

# ロボットとのインタラクションにおける ユーザの嗜好の類似性を用いた行動学習

Learning method based on the similarities of interaction between a robot and humans

大村 英史\*<sup>1</sup>      片上 大輔\*<sup>1</sup>      安村 禎明\*<sup>1</sup>      新田 克己\*<sup>1</sup>  
Ohmura Hidefumi      Katagami Daisuke      Yasumura Yosiaki      Nitta Katsumi

\*<sup>1</sup>東京工業大学大学院総合理工学研究科  
Tokyo Institute of Technology

In this report, we propose a learning method for a robot to interact with various types of people effectively. This method adopts an extended classifier system which uses variable set of classifier and a user profile for each people. Each person's classifier strength is enhanced by the interaction between the robot and similar type of people, when the classifier fires. We have experimented the effects of the learning method using AIBO ERS-220A.

## 1. はじめに

近年、多数のロボットが人間の生活空間で活躍してきており、画像認識や音声認識の技術を駆使してオーナーを特定しそれに応じたインタラクションが行われている。例えば、PaPeRo[PaPeRo 99] は人間を識別し状況にあわせてコミュニケーションする機能を持っている。また AIBO[AIBO 99] は画像と音声によりオーナーを識別することができる。しかし、これらのロボットの多くは、個別の対応を学習し、特定の個人に対しての動作を行うことが多い。従って、初めての環境やインタラクションに対しては一から学習しなくてはならないため学習に長い時間を必要とし効率的ではない。

一方、人間はこのような場合でも、適した対応をとることができる。例えば、初対面の人とコミュニケーションをとるといった場合においても、過去の経験から現在の状況と似た人を連想し、そこで学習した情報を用いることにより、うまく対応することができる。

本研究では、複数のユーザとのインタラクションにおいて、各ユーザの嗜好の類似性に基づいて仮想的に他のユーザとのインタラクションに対する学習を行い、複数のユーザに対して対応するシステムの構築を目的とする。

## 2. ユーザの嗜好の類似性を用いた学習

人間は他の人間とコミュニケーションを行う場合、過去に積み重ねた経験を学習しそれをもとに現在の行動を決定していると考えられる。たとえ、それが初めてコミュニケーションをとる人だとしても、数回インタラクションを行うことで、過去コミュニケーションをとった人と似たタイプだと判断できれば、その人との経験に基づいて、初めての人に対してもある程度の対応が可能である。

たとえば、図 1 のような状況を考える。ここで、矢印の向きはインタラクションの方向、太さは経験の量、線の種類はインタラクションの種類を表すものとする。

(a) インタラクション: ユーザ H は日常様々な人達と多くのインタラクションを行っている。インタラクションの内容は、その人に対して異なり、類似のものであったり、まったく異なるものであったりする。

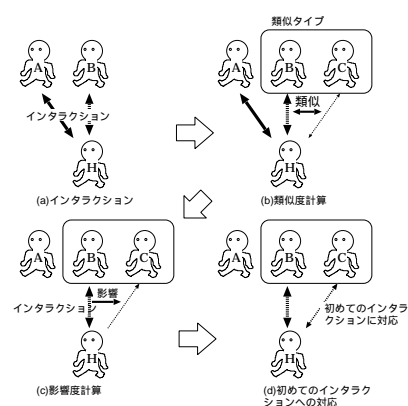


図 1: 人間の類似性に基づいた行動学習

(b) 類似度計算: ここで、初めて C に会った場合、C がどのような人物かわからないので、当たり障りのない常識的なインタラクションから始める。数回のインタラクションを行いながら、C との経験を過去に経験してきた様々な人とのインタラクションと比較し、過去のどの人物とタイプが似ているのかを判定する。

(c) 影響度計算: 人 H に対する人 B と人 C のタイプが似ている場合、B もしくは C とのインタラクションの経験はもう一方とのインタラクションに利用することができる。

(d) 初めてのインタラクションへの対応: C から初めてのインタラクションがあった場合、通常対応することができないが、類似タイプである B とのインタラクションの経験から、そのインタラクションに対して B のようなタイプの人とはどう対応すればいいかはすでにわかっているので、円滑な対応を行うことができる。

本研究では、以上のような仮説に基づき、個々のインタラクションをユーザの嗜好としてユーザの嗜好の類似性を測ることとした。我々はこのような学習を嗜好の類似性を用いた学習と呼び、これをロボットに搭載することで、複数のユーザに学習初期段階から対応する学習システムを構築する。

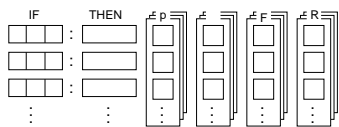


図 2: 個別変数

### 3. 本研究の手法

#### 3.1 クラシファイアシステム

本研究では、機械学習システムの中からクラシファイアシステム（以下 CS）の一つである XCS[Wilson 95] を実現の手法として用いる。CS を用いた主な理由は (1) ロボットの学習結果がルールとして獲得されるため獲得された行動の可読性が高い (2) 環境の動的な変化に対しても、獲得された知識に遺伝的操作を加えることで比較的容易な適応が可能である等があげられる。

XCS ではクラシファイアの強度パラメータの代わりに各クラシファイアに以下の 3 つの新しいパラメータが用いられている。予測値  $p$ : 受け取った内部または外部の報酬の数回分の平均。予測誤差  $\epsilon$ :  $p$  パラメータの誤差の計測の平均。適応度  $F$ :  $\epsilon$  の逆関数。予測値  $p$  の更新は報酬  $B$  に基づく次式 ( $p_n = p_{n-1} + \beta(B - p_{n-1})$ ) によって行われる。ここで、 $\beta$  は学習率を定める定数である。

#### 3.2 個別変数

本研究では複数の人とインタラクションを行うため、クラシファイアに関する変数  $p, \epsilon, F$  をそれぞれインタラクションを行う各ユーザごとに持つ。それぞれ個別予測値、個別予測誤差、個別適応度と呼ぶ。また、過去のインタラクションの経験を記録する個別評価値  $R$  を設定する。この値は各クラシファイアにそれぞれ設定されており、正の評価を受けたとき、1 ずつ増加し、負の評価を受けたとき、1 ずつ減少する。これらの値を個別変数（図 2）と呼ぶ。システムは、インタラクションを行うユーザ毎に個別変数を持つ。

#### 3.3 プロファイル

本稿では、前節で述べた個別変数である個別評価値  $R$  を成分として持ったベクトルをプロファイルと設定する。そのため人  $A$  に対するプロファイルにおける個別評価値  $R$  は存在するクラシファイア（ルール）の数だけ存在する。プロファイルは次節の類似度の計算で用いるものである。クラシファイアが多すぎると適切な類似度が求められないことがあるため、適切なフィルタリング方法によりプロファイルを作成する必要がある。これについては、4.2, 4.3 章で詳しく説明する。

#### 3.4 類似度の計算

本稿では、類似度計算の最も単純な方法のひとつとしてベクトル空間法を用いる。システムは、行為に対してユーザの好きか嫌いかのインタラクション情報を個別評価値としてクラシファイア毎に獲得する。ユーザ  $A$  とユーザ  $B$  の類似度  $S$  は、個別評価値をまとめた各ユーザのプロファイルから以下の計算式でもとめる。 $f, g$  は比較する 2 つのベクトル（プロファイル）である。

$$S(f, g) = \frac{(f, g)}{\|f\| \|g\|} = \frac{\sum_{m=1}^K f_m g_m}{\sqrt{\sum_{m=1}^K f_m^2} \sqrt{\sum_{m=1}^K g_m^2}}$$

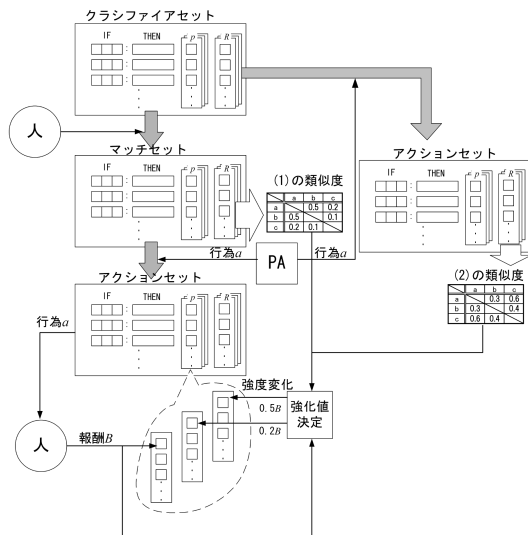


図 3: ユーザの嗜好の類似性を用いた行動学習システム

### 4. ユーザの嗜好の類似性を用いた行動学習

#### 4.1 システムの概要

本研究で提案するシステムの概要図を図 3 に示す。基本的な学習機構は 3.1 節の XCS を用いている。システムは 3.2 章で説明した個別変数を各ユーザに対して用意し、個別変数は通常の XCS と同様に更新をする。ただし、個別予測値  $p$  においては、他のユーザがインタラクションを行った際にも、そのユーザと他のユーザとの類似性に従い更新される。また、個別評価値は、ユーザとのインタラクション（評価）に基づき更新され、その都度類似度が計算される。

ここで、プロファイルの作成に全てのクラシファイアの個別予測値を用いると、冗長性が高く正確な類似度を計算できないといった問題が起こることが予想される。本稿では、適切な類似度を作成するために、システムの入力である環境の状態とシステムが出力する行為に注目した 2 通りのフィルタリング方法を提案する。

#### 4.2 新しい行為の評価を予測するためのフィルタリング

ユーザに対してまだ行ったことのない行為の評価を予測するために行うフィルタリングである。ここで使用するプロファイルは、その時の「状態」で選び出したマッチセットの評価値をプロファイルとする。これによりその状態に関係のないクラシファイアを類似度計算から省くことになり、より正確な類似度を生成することが可能である。図 3 の (1) の類似度に対応している。

具体的な例を図 4 に示す。A は行為「踊る」と「歌う」を好んでおり、「歩く」を嫌っている。また、B は「踊る」を好み、「歩く」を嫌い、「歌う」に関しては未経験である。状態「たたく」で固定したセット（マッチセット）で類似度を計算することにより、B は A と類似度が高くなり、A と似たタイプだと判断される。この結果 B が「歌う」を好むことを予想することができる。

#### 4.3 新しい状態に対して適切な行為を予測するためのフィルタリング

ユーザからのまだ経験したことのない状態に対して、適切な行為を予測するために行うフィルタリングである。ここで使

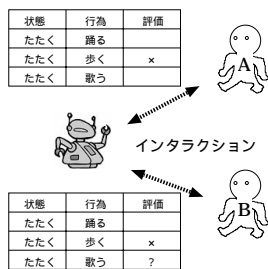


図 4: 新しい行為の評価の予測

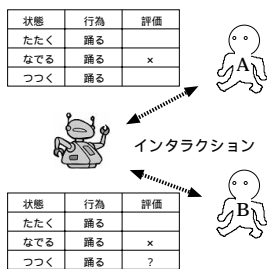


図 5: 新しい状態の行為の評価の予測



図 6: AIBO「EPS-220A」

用するプロフィールは、その時に選択された「行為」で選出されたクラシファイアセットの評価値をプロフィールとする。図 3 の (2) の類似度に対応している。

具体的な例を図 5 に示す。この場合も、行為を固定したアクションセットで類似度を計算することにより、4.2 節と同様、B は A と類似度が高くなり、B と似たタイプだと判断される。この結果 B は A と同様、「つつく」のとき「踊る」を好むことを予想することができる。

#### 4.4 個別予測値の更新

クラシファイアの個別予測値  $p$  が、人間とロボットのインタラクションによって更新する方法を図 3 において説明する。(1) システムは人 A とインタラクションを行うことによりセンサ情報を獲得する。(2) センサ情報よりマッチセット  $[M]$  が生成され、行為選択法より選ばれた行為  $a$  を出力する。(3) 人 A から評価をもらい予測値  $p$ 、個別変数を更新する。(4) 人 A の更新されたプロフィールとその他の人のプロフィールを利用して類似度  $S$  を計算する。(5) 他の方の個別予測値  $p$  を類似度  $S$  に従って以下のように更新する。

予測値  $p$  の更新は類似度が正のときと負のときで異なり、以下の式で行う。ここで、 $p = n$  は人 A に対する step  $n$  における予測値、 $\beta$  は係数、 $B$  は人 A からの報酬とする。

$$p_n = \begin{cases} p_{n-1} + \beta \cdot S \cdot (B - p_{n-1}) & (S > 0) \\ p_{n-1} + \beta \cdot |S| \cdot (|B - 100| - p_{n-1}) & (S \leq 0) \end{cases}$$

## 5. 実験

### 5.1 実験設定

本研究は人間の活動空間で共に活動するロボットを対象としている。そこで実験には人間の活動空間に浸透しつつあるペットロボットを用いる。使用するロボットは SONY のエンタテインメントロボット AIBO「EPS-220A」(図 6) をである。AIBO は無線 LAN を通じて PC と通信を行うことができる。全ての実験は、以下の被験者と AIBO のインタラクションを通して行う。(1) 被験者が AIBO のボタンを押し、AIBO が行為を出力する。(2) AIBO の行為に対して被験者の好みで「ほめる」「しかる」の 2 種類のうちどちらかの評価を与える。(3) AIBO が評価に対してリアクションをとる。(4) (1) にもどる。上記の手続きを一回のインタラクションの手続きとする。

### 5.2 実験 1, 2: 新しい行為の評価を予測

被験者は以下の A~D の 4 タイプに分ける。A: 新しい行為を先に学習し、他の人に影響を与える人。B: A と類似度  $S$  が高く、A の大きな影響を受ける人。C: A と類似度  $S$  の低く、A の小さな影響を受ける人。D: A の新しい行為の学習期間に A の影響を受けない人。これにより、B と D を比較する

ことで、新しい行為の評価を予測するフィルタリングの効果を見ることが可能である。また、B と C の学習後の個別予測値を比較することで、A との類似度に応じた影響によって正しく学習できたかどうかを解析することが可能である。実験は以下の 3 段階に分かれている。

1. プロファイル作成期間  
被験者: 全員  
状態(固定): 1 種類(背中のボタン)  
行為: No.1~No.10 の 10 種類(ランダムで出力)  
学習回数: 30 回
2. 新しい行為の学習期間  
被験者: A  
状態(固定): 1 種類(背中のボタン)  
行為: No.11~No.15 の 5 種類(ランダムで出力)  
学習回数: 30 回
3. テスト期間  
被験者: B, C, D  
状態(固定): 1 種類(背中のボタン)  
行為: No.11~No.15 の 5 種類(ルーレット選択で出力)

プロフィール作成期間の終了後、行為 No.11~No.15 の評価値をプロフィールとして、被験者の全組み合わせの類似度を算出する。その値よりタイプ A~D のどこに分類するかを決定する。新しい行為の学習期間では、まず A となった人が新しい行為を学習する。その期間中に、類似度に応じて B, C の個別予測値が更新される。最後にテスト期間において、B~D のインタラクションの手続きを行い評価とする。

ここで、被験者 A からの影響を調べるためにプロフィールから被験者 A, B, C, D のグループを 2 組、複数人の被験者 A1, A2 からの影響を調べるために A, B, C 各二人ずつ D を一人を 1 組とし、計 3 回の実験を行った。それぞれ実験 1, 2 とする。

### 5.3 実験 1 結果: 新しい行為の評価予測の効果

行為 13 における個別予測値の変化を図 7 に示す。A と類似度の高い B は初期値が 70 あたりから始まっている。それに比べ、C, D の予測値の初期値はほぼ 50 である。B と D を比較すると収束速度が明らかに B のほうが速いことがわかり、類似度によりうまく仮想学習が働いているといえる。

### 5.4 実験 2 結果: 複数人の影響下における効果

複数人数から影響を受ける環境下での個別予測値の変化を図 8 に示す。C, D に比べ B が圧倒的に早く収束している。この実験でも B において学習が促進されているといえる。

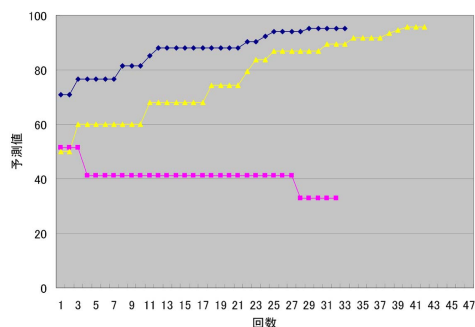


図 7: 実験 1: 個別予測値の変化 (行為 13)

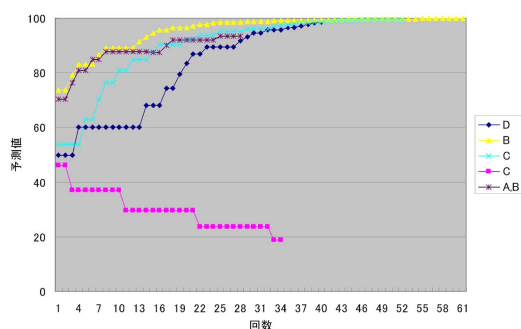


図 8: 実験 2: 個別予測値の変化 (行為 13)

すべての B においての行為と予測値のグラフを図 9 に示す。それぞれの予測値が正に近づくもの「行為を好む」と、負に近づくもの「行為を嫌う」に分けた。ここで、閾値を+10, -10 とすると、予測が成功し学習を促進させたものが 40%、予測がうまくいかなかったものは 10% であった。ユーザの嗜好の類似度により新しい行為を予測できたと言える。

### 5.5 実験 3: 新しい状態の適切な行為を予測する実験

被験者は a~f の 6 人で行う。はじめにプロフィールを作成期間を全員行い、順番にテスト期間に入る。テスト期間では、その人より前にこの期間を行った人の影響を受けることになり、6 番目の人は 5 人の影響を受けていることになる。

#### 1. プロファイル作成期間

被験者: 全員  
状態: 2 種類 (背中, 頭のボタン)  
行為 (固定): 16~20 の 5 種類 (ランダムで出力)  
学習回数: 命令ひとつにつき 15 回, 全部で 30 回

#### 2. テスト期間

被験者: 全員  
状態: 1 種類 (あごのボタン)  
行為 (固定): 16~20 の 5 種類 (ルーレット選択で出力)

### 5.6 実験 3 結果 ~ 新しい状態の適切な行為を予測 ~

実験 1,2 と同様、テスト期間中における個別予測値の変化を見てみると、b, c に比べ e, f のほうが的確な学習の促進を行っていた。これは、より多くの人からの影響を受けたほうが学習の促進につながり正しい行為を強化することができることを示している。

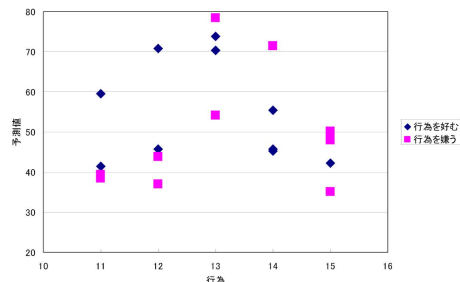


図 9: 予測値の初期値

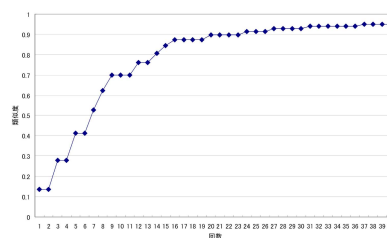


図 10: 実験 3 の a, b の類似度の変化

学習の効率を確認するために 5.4 節と同様、b~f のそれぞれの行為の予測値を調べた。それぞれ「行為を好む」と「行為を嫌う」と、「どちらでもない」に分けた。学習を促進させたものが 64% であり、うまくいかなかったものは 4% であった。

### 5.7 類似度の変化に対する追従性

本実験における類似度はロボットとユーザのインタラクションによって常に動的に変化している。実験 3 の行為 19 における a, b の予測値の変化の様子を図 10 に示す。この図は、初めは類似度が低かったものがインタラクションを重ねるたびに高くなっていくことを示しており、類似度が初めと終わりで大幅に変化していることがわかる。このように、本システムでは類似度の変化に追従するため、最終的に実験結果のような正確な予想ができたと考えられる。

## 6. まとめ

本論文では、ロボットがユーザの嗜好の類似性を見つけることで、複数のユーザとのインタラクションにおいて効率の良い学習ができる手法を提案した。

このシステムを用いた新しい行為の評価を予測する実験と新しい状態に対して適切な行為を予測する実験において学習の促進をすることを確かめた。また、複数のユーザの影響を受ける場合における効果を検証し、複数の影響を受けることで、より正確な行為を推薦できるということがわかった。

## 参考文献

[PaPeRo 99] <http://www.incx.nec.co.jp/robot/>  
[AIBO 99] <http://www.jp.aibo.com/>  
[Wilson 95] S. W. Wilson : Classifier Fitness Based on Accuracy, Evolutionary Computation, 3(2), pp.149-175, 1995.