実機とシミュレータを用いたオンライン複数人教示手法の提案

生和　太郎†　片上　大輔†　新田　克己†

† 東京工業大学大学院 総合理工学研究科 〒 226-8502 神奈川県緑区長津田町 4259

あらまし　ロボットに同一のタスクに対して複数の方針を与えることによって，ロボットの身体性，環境に対し最適な方針を選択する手法を提案することを目的とする．本研究では，複数人による教示を利用し，教示によって得た教示データをクラスタリングすることにより，同一の方針の教示をクラスタ化し，ロボットの行動の評価値を基に各ロボットの身体性や環境に最適な方針を選択するシステムを考案する．従来は，時系列データで構成される教示データの波形の類似度を基にクラスタリングを行っていたが，さらに身体性や環境の評価値を含むことで，複数の方針からロボットが自己の身体性や環境を基準に，最適な方針の選択が可能になる．実験では，各評価値を含めたクラスタリングを行うことにより，各クラスタ内の教示データを平均化した代表教示が改善されることを確認した．

キーワード　複数人教示，直接教示，DP マッチング，クラスタリング，シミュレータ Webots

A Proposal of Online Multiple Human Teachings System with Real Robot and Simulator

Taro NYUWA†, Daisuke KATAGAMI†, and Katsumi NITTA†

† Tokyo Institute of Technology，4259 Nagatsuta-cho, Midori-ku, Yokohama, 226-8502, Japan

Abstract　The proposal of this research is to develop a system to select the best policy for an embodiment and an environment of a robot by giving two or more policies to the same task. In this research, we develop a system that selects the best policy based on the evaluation value of an action of a robot. In this system, the representative instruction data of the same policy is made a cluster by clustering based on multiple human teaching. The conventional method clusters based on the degree of similarity of waveform instruction data, which composed by the time series data. The proposed method enables to select one's own best policy by including the evaluation values based on the embodiment and the environment of each robot in addition the degree of similarity of data. In the experiment, we confirmed that the representative instruction data in each cluster was improved by clustering with the adding evaluation values.

Key words　multiple human teaching，direct teaching，DP matching，clustering，simulator Webots

1. はじめに

近年，人間と同じ環境で活躍するロボットの普及が高まっている．このようなロボットは，様々な環境に適応した行動が必要とされる．ロボットに行動を学習させるに当たり，ロボットを学習者と見立て人間が教示を与えて行動を学習させる，というアプローチが有用とされている．現在，このような研究の多くはインタラクションする人間は 1 人である．しかし，実環境ではロボットがインタラクションする人間は必ずしも 1 人とは限定できなく，複数の人間とインタラクションする場合が十分に考えられる．1 人のみで教示行動行った場合，学習した行動が偏ってしまい，様々な環境には適応できない．様々な環境に適応した行動をとるためには，複数の方針を持つことで環境に合わせて方針を選択する方法が有用であると考えられる．そこで，複数の方針が存在する可能性のある複数人による教示に着目した．

小竹ら [1] は，同一目的に対し様々な方針が存在する可能性のある複数人による教示 (以降複数人教示と呼ぶ) を利用し，方針を競合することなく同一の目的に対し複数の方針を持つことによって，身体性が変化した際にその変化に適した方針の選択が可能であることを示した．

しかし，この研究では，複数の教示データから同一の方針をクラスタ化する際に教示データの時系列の形のみを基にクラスタリングを行っていたため，同クラスタ内での評価値の分散が大きくなっていた．また，オフラインの直接教示を取り扱っており，人間側が一方的に教示を行うのみであり，教示行動の評

価を行う判断基準が無く，教示中に行動の修正を行う事が出来なかった．

そこで，本研究では複数人からの教示データが与えられた際に複数の方針を得て，そこからロボットの身体性，環境に最適な方針を選択することを目的とする．そのために，評価値を含めたクラスタを構築する，シミュレータ Webots を用いた教示手法を提案する．

本論文では，第 2. 章にて関連研究について言及し，第 3. 章にて複数人教示に対応できる手法を提案する．第 4. 章にて提案手法をシステムに実装して行った実験概要と実験結果について述べ，第 5. 章にて結論をまとめる．

2. 関 連 研 究

本研究で達成しようとする同一目的に対し複数の方針が存在する場において，最適な方針を環境や自身の特徴に合わせ最適な方針を選択するという目的の背景には，知的な振舞をロボットに発現させようという目標が存在する．

複数の教示を統合する研究では，スキルサイエンスとして知られる研究の分野がある．これは，複雑な動作をロボットに学習させるために，複数回に渡り教示動作を与え重要な部分を抽出し汎化することを目的としている．Terada らは人間とほぼ同じサイズのヒューマノイドロボットに，起き上がり動作をさせている [3]．この研究では，教示者が複数回にわたり起き上がり動作をモーションキャプチャでサンプリングし，人間が手動でロボットへの制御動作を生成させている．Ogawara らはマニピュレータロボットに，人間からの複数回の直接教示により目的のタスクを達成する動作を汎化し獲得させている [4]．この研究ではタスクを自動的にサブタスクへと分割し，DP マッチングにより冗長な要素を省き必須要素のみを抽出している．いずれの研究においても教示は全て同じ方針で目的を達成すると設定しており，本研究で想定する複数の方針が存在する場に適用するのには不十分である．

人間が目的を達成するための動作を学習していく 1 つの過程として挙げられる見真似を模した学習方法では，Takano らの研究 [5] と杉本らの研究 [2] をここに挙げる．Takano の研究では，挙手や歩行等の 5 種類の動きを各 10 個ずつ用意し，50 の動作を教示として与え，5 種類の動作へと隠れマルコフモデルを利用し運動パターンの記号化を行いロボットに動作を獲得させている．しかし，この研究では教示動作が複数の目的で用意されていることと，運動パターンの数が既知であることを条件としているので，本研究が想定する同一目的に対し複数の方針が存在する場では方針の数が未知のために適用することが難しい．杉本らの研究では，教示者と学習者の間で身体差がある条件の中で，見真似学習により単振り子の振り上げ課題を成功させている．本研究では学習を焦点に当てるわけではなく，学習者の身体性の差により様々な最適行動が現れることを示すので，通常の見真似学習の視点とは異なっている．

3. 複数人教示システム

同一目的に対し複数の方針が存在しうる複数人教示では，従来研究のような全ての教示が単一の方針であると仮定しているような手法では対応できない．そこで本研究では，複数人教示によって得た教示データ（図 2）をクラスタリングすることにより，同一の方針の教示をクラスタ化し，異なる方針の教示は異なるクラスタに属させるということを試み，この問題点を解決する．また，クラスタリングすることにより全体的な教示数を圧縮し，最適な教示を選択する段階において早い選択を可能にすることも視野に入れる．

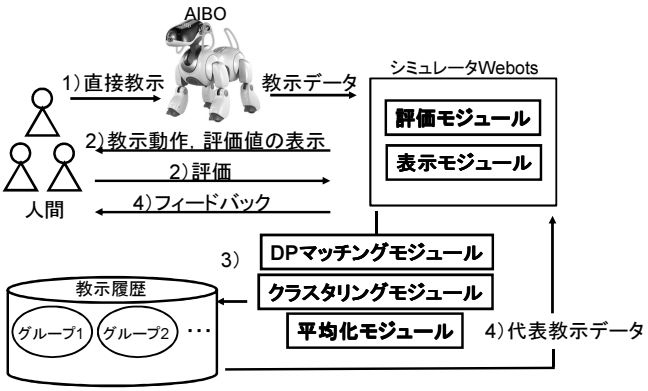

図 1　システム概要

また，教示データに対し評価値を与えることで，教示データ間の形の類似度と評価値を含めた類似度を基にクラスタリングを行う．これにより，データの形のみでなく評価値を含んだクラスタを生成する．複数の方針からロボットの身体性や環境に対し，評価値を基にした方針の選択が可能になる．

3.1 システムの概要

提案するシステムの概略図を図 1 に示し，以下に説明を行う．

1）複数人により，ロボットに対し教示を与える．本研究では教示する対象として動作を設定し，教示方法として直接教示を採用する．教示の対象として動作を扱う場合のほとんどは，時系列データを扱うことになる．

2）シミュレータ Webots では，教示データを入力する．シミュレータ Webots は評価モジュールと表示モジュールの二つのモジュールを含む．評価モジュールでは教示データを入力し，シミュレータ上で教示データの再現を行うことでセンサ情報を取得し，前進距離，衝撃値といった各教示データに対する評価値を算出する．また，表示モジュールではシミュレータ Webots の画面上に実機の教示時の行動の表示を行うことにより，教示者から評価を受ける．

3）DP マッチングモジュールでは，複数人からの教示を入力する．一般に人間が同一だと考える時系列データを与える場合には形と値が重要視され，時間に依存するタイミングや静止時間等の差については重要視されないことが多い．よって，人間が重要視していない時間に依存する誤差については最小化し，人間が重要視する形と値に関しての差を考えなければいけない．そこで，その問題を解決するために DP マッチングモジュールにて DP マッチング (動的計画法) を利用する．

クラスタリングモジュールでは，DP マッチングモジュールの出力である教示データの形の類似度と，シミュレータ Webots

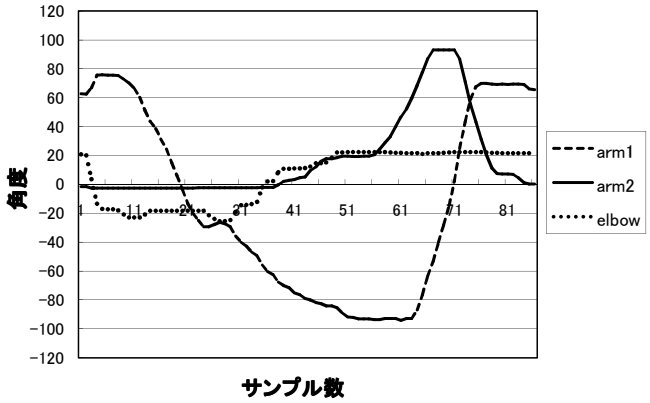

図 2　前足での前進運動を教示した際の教示データ

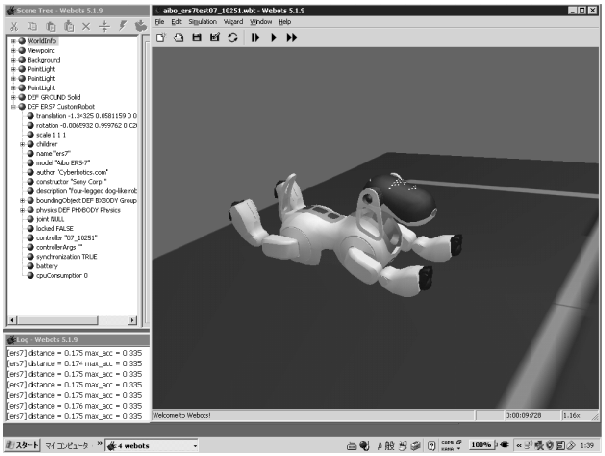

図 3　シミュレータ Webots

の出力である各教示データの評価値を入力とする．このモジュールでは形の類似度と評価値を基に各教示データ間の類似度を算出し，クラスタリング手法を用いその類似度から各教示データをクラスタ化し出力する．

平均化モジュールでは，クラスタリングモジュールによって出力した教示群をクラスタごとに平均化し統合する．このモジュールでは時系列平均化により各教示者特有の癖やノイズを除去し，教示数削減のために各クラスタごとに代表教示データを出力する．

4）平均化モジュールで作成した代表教示データをシミュレータ Webots で表示することにより，教示者へのフィードバックを行う．

各モジュールの説明は次節以降にて行う．

3.2　教示データ

教示データは複数次元の時系列データからなる．図 2 に dilect teaching の一例として，AIBO に対し前足のみの前進運動を教示した際の，右前足の教示データを示す．この際，教示データは肩 2 自由度，肘 1 自由度の計 3 自由度の時系列データからなる．この教示データをシミュレータ Webots の評価モジュール，表示モジュール，DP マッチングモジュールへ入力する．

3.3　シミュレータ Webots

シミュレータ Webots は cyberbotics 社が提供する，ロボットシミュレータである．(図 3）khepera や AIBO などのロボットがプロジェクトとして組み込まれており，シミュレータ上の動きを実機で再現するクロスコンパイルの機能を持つ．

評価モジュールでは，教示データを入力することで，シミュレータ Webots 上で教示行動の再現を行う．同時にシミュレータ上で得られるセンサ情報からその教示行動に対する評価値（前進距離 E_{Fd}，衝撃度 E_{Is}）を算出する．前進距離は GPS データから算出する．この GPS データは実機にないセンサ情報であり，シミュレータ Webots は実機では得られないセンサ情報を扱うことができる．衝撃値は教示中で最大の加速度と定義する．

3.4　DP マッチングモジュール

本節では DP マッチングモジュールにて用いる DP マッチングを説明する．

DP マッチングは，時系列データにおける値と形が重要視される場合に利用される類似度計算法である [6]．この手法は 1 つの点を複数の点に対応させる，時間軸非線形伸縮を特徴とする．

まず，2 つの時系列データ A，B を時間軸に沿ってサンプリングする．

$$\begin{aligned} A &= \{a_n | n = 0, 1, \cdots, I\} \\ B &= \{b_n | n = 0, 1, \cdots, J\} \end{aligned} \tag{1}$$

次に，a_n, b_n からなる 2 次元平面を考え，ワーピングパスで両データの点を対応付けていく．つまり，式 (2) のように定義することにより，$f = (i, j)$ の系列で格子点の移動が表現できる．

$$\begin{aligned} F &= f_1, f_2, \cdots, f_k, \cdots, f_K \\ f_k &= (i_k, j_k) \\ \forall k_1 < k_2 &\Rightarrow (i_{k_1} \leqq i_{k_2}) \wedge (j_{k_1} \leqq j_{k_2}) \end{aligned} \tag{2}$$

また，極端な時間伸縮を避けるために，整合窓を設定することが多い．本研究でも整合窓を設定し，データ長の何割まで伸縮できるかを設定している．図 4 に DP マッチングのワーピングパスと整合窓の例を示す．

ここで，データ間の格子点上での距離を式 (3) で定義すると，これを総和することによりデータ間のユークリッド距離を式 (4) のように求めることができる．この $D(A, B)$ を最小にするような時間伸縮関数をトレースバック手法により求め，式 (5) に示した $D(A, B)$ を時間伸縮関数の長さで割り正規化したものをデータ A とデータ B の形の類似度とすることができる．これを DP マッチングモジュールの出力とする．

$$d(f_k) = (a_{k_i} - b_{k_j})^2 \tag{3}$$

$$D(A, B) = \sum_{k=1}^{K} d(f_k) \tag{4}$$

$$\hat{D}(A, B) = \frac{D(A, B)}{K} \tag{5}$$

3.5　クラスタリングモジュール

本節ではクラスタリングモジュールにて用いるクラスタリングについて説明する．

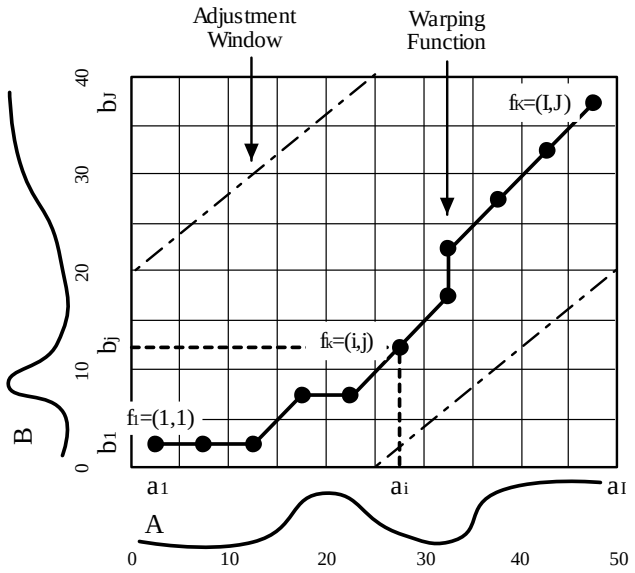

図 4　DP マッチングにおけるワーピングパスと整合窓

本研究では階層的手法を用いる．ここで，階層的手法によるクラスタリングを説明する．まず，1 個の対象のみを含むクラスタを n 個作成する．その後，各クラスタ間の類似度を距離関数により全て算出し，最も類似度が高いクラスタを 1 つのクラスタに統合する．統合してできたクラスタと，他のクラスタの類似度を再び全て計算し，類似度が最も高かったものを統合させる．これをクラスタが 1 つになるまで，再帰的に繰り返すのが階層的手法によるクラスタリングである．

教示データ間の類似度は式 (6) によって求める．類似度を $S(A,B)$，DP マッチングモジュールの出力である形の類似度 $\hat{D}(A,B)$ を正規化した値を $\hat{D}'(A,B)$，シミュレータ Webots の出力である評価値を基に算出した前進距離の差を正規化した値を $E'_{Fd}(A,B)$，衝撃値の差を正規化した値を $E'_{Is}(A,B)$，各評価値に付与する重みを α と置く．

$$S(A,B) = \hat{D}'(A,B) + \alpha E'_{Fd}(A,B) + (1-\alpha)E'_{Is}(A,B) \quad (6)$$

階層的手法のクラスタリングの種類として，統合して生成されたクラスタと他のクラスタとの距離を計算をする距離関数で様々なパターンが存在する．本研究では，統計的な処理でよく利用されている群平均法 (group average method) を距離関数で用い，クラスタリングを行う．群平均法は 2 つのクラスタに属する対象間のすべての類似度を求め，その平均値をクラスタ間の類似度とする．クラスタ p とクラスタ q を結合し，クラスタ t を作成した場合，任意のクラスタ r との類似度 S_{tr} を考える．各クラスタ間の教示データ数を n_p，n_q，n_r とすると，類似度 S_{tr} は式 (7) から求まる．

$$S_{tr} = \frac{n_p S_{pr} + n_q S_{qr}}{n_p + n_q} \quad (7)$$

3.6　平均化モジュール

クラスタリングモジュールでは類似度の高い教示データをクラスタ化したが，クラスタに格納された教示データは同じ方針で教示を与えているはずである．しかし，同じ方針といえども個人差があるために多少の差や個人の癖があることが予想され，その影響を取り除いてやる必要がある．また，教示数圧縮という観点においても，各クラスタから 1 つの教示データを出力する必要がある．本研究では，各クラスタごとに教示データを平均化することにより，クラスタを代表する教示データである代表教示データを出力する．

一般に時系列データの場合，単純に時間に沿って平均化するのでは直感に反してしまう平均化をしてしまうことが多い．これは第 3.4 節でも述べたが，時系列データを扱う上で厄介な問題である．Yamada や Nakamoto らはその問題に対し，DP マッチングにより時系列データの平均化を行っている [7] [8]．そこで，データの平均化に関しても DP マッチングにより，教示動作データの動的時間伸縮を行い平均化を行う．

クラスタ p,q の平均教示データを $average(p,q)$ とし，各クラスタ内に統合された教示の数を N_p, N_q，教示データを A,B とすると，平均教示データは以下のように求めることができる．ただし，クラスタ内の教示の数が 1 つの場合は，その教示自身を平均教示データとする．

$$average(p,q) = \frac{n_p\ A + n_q\ B}{n_p + n_q} \quad (8)$$

表 1　AIBO ERS-7 における各部位が有する自由度

component	degree of freedom
head	3
four limbs	$3 \times 4 = 12$
mouth	1
ear	2
tail	2
total	20

4.　実　　験

本章では，本提案手法をシステムに実装し，複数人教示をシステムに入力した際にどのように最適な方策を選択するかを示す．

4.1　実験概要

4.1.1　実験に使用するロボット

本実験では，具体的な身体性を有する犬型ペットロボットである AIBO ERS-7 [9] を用いる．

AIBO は表 1 のように，計 20 自由度を持つロボットである．また，OPEN-R という開発環境を用いることで比較的容易に AIBO の挙動などをプログラムにより制御することができる [10]．

4.1.2　実験におけるタスク設定

実験は前節にて述べた AIBO を用い，被験者 40 人に対し AIBO の前足のみで可能な限り前に進むことを目的とする前進運動を考えてもらい直接教示 (AIBO の手を直接掴み動かしてもらう) にて教示をさせた．

また，本実験では AIBO の前進運動において前腕のみを利用するということを条件とした．実際の犬は前足でのみ歩くということはしないが，直接教示で人間が AIBO に教示する場合に

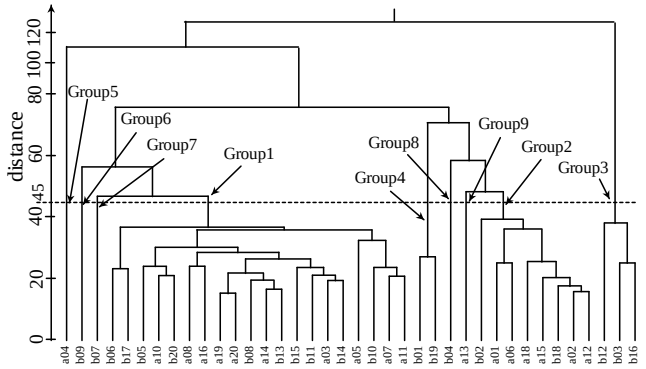

図 5　形の類似度を基にした全体のクラスタリング結果

は，計 12 自由度持つ 4 本の足を 1 人の人間では扱いきれないと判断したためである．このことから，本実験で AIBO が教示される際にサンプリングした関節角度は両前足の 6 つのサーボモータのセンサのみとした．なお，サンプリング間隔は 128ms で実験を行った．

複数人の直接教示によって得られた教示データをシミュレータ Webots 上で再現し，前進距離，衝撃値を評価値として各教示データに与え，形の類似度と評価値を含んだ類似度を基にクラスタリングを行った．各教示データ間の類似度計算の際に用いた式 (6) の重み α を 0.0〜1.0 まで変化させ，デンドログラムを作成した．また，各重み α でできたデンドログラムの中で最大のクラスタの代表教示データを作成した．シミュレータ Webots に代表教示データを入力すし，各代表教示データに評価値を与えることで各重みに対して評価値の比較を行った．

4.1.3　被　験　者

被験者は男子理系学生が 25 名，女子理系学生 3 名，男子文系学生 5 名，女子文系学生 3 名，女性事務 3 名，男子高校生 1 名の計 40 名である．

4.1.4　実験システムの各種設定

実験システムに教示を入力する際には，繰り返し部分における一区間を抜き出し入力とする必要がある．本実験では，左右対称の動作であれば両前足が胸の前にある状態からを一区間と定義し，左右非対称の動きであれば左足が前にある状態からを一区間と定義した．

DP マッチングにおける整合窓は，事前実験にてデータ長の 30%程度で妥当という結果が出ているので，本実験でもデータ長の 30%と設定した．

4.2　実 験 結 果

図 5 に形の類似度のみでクラスタリングを行った結果のデンドログラムを示す．横軸の a01 から b20 までは各教示データ，縦軸の distance は各クラスタの結合距離を表している．図中の group 1 から group 9 は各クラスタを示し，このクラスタ内の教示データを平均化することにより各クラスタの代表教示データを算出する．

図 6 に図 5 のデンドログラムの中で最大のクラスタ group 1 のデンドログラムを示す．各教示データ番号の下に前進距離，衝撃値の 2 つの評価値を示した．また，図 7 に，評価値を含めたクラスタリング結果のデンドログラムを示す．類似度計算の

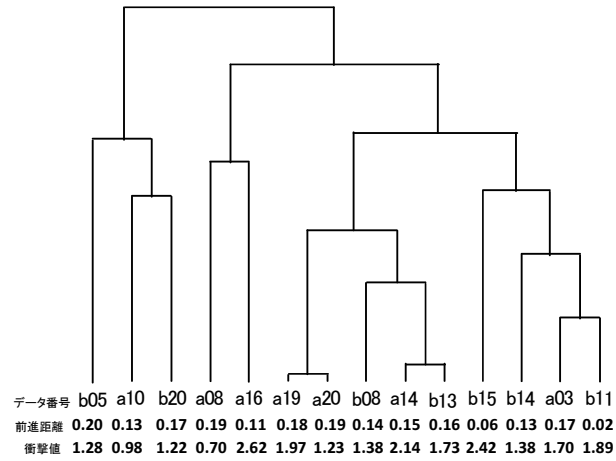

図 6　形の類似度を基にしたクラスタリング結果で最大のクラスタ

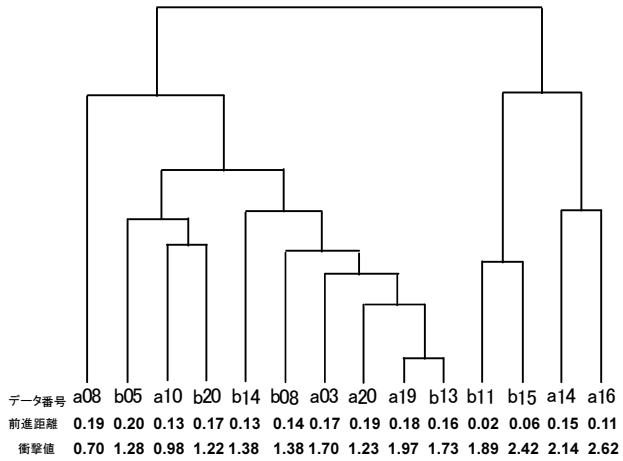

図 7　評価値を含めたクラスタリング結果（$\alpha = 0.3$）

表 2　重み α を変えた際の最大クラスタにおける評価値の比較

	n	E_{Fd}	分散 (E_{Fd})	E_{Is}	分散 (E_{Is})
形の類似度	20	0.141	0.00238	1.35	0.482
α=0.0	13	0.161	0.00135	1.05	0.053
α=0.5	10	0.173	0.00095	1.07	0.067
α=1.0	14	0.166	0.00031	1.27	0.334

際に用いた式 (6) で重み α を 0.3 に設定した．

また，重み α を 0.0，0.5，1.0 に変化させた際に得られた代表教示データのデータ数 n，データに対する評価値 (E_{Fd}，E_{Is})，評価値の分散を表 2 に示す．

4.3　考　　察

4.3.1　クラスタ内の評価値の分散

教示データの形のみを基にしたクラスタリング結果（図 6）では，クラスタ内の評価値の分散が大きく，代表教示データに著しく評価値の低い教示データが含まれてしまう可能性がある．例えば前進距離に関しては，図 6 の b15-b11 のクラスタでは，前進距離の最小値が 0.02，最大値が 0.17 と同クラスタ内で評価値が大きく異なっている．衝撃値に関しては，a08-a16 で全体のクラスタ内の衝撃値の最大値と最小値が含まれているにもかかわらず，類似度が一番高いと判断されている．

これに対し，評価値を含んだクラスタリング結果（図 7）で

は，各クラスタ内の評価値の分散が少なくなっている．前進距離に関しては，評価値が著しく低い教示データ b11 と b15 がクラスタ化され，他のクラスタと分離されている．衝撃値に関しては，図 6 で同じクラスタだった教示データ a08 と a16 が他のクラスタに分かれ，類似度の差が大きくなっている．このように，各クラスタ内での評価値の分散が少なくなっている．

4.3.2 代表教示データの比較

表 2 の各最大クラスタの分散をみると，形の類似度の最大クラスタの分散が前進距離，衝撃値両方に対して一番大きな値をとっている．$\alpha = 0.0$ では衝撃値の分散が最小値，$\alpha = 1.0$ では前進距離の分散が最小値となっているため，重みづけにより各評価値に対して分散を少なくさせるクラスタリングが行われている．また，形の類似度と $\alpha = 1.0$ を比べると，前進距離の値は 0.141 と 0.166 と $\alpha = 1.0$ の方が大きく，分散は 0.00238 と 0.00031 と $\alpha = 1.0$ の方が小さくなっており，評価値は保ったまま分散が少なくなっているため，代表教示データが改善されている．また，$\alpha = 0.5$ では形の類似度に比べると前進距離，衝撃値両方の分散が少なくなっており，α を変化させることで前進距離，衝撃値両方の分散が少ない代表教示データを含むクラスタを作成することができると考えられる．

5. おわりに

本研究はロボットに同一のタスクに対して複数の方針を与えることによって，ロボットの身体性，環境に対し最適な方針を選択する手法を提案することを目的とした．複数人による教示を利用し，教示によって得た教示データをクラスタリングすることにより，同一の方針の教示をクラスタ化し，ロボットの行動の評価値を基に各ロボットの身体性や環境に最適な方針を選択するシステムを考案した．

実験では，各評価値を含めたクラスタリングを行うことにより，クラスタ内での分散が少なくなり，各クラスタ内の代表教示データが改善されることを確認した．

文　献

[1] M.Kotake, D.Katagami, K.Nitta, ”Acquisition of Motion Skills by Multiple Human Teaching” *SCIS ISIS2006*, pp.1048-1053 (2006)

[2] 杉本徳和, 鮫島和行, 銅谷賢治, 川人光男. ”複数の状態予測と報酬予測モデルによる強化学習と行動目標の推定”, 電子情報通信学会論文誌, J87-D- (2), 683-694 (2004)

[3] K. Terada, Y. Ohmura, Y. Kuniyoshi, ”Analysis and Control of Whole Body Dynamic Humanoid Motion – Towards Experiments on a Roll-and-Rise Motion” *International Conference on Intelligent Robots and Systems*(2003)

[4] K. Ogawara, J. Takamatsu, H. Kimura, K. Ikeuchi, ”Extraction of Essential Interactions Through Multiple Observations of Human Demonstrations” *IEEE Trans. on Industrial Electronics*, Vol. 50, No.4, pp.667-675 (2003)

[5] W. Takano, Y. Nakamura, ”Segmentation of human behavior patterns based on the probabilistic correlation” *The 19th Annual Conference of the Japanese Society for Artificial Intelligence* (2005)

[6] H.Sakoe, S.Chiba, ”Dynamic Programing Algorithm Optimization for Spoken Word Recognition”, *IEEE Transaction on Acoustics, Speech, and Signal Procdssing*, Vol.ASSP-26,No.1,pp.43-49 (1978)

[7] Y. Yamada, E. Suzuki, H. Yokoi, K. Takabayashi ”Decision-tree Induction from Time-series Data Based on a Standard-example Split Test”, *Proc. Twentieth International Conference on Machine Learning (ICML)*, pp.840-847 (2003)

[8] K. Nakamoto, E. Suzuki ”Fast Clustering for Time-series Data Based on a TWS Tree”, *Proc. 48th SIG-FAI, Japanese Society for Artificial Intelligence*, pp.9-14 (2002)

[9] http://www.jp.aibo.com/index.html

[10] http://openr.aibo.com/