m\,l^2\,\dot{\theta}^2 + 9.8\,m\,l\,(\cos\theta - 1)$$

力学的エネルギーを用いると，倒立静止状態は $E(0,0) = 0$ と表せる．

2.3. 成功条件と失敗条件

倒立振子課題において，異なる制御で一見よく似た動作を生成するため，本稿では制約振子を導入する．制約振子（図１右上）は，標準的な振子（標準振子）と異なり，倒立状態を含むある範囲の角度 $\theta \in \pm\pi/8$ が可動域から除外されている（この範囲に反発のない「壁」があるものとみなせる）ほか，角度 $\theta \in \pm\pi/6$ では制御トルクが無効となる．

---

[1] 制御理論では，制御器は「方策」とも呼ばれる．本稿では「方策」は用いず「制御」に統一する．

[2] 制御理論では，課題（目的）とは評価関数，制御対象・制約条件（身体・環境）等で記述される．また，制御（意図）は制御対象の状態から制御命令への関数である．

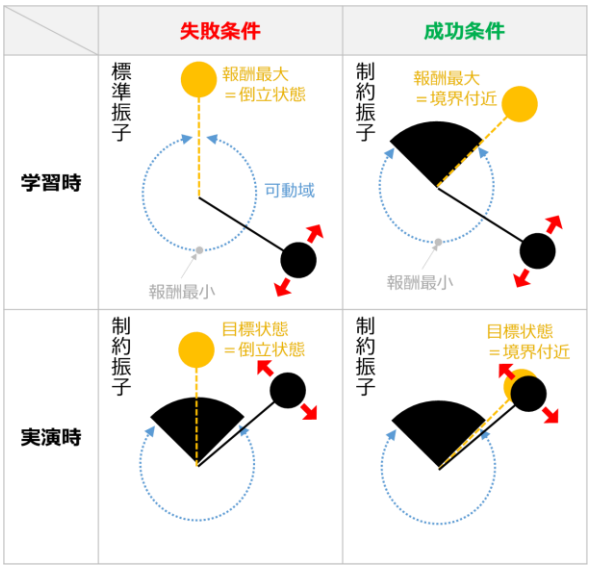

図 1. 本稿の数値実験の実験設計．学習用と実演用で２種類の振子を用いることで，実演時に一見似ている動作を生みだす異なる制御を学習させる．（左上）標準的な倒立振子課題．（右上）可動域に制約のある倒立振子課題

本稿の数値実験[3]では，標準振子と制約振子を用い，2.1 節の実験設計を実装する．具体的には，可動域に関する制約の有無を学習時と実演時で操作する，つまり同じ課題か異なる課題かを操作することで，実演時の課題に対して最適な制御（成功条件）と最適でない制御（失敗条件）を作りだし，一見同じ動作を生みだす異なる２つの制御とする．2.1 節と同じく，ある課題（目的）を異なる制御（意図）で達成しようとする２体のエージェントを考える．それぞれのエージェントは成功条件と失敗条件に対応する．まず，成功条件では，「成功」エージェントは**制約振子**を用いて最適な制御を学習し，**制約振子**を用いて実演を行う（図１右上→右下）．したがって，成功エージェントは最適な動作を行う．他方，失敗条件では，「失敗」エージェントは**標準振子**を用いて最適な制御を学習し，**制約振子**を用いて実演を行う（図１左上→左下）．したがって，失敗エージェントは最適でない（準最適な）動作を行う．学習用の振子（標準振子か制約振子か）は各条件で異なるが，実演用の振子（制約振子）は両条件で一致している．この実験設計により，本研究では，異なる制御から類似した動作を生成可能にした．この実験設計を用いて，オブザーバが類似した動作の背後にある制御（意図）の違いを識別できるかを本稿では検討する（図１左下・右下）．

本稿では，一見類似した動作（軌道）の背後にある意図（制御）の相違の認識を問題とするので，振子の動作を人間が見たとき直ちに制御の相違を認識できるような非可動域の設定は意味をなさない．本稿では，一見区別できない動作を生みだす非可動域の設定は角度の時系列（図2）を著者らが目視して選んだ．本稿では非可動域 $\theta \in \pm\pi/8$の場合を報告するが，後述する本稿の結果から，他の非可動域の設定でも本稿の主題であるフラクタル次元（3.1節）に関しては定性的に同様の結果がえられると考えている．

2.4. 強化学習を用いた制御器の最適化

倒立振子課題において，累積報酬 $\sum_t \cos\theta_t$ の最大化を目的とする動作を生みだす制御器 $u = g(\theta, \dot{\theta})$ は強化学習（アクター・クリティク法） [14] を用いて構築した（学習アルゴリズムやパラメータの詳細は付録 A を参照）．状態空間 $(\theta, \dot{\theta}) \in [-\pi, \pi] \times [-2\pi, 2\pi]$ を 40×40 分割のタイルコーディングとし，正規乱数を用いた方策勾配降下で制御トルク $u = g(\theta, \dot{\theta})$ を求めた．学習は 5000 試行（各試行 10000 ステップ）以内には収束し，ほぼ最適に近い期待報酬をだす制御関数を構築できた．

3. フラクタル次元による制御の識別

3.1. 本研究の仮説

本稿では，複数の制御の間の相違が，それらの制御から生成される動作の「構造的複雑さ」の相違として観察されるという仮説を検討する．本稿の実験設計では，学習時と実演時で同じ課題・制約とする場合を成功条件，異なる課題・制約とする場合を失敗条件と呼んだ．成功条件では，その制御器は実演時にも同じ課題の下で最適だとみなせる．他方，失敗条件では，実演時に異なる可動域制約をもつので，その制御器は異なる課題の下で最適ではない．最適制御理論の知見から，最適な制御器の生みだす動作は不必要な動作のばらつきを排除していることが予想される [15]．他方，準最適な制御器の生みだす動作は不必要な動作のばらつきを伴うことが予想される．この予想を踏まえ，本稿では，最適・準最適に由来する動作のばらつきの構造が，一見類似した動作の背後にある，制御器の相違を見抜く手がかりとなるという仮説を立てた．

具体的には，本稿の実験設計では，成功条件と失敗条件はどちらも制約振子（図１左下・右下）を用いるが，失敗条件の制御器は可動域**制約なし**の前提で最適化されている点で，成功条件の制御器（可動域**制約あり**の前提）と異なる．この実演時に突如出現した可動域制約のため，失敗条件の制御器は可動域の境界（ある種の「壁」）に接触・衝突する動作を示す．この接触・衝突といった挙動は，可動域制約の下で学習した成功条件の制御器にはほとんど見られない（罰を避ける）．本稿の仮説では，課題の変更に伴う可動域制約の追加という系の自由度の

[3] 数値実験について抽象的に述べる前に具体的な状況で述べる．標準振子を人の腕とみなす．可動域に制約をもつ制約振子は“五十肩”のようなの症状（肩関節の可動域が狭くなる症状）を患った腕とみなせる．ここで２人の大人を考える．一方は五十肩を数十年経験し慣れた五十肩の熟練者（すでに付与の肩関節可動域を熟知した者）であり，他方は五十肩の発症から間もない五十肩の初心者である．熟練者は五十肩の制約を前提とした腕の制御を行える（成功条件）．他方，初心者はあたかも肩関節の可動域が実際に可能な領域より広いかのように腕を動かそうとして五十肩を痛感する（失敗条件）．腕を上げたときの両者の動作は一見すると類似しているが，熟練者は課題「五十肩の制約あり」で学び直した制御なのに対し，初心者は課題「五十肩の制約なし」で学んだ制御をしかし「五十肩の制約あり」の身体で実行する．どちらも現在は課題「五十肩の制約あり」に直面しているが，そのときに適用する制御が同じ課題で学習されたものか，異なる課題で学習されたものかで異なる．この点で，一見同じ動作を生みだす異なる２つの制御（≒意図）が実現でき，実行結果が制御（意図）に沿うか・沿わないかで成功・失敗を定義する．

変化が系の「構造的複雑さ」の違いとして動作データから検出できると考えた．

この仮説を検討するため，後続の節では，オブザーバの立場から，成功エージェント（成功条件）の動作と失敗エージェント（失敗条件）の動作だけから，その動作データをその背後にある制御の違いと一致するように区別できるかを分析する．

3.2. フラクタル次元

振子の軌道（動作）に見られる「構造的複雑さ」を特徴づけるため，本研究では，単振子と制御器をある種の力学系とみなし，そのアトラクタのフラクタル次元を分析する．フラクタル次元のひとつに点次元 [4] がある．非形式的にいえば，点次元は点集合（たとえばアトラクタの軌道）の各点についてその各点の「自由度」を特徴づける．本稿の仮説に照らせば，可動域の境界（壁）に接触・衝突するという追加の「自由度」がこのフラクタル次元によって検出されるだろうと予想される．著者の一人によって，点次元の推定方法が提案されている（詳しくは [16]）．本稿ではこの点次元推定法を用い，単振子の軌道のデータからそのフラクタル次元を推定する．

データから点次元を推定するときには，オブザーバの観察する軌道のデータ $(x_t, y_t) = (\sin\theta_t, \cos\theta_t)$ に対し，その 10 次元の埋め込み座標系 $(x_t, y_t, x_{t+1}, y_{t+1}, \ldots, x_{t+9}, y_{t+9})$ を用いた．力学系の理論では，滑らかな力学系の場合には埋め込み座標系をとることで低次元の観察からでも本来の高次元のアトラクタの構造を再構成できることが知られている [17]．本稿では埋め込み次元を 10 としたが，他の埋め込み次元でも定性的に類似の分析結果をえており，一定以上の埋め込み次元であれば，同等の結果が再現可能だと考えている．

3.3. 制約振子の軌道とフラクタル次元

最適制御理論の知見 [15] から，倒立振子課題の成功条件では，成功エージェントの制御が生みだす動作は制約振子に対して最適化されており，「なめらかで」自由度の低い（次元の低い）軌道を生成すると予想される．他方，失敗条件では，失敗エージェントの制御が生みだす動作は制約振子に対して最適化されておらず，「ぎこちなく」自由度の高い（次元の高い）軌道を生成すると予想される．

図 2 は，成功エージェント（成功条件）と失敗エージェント（失敗条件）それぞれの制御する制約振子の軌道と，その軌道から計算された点次元の時系列を示す．初期条件は一致させてある．制約振子の軌道はエージェント間（条件間）で見かけ上類似しており，軌道（動作）を観察するだけでは一見区別がつかない．しかし，図 2 から，軌道のデータから推定されたフラクタル次元を比較するとより区別がつきやすい．成功エージェントの動作は，約 2000 ステップで最大振り幅に到達した以降，フラクタル次元が平均的に低い．他方，失敗エージェントの動作は，約 2000 ステップ以降，逆にフラクタル次元が平均的に高い．この結果は，制御の相違がその制御で生成される軌道から推定されたフラクタル次元の相違として検出されるという仮説を支持する．また，最適な制御器の生成する軌道はフラクタル次元が低く，他方，準最適な制御器の生成する軌道は相対的にフラクタル次元が高いという予想とも一致している．

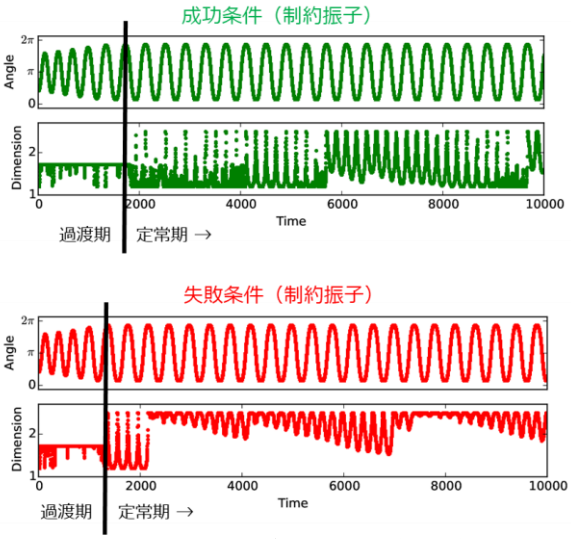

図 2. 実演時の典型的な角度の時系列（Angle）とフラクタル次元の時系列（Dimension）．本研究で設定した条件では，成功条件と失敗条件が類似しているため，角度（Angle）を一見しただけでは区別がつかない

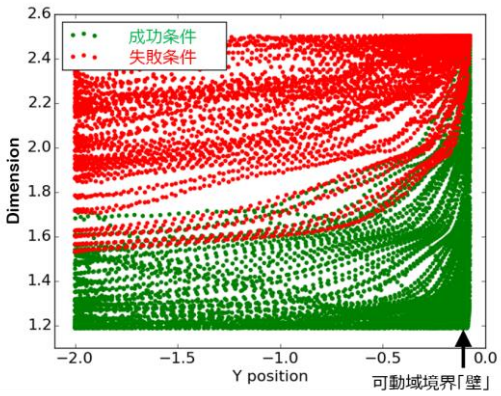

図 3. 振子頭部の Y 座標（Y position）とフラクタル次元（Dimension）の関係

成功エージェントと失敗エージェントの制御の違いは，失敗エージェントの制御が可動域制限なしを前提としたものであることから，制約振子の可動域境界（壁）付近にあると考えられる．

図 3 には，図 2 と同一の時系列データを用い，制約振子の高さ方向の位置 $\cos\theta - 1$ の関数としてフラクタル次元を図示した．可動域の境界の Y 座標は $\cos\theta - 1 = -0.076$ である．図から，図 2 と同様に，失敗エージェントの制約振子の軌道のほうが成功エージェントの制約振子の軌道よりも平均的なフラクタル次元が高いことがわかる．また図から，どちらのエージェントの制御でも可動域の境界付近ではフラクタル次元が上昇している．この結果から，フラクタル次元は制約振子の可動域境界（壁）に対する衝突を捉えていると考えられる．

制約振子の制御では，可動域境界に達すると力学的エネルギー（とくに運動エネルギー）が減少し，可動域境界を離れたあとは次の振り上げに向けて運動エネルギーを補充するといったパターンが予想される．このパターンから，制約振子の軌道

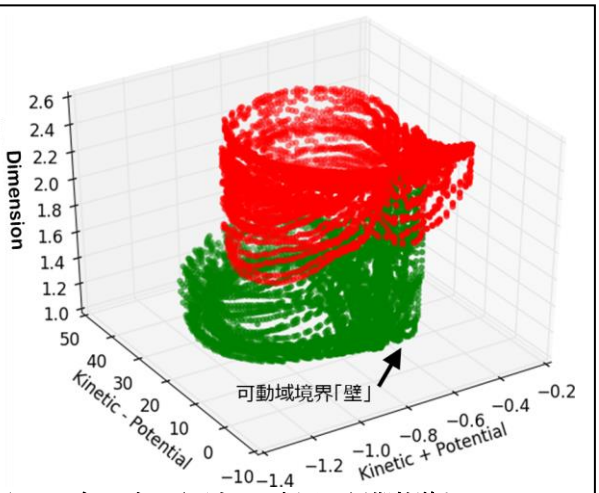

図 4. エネルギー平面上での振子の周期軌道とフラクタル次元の関係．底平面の軸は運動エネルギー（Kinetic）と位置エネルギー（Potential）の差と和

とフラクタル次元の関係は，振子の状態を記述する量である力学的エネルギーを経由して捉えられると考えられる．力学的エネルギー E は運動エネルギー K と 位置エネルギー U に分解できる．図 4 には，可動域境界（壁）と振子の相互作用を含めて振子の周期軌道を視覚的に捉えることが容易な平面の一例として，$K + U$ と $K - U$ の平面上にフラクタル次元を図示した．周期軌道を捉える平面は $K + U$ と $K - U$ に限らないが，最低 2 次元の線形独立な空間の一例を選んだ．点の色は図 3 と同じくエージェントの違いを表す．図から，両エージェントの制御する制約振子は運動エネルギーと位置エネルギーの平面上で周期的な軌道（エネルギー平面上でのアトラクタ）を描いており，可動域境界（壁）に衝突する付近で次元が高まることがわかる．この結果から，フラクタル次元が振子の力学的エネルギーの変化に関して重要な特性を捉えていると考えられる．

3.4. 動作を生成した未知なる制御の識別

前節までで，未知なる制御の違いが動作から推定されたフラクタル次元の違いとして観察できることを定性的に確認した．本節では，オブザーバの立場から，成功エージェント（成功条件）の動作と失敗エージェント（失敗条件）の動作だけから，その動作の集合をその背後にある制御器と一致するように分類できるかを定量的に分析する．具体的には，角度，角速度，周波数，力学的エネルギー，フラクタル次元という 5 種類の特徴量を用意し，これらの特徴の違いによる分類性能（正答率）を調べる．単振子に関しては XY 座標と角度は 1 対 1 対応しており，分類課題の特徴量として本質的な違いはないと考えられるので今回は角度だけを用いた．力学的エネルギーは振子の状態を特徴づける直接的な特徴量であり，前節の結果（図 4）からも高い分類性能を与えることが予想される．しかし，オブザーバの立場から言えば，力学的エネルギーは振子の運動方程式および振子の身体パラメータ（質点の位置，質量，支柱の長さなど）を既知としないとえられない．他方で，フラクタル次元は振子の軌道のデータ（XY 座標の時系列）から計算でき，振子の運動方程式等の知識を前提としない．Warneken & Tomasello [3] の心理実験の子供は後者の「無知な」オブザーバ

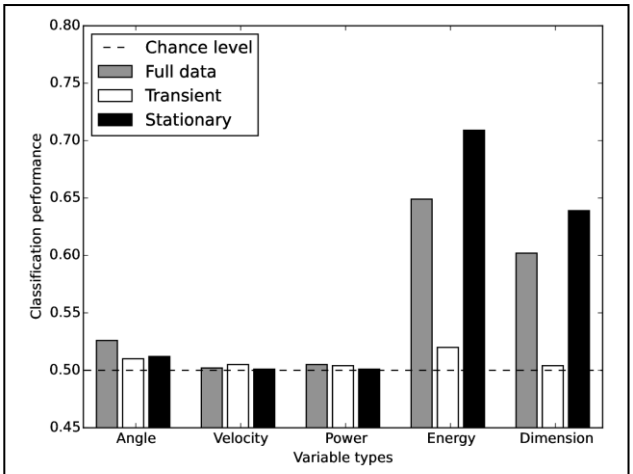

図 5. 動作を生成した未知なるエージェントタイプ（異なる制御）の識別正答率．5 種類の特徴量を用いたときの正答率を比較した．角度（Angle），角速度（Velocity），周波数（Power），力学的エネルギー（Energy），フラクタル次元（Dimension）

に相当すると考えられる．本節では，観察対象に関して無知なオブザーバであっても，どの程度まで動作のデータから未知なる制御の違いを識別できるかに関心がある．そこで，実演者の身体パラメータなどの知識の必要な力学的エネルギーを用いて制御器の違いを識別する「博識な」オブザーバ（すなわち，識別精度の上限を与える）と対比して，無知なオブザーバがどの程度まで博識なオブザーバに迫る識別精度をもつかを調べる．

制御器の識別には正規混合モデル（Gaussian Mixture Model）を用いた．正規混合モデルはデータの密度分布を複数の正規分布の混合で近似する方法である．本数値実験は制御器が確定したあとは決定論的であるため，初期角度の範囲を等間隔に 30 分割し，その30 点の初期値から生成された時系列 30 本を用いる．2 条件 × 時系列 15 本分の点をラベル付き訓練データとし，同じく 2 条件 × 時系列 15 本分の点を試験データとして分類させ，尤度最大のエージェントタイプ（成功エージェントか失敗エージェントか）を選ばせてその正答率を求めた．正規混合モデルの分布数はベイズ情報量基準で最適なものを選んだ（分布数は，フラクタル次元の場合約 5 〜10，他の特徴量の場合約 5〜20 であった）．訓練・試験データには，可動域境界（壁）に最初に到達する前後（約 2000 ステップ）で，過渡期と定常期に分け，フルデータ（過渡期＋定常期），過渡期のみ，定常期のみ，の 3 種類を用意した（図 2 参照）．

図 5 は各特徴量（角度，角速度，周波数，力学的エネルギー，フラクタル次元）を用いた場合の分類の正答率を示す．特徴量としての「周波数」には角度の時系列をフーリエ変換してえた周波数の時系列を用いた．図から，角度，角速度，周波数では，フルデータ，過渡期のみ，定常期のみによらず，ほぼチャンスレベル 50% である．力学的エネルギーでは，定常期のみの場合には約 71% の正答率である．この結果は，力学的エネルギーのように，振子の知識を前提とした特徴量を用いれば（博識なオブザーバ），動作の背後にある制御の違いを識別できることを示唆する．一方で，フラクタル次元では，力学的エネルギーほどではないが，定常期のみの場合には約 64% の正答率で

ある．この結果は，フラクタル次元という動作のみから計算可能な特徴量に着目すれば，振子の知識を前提としなくとも（無知なオブザーバ），動作の背後にある未知なる制御の違いを識別できることを示唆する．

過渡期のみの場合は，どの特徴量を用いた場合でも概ね正答率 50% 程度に留まる．これは，課題の性質上，成功条件（成功エージェント）と失敗条件（失敗エージェント）の違いが基本的には可動域境界以上の振り上げ角度にあり，可動域境界に到達する前の過渡期では動作の違いがほとんどないことを示している．過渡期の動作に関しては振子の知識を前提とした力学的エネルギーでも制御の違いを識別できないので，振子の知識を前提としないフラクタル次元で制御の違いを識別できないことは妥当な結果だと考えられる．

4. 議論

本稿では，目的や意図に関する事前知識を所与とせずに，いかにして一見よく似た動作の背後にある意図（制御）の違いを識別できるかを扱った．本稿では，Warneken & Tomasello [3] の目的推定の心理実験に着想をえた数値実験を設計し，制御器の学習時と実演時の課題や制約を変えることで一見類似した動作を生みだす異なる２つの制御を実現した．その実験設計の下，実演時の課題に最適な制御（意図）と最適でない制御（意図）をそれぞれ目的達成（成功条件）と目的未達成（失敗条件）とみなし，一見類似した動作データからその背後にある目的の成功・失敗の違い，すなわち制御（意図）の違いを識別できるかを調べた．

本稿では，上記２種類の制御の違いがそれらの制御で生成された動作（軌道）のみから推定されるフラクタル次元（自由度を特徴づける）の相違として検出できる，という仮説を検討した．数値実験の結果，フラクタル次元という力学系の不変量を特徴とした場合，振子の知識を前提としなくとも，高い正答率で制御の違いを識別できることが示された．フラクタル次元を特徴とした場合の正答率は，振子の知識を前提とする力学的エネルギーを特徴とした場合に匹敵する．この結果は，力学的不変量であるフラクタル次元に着目することで，ある種の意図推定がほとんど事前知識なしに身体動作の観察から可能であることを示唆する．

本稿では２つの問題をあげた．第一は，観察不能な変数の存在であり，第二は目的と動作の多対多対応である．第二の問題は実験設計上，複数の目的に対して一見類似した動作がある状況（目的対動作で多対一）を作りだすことで，その場合を検討した．第一の問題は，本稿の数値実験では各成功・失敗エージェント（実演者）がその制御器に用いる状態変数（角度と角速度）を，フラクタル次元を用いるオブザーバ（観察者）が直接観察できないことで検討された．フラクタル次元を用いるオブザーバが観察したのは振子頭部の XY 座標データであり，角度や角速度の情報は抜け落ちている．本数値実験では，こうした観察不能な変数の存在にもかかわらず，フラクタル次元を用いるオブザーバは XY 座標データから未知なる制御器の違いを識別できることを示した．これが可能であるのは，フラクタル次元が滑らかな変換（観察）に対する不変量であり，表面上のデータ表現の違いに不変な特徴を捉えているためだと考えられる．

本稿の範囲は意図（制御）の識別に留まるが，本稿の成果は目的の推定や行為の模倣あるいは真の模倣 [4] に向けた第一歩といえる．詳細は別の論文 [18]に譲るが，著者らは人間の被験者に対する行動実験で，フラクタル次元を用いた投球動作の特徴づけで，熟練者と初心者（運動制御の最適性に違いがあるとみなす）を識別できる可能性を示す結果をえている．今後は，本研究の拡張として，意図の推定と合わせてオブザーバによる目的の推定や動作の生成を検討していくほか，人間の行動データにも本研究のアイデアが適用できることを実証していきたい．

謝辞

本研究は科学研究費補助金 JP16H05860，JP17H06713 の助成を受けて行われた．

参　考　文　献

[1] M. Tomasello, The Cultural Origins of Human Cognition, Harvard University Press, 2001.

[2] N. A. Bernstein, Dexterity and Its Development, M. L. Latash and M. T. Turvey, Eds., Psychology Press, 1996.

[3] F. Warneken, M. Tomasello, "Altruistic helping in human infants and young chimpanzees," *Science*, vol. 311, pp. 1301-1303, 2006.

[4] C. Breazeal, B. Scassellati, "Robots that imitate humans," *TRENDS in Cognitive Sciences*, vol. 6, no. 11, pp. 481-487, 2002.

[5] 石川友紀也, 末松圭史, 大坪正和, 吉田香, "上半身動作を用いたロボットの回避方向意図の伝達," *知能と情報*, vol. 30, no. 5, pp. 709-716, 2018.

[6] 谷嵜悠平, ジメネス・フェリックス, 吉川大弘, 古橋武, "教育支援ロボットにおける身体動作と表情変化による共感表出法の印象効果," *知能と情報*, vol. 30, no. 5, pp. 700-708, 2018.

[7] 河合祐司, 長井志江, 浅田稔, "視覚の精緻化が導くミラーニューロンシステムの発達モデル," in *第29回日本ロボット学会学術講演会予稿集*, 2011.

[8] 小嶋秀樹, "発達ロボティクスからみた模倣とコミュニケーションのなりたち," *バイオメカニズム学会誌*, vol. 29, no. 1, pp. 26-30, 2005.

[9] 長井志江, "認知発達の原理を探る：感覚・運動情報の予測学習に基づく計算論的モデル," *ベビーサイエンス*, vol. 15, pp. 22-32, 2015.

[10] D. Marr, Vision, Cambridge: MIT Press, 1982.

[11] 川人光男, 脳の計算理論, 産業図書, 1996.

[12] K. J. Astrom, K. Furuta, "Swinging up a pendulum by energy control," *Automatica*, vol. 36, no. 2, pp. 287-295, 2000.

[13] K. Doya, "Reinforcement learning in continuous time and space," *Neural Computation*, vol. 12, pp. 243-269, 1999.

[14] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, MIT Press, 1998.

[15] E. Todorov, M. I. Jordan, "Optimal feedback control as a theory of motor coordination," *Nature Neuroscience*, vol. 5, pp. 1226-1235, 2002.

[16] S. Hidaka, K. Neeraj, "On the estimation of pointwise dimension," *ArXiv:1312.2298*, 2013.

[17] L.-S. Young, "Dimension, entropy, and Lyapunov exponents,"

Ergodic Theory and Dynamical Systems, vol. 2, no. 1, pp. 109-124, 1982.

[18] 鳥居拓馬 , 日髙昇平, “利き手と逆の手の比較に基づく熟達技能への実験的アプローチ (大会発表賞受賞論文) ,” *認知科学*, vol. 25, no. 1, pp. 122-125, 2018.

[19] I. Grondman, M. Vaandrager, L. Busoniu, R. Babuska, E. Schuitema, "Efficient model learning methods for actor-critic control," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 3, 2012.

[20] S. Schaal, "Learning from demonstration," M. C. Mozer, M. Jordan and T. Petsche, Eds., Cambridge: MIT Press, 1997, pp. 1040-1046.

[21] C. D. Cutler, "A review of the theory and estimation of fractal dimension," vol. 1, World Scientific, 1993, pp. 1-107.

[22] E. Todorov, "Optimality principles in sensorimotor control," *Nature Neuroscience*, vol. 7, pp. 907-915, 2004.

（****年**月**日　受付）
（****年**月**日　採録）

[問い合わせ先]
〒923―1292　石川県能美市旭台1-1
北陸先端科学技術大学院大学
鳥居 拓馬
TEL：0761―51―1728
E-mail: tak.torii@jaist.ac.jp

付録

A. 強化学習を用いた制御器の最適化

強化学習 [14] は行動心理学や制御理論に起源をもつ枠組みである. 強化学習の学習者は状態 s で行動 a を選び, その結果, 報酬 r と次の状態 $s' = P(s'|s)$ をえる．ここで，P は状態遷移関数である. 学習者の課題は報酬の累積値を最大化するような制御器 $g(a|s)$ を見つけることである.

倒立振子課題は連続空間・連続時間を扱う古典的な制御課題である [13]. この課題を効率的に解くアルゴリズムは近年でも開発されている [19]. 初期のアルゴリズムはアクター・クリティク法 [14] と呼ばれ，制御器 $g(a|s)$ を表すアクターと価値関数 $V(s)$ を表すクリティクを異なる部品としてもつ.

連続的な空間と時間を扱う課題では工学的に厄介な問題がある. 典型的な解法は連続的な空間と時間を離散化して近似する方法である．本稿では，時間に関しては，オイラー積分（ステップ幅 $dt = 0.01$）を用い，制御命令を 3 ステップ間隔で受けた．空間に関しては，タイルコーディング [14] と呼ばれる連続的な状態空間 $(\theta, \dot{\theta}) \in [-\pi, \pi] \times [-2\pi, 2\pi]$ を等間隔に 40×40 分割したもの（各セルがひとつの状態 s に対応する）を採用した.

強化学習は，状態価値 $V(s)$ を推定し，状態価値の高い状態への訪問確率を高めるように制御器 $g(a|s)$ を更新することで進行する．いま制御対象である単振子が状態 s にあり，制御器が制御命令 u を平均 μ_s と分散 σ_s^2 の正規分布 $u = N(\mu_s, \sigma_s)$ に従って与え, その結果, 報酬 r と次の状態 s' を受けたとする．そのとき，学習者（クリティク）は価値関数 $V(s)$ を次の更新量だけ増やす（TD 誤差）.

$$\Delta V(s) = \alpha_c \left[r + \gamma_c V(s') - V(s) \right] E_c(s)$$

ここで，$\alpha_c = 0.1$ は学習率，$\gamma_c = 0.97$ は割引率，E_c は適格度トレース（減衰率 $\lambda_c = 0.65$）といい，直近の状態ほど高い荷重を割り当てる，連続状態での学習を促進する工夫である.

連続値の制御命令を扱うため, 学習時には制御器は状態 s に依存した正規分布 $u = N(\mu_s, \sigma_s)$ に従う制御命令を出す．制御器の学習では状態毎に平均 μ_s と分散 σ_s^2 を調整する. 形式的には，学習者（アクター）は正規乱数の勾配に基づき，各パラメータを次の更新量だけ増やす（方策勾配降下法）.

$$\Delta\mu_s = \alpha_a \left[r + \gamma_a V(s') - V(s) \right] \frac{\partial N(\mu_s, \sigma_s)}{\partial \mu_s}(u)\, E_a(s)$$

$$\Delta\sigma_s = \alpha_a \left[r + \gamma_a V(s') - V(s) \right] \frac{\partial N(\mu_s, \sigma_s)}{\partial \sigma_s}(u)\, E_a(s)$$

ここで，$\alpha_a = 0.001$ は学習率，$\gamma_a = 0.65$ は割引率，E_a は適格度トレース（減衰率 $\lambda_a = 0.0$）である.

報酬関数 $r = f(s, a)$ は課題の目的を指示するよう注意深く設計する必要がある．先行研究 [13] に従って，本研究では $r = \cos\theta$ を用いた．この報酬関数は倒立振子課題の目的を指示できる．振り上げた状態 $\theta = 0$ のときの報酬 $\cos 0 = 1$ は最大となり，他方，吊り下げた状態 $\theta = \pm\pi$ のときの報酬 $\cos\pi = -1$ は最小となる． 強化学習は累積報酬 $\sum_t \cos\theta_t$ を最大化するので,「素早く倒立姿勢まで振り上げてその姿勢を維持すること」を指示できる.

著　者　紹　介

とりい　たくま
鳥居　拓馬［非会員］

2014年北陸先端科学技術大学院大学博士課程単位取得退学．2016年博士（知識科学）．2012年ハートフォードシャ大学にて訪問研究員．2014年東京大学大学院工学系研究科にて研究員．現在，同大学助教．日本認知科学会，人工知能学会会員．人の意味認知とくに計画・意図に関する研究に従事.

ひだか　しょうへい
日髙　昇平［正会員］

2007年京都大学大学院情報学研究科博士課程修了．博士（情報学）．同年インディアナ大学にて博士研究員．2010年北陸先端科学技術大学院大学助教．現在，同大学准教授．言語・認知発達，意味認知の計算論的メカニズムを解明することを目的に，心理学実験・情報理論・機械学習・非線形時系列解析などを駆使した研究を行う．日本認知科学会，人工知能学会など会員.

Modeling intention inference as a basis of imitation learning: analysis of pendulum swing-up task

by

Takuma TORII, and Shohei HIDAKA

Abstract:

Learning an action from others require to infer their underlying intentions. Psychological studies have reported behavioral evidences that young children do infer others' underlying intentions by observing their actions. The objective of the present study is to propose a mechanistic account for how intention inference is possible by observing others' actions. For this purpose, we performed a series of simulations in which two agents control pendulums for different tasks and goals, and analyzed which types of features is informative to infer their latent intentions. Our analysis showed that a type of fractal dimension of the pendulum movements is sufficiently informative to classify the types of agents. With respect to its invariant nature, our results suggest that the fine-grained movement patterns such as the fractal dimension reflect the structure of the underlying intentions.

Keywords: imitation learning, intention inference, dynamical systems, pendulum swing-up

Contact Address: Takuma TORII

Japan Advanced Institute of Science and Technology

1-1, Asahidai, Nomi, Ishikawa, 923-1292, Japan

Phone : 0761-51-1728

E-mail: tak.torii@jaist.ac.jp