日本知能情報ファジィ学会誌「IDEA」特集解説

IDEA: 適応のためのインタラクション設計
IDEA: Interaction DEsign for Adaptation

山田誠二
国立情報学研究所
〒 101-8430 東京都千代田区一ツ橋 2-1-2
Tel&Fax: 03(4212)2562
E-mail: seiji@nii.ac.jp

Seiji Yamada
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda, Tokyo 101-8430, JAPAN
Tel&Fax: +81-3(4212)2562
E-mail: seiji@nii.ac.jp

角所 考
京都大学 学術情報メディアセンター
〒 606-8501 京都市左京区吉田本町
Tel: 075(753)9062
Fax: 075(753)9056
E-mail: kakusho@media.kyoto-u.ac.jp

Koh Kakusho
Academic Center for Computing and Media Studies, Kyoto University
Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, JAPAN
Tel: +81-75(753)9062
Fax: +81-75(753)9056
E-mail: kakusho@media.kyoto-u.ac.jp

1 はじめに

ここ 10 年ほどで，マイクロソフトエージェント [9] に代表される擬人化エージェント（life-like agent）[7]，ECA (Embodied Conversational Agent)[3] と呼ばれるソフトウエアエージェントから，AIBO[1] や Roomba[12] に代表される身体をもったペットロボットや掃除ロボットまで様々なエージェントが，オフィスや一般家庭に普及してきた．このようなエージェントは，身体をもつか否かに関わらず，人間のユーザとのインタラクションを持ちながらも，自分自身で状況判断と行動決定を行い，実行していく能力をもっている．そのような人工物と一般のユーザが日常的にインタラクションを持ち続ける状況は，これまでの科学技術の歴史からみても，前例がない．そして，このような状況では，人間のユーザとエージェントがどのように付き合っていけばいいのか，さらにうまく付き合っていくにはどのようなインタラクションを設計すればいいのかが，重要な工学的課題となってくる．人間とエージェントが自然に無理なく付き合っていけることが，ユーザにとっては負担が軽減する意味でメリットがあり，エージェントのさらなる普及のためにも必須である．

しかし，これまで，エージェントとユーザ間のインタラクションをどのように設計すれば，人間とエージェントがよりよい付き合いを実現できるかについて，深く議論する研究分野はなかったと言ってよい．このような人間のユーザとエージェント間のインタラクション設計問題は，HAI(Human-Agent Interaction)[21] と呼ばれ，近年特に日本において活発に研究されている．本特集では，そのような研究の論文を集めて掲載しており，先端的な HAI 研究の一端を知ることができるだろう．

これまでの HCI（Human-Computer Interaction），ユーザインタフェースの研究では，主にユーザにとっての使いやすさ（usability）の向上をめざし，様々なデバイス，メタファなどが研究，開発されてきた [13]．一方，HAI を考えた場合の重要な特徴として，人間と接する対象がエージェントと呼ばれる自律性をもつ，行動主体であることが挙げられる．人間は不可避的に対象（エージェント）に適応するし，エージェントは個人適応（personalization）するために，人間に適応する必要がある．つまり，人間はエージェントに，エージェントは人間にお互いが適応しあうという相互適応が生じる．しかし，そこでは，エージェントが適切なタイミングで適切な情報を提示してくれない，エージェントがユーザ適応しているのかどうかわからずに戸惑ってしまう，また，エージェントが余計なユーザ適応をするので困る等のお互いの不適切な適応に起因する問題が多く生じる，また，“適応” は，人工知能，ロボティクス，認知科学における機械学習，ユーザモデル，認知モデルの研究において，今日においてもさらに重要な研究課題である．よって，人間とエージェントの上手な付き合いを実現するには，人間とエージェント間のインタラクションを,「人間とエージェント間の適切な適応」の実現を目的として，設計する必要があると考えられる [20].

このような背景から，我々は人間とエージェント間の “適切な適応” を実現するために，その “適応を促進する HAI の設計” を共通の目的とした研究分野を「適応のためのインタラクション設計」IDEA(Interaction DEsign for Adaptation) と呼び，同名の研究グループ IΔEA を中心に，これまで議論を重ねてきた．本稿では，その議論を基に関連する研究例を紹介しながら，IDEA の目指すもの，それを実現するための方法論について議論する．以下では，まず HCI や HAI に対する IDEA の位置付けについて，2 章で議論し，それを踏まえて，3 章で，人間とエージェントの適応の仕方に基づいて HAI を分類・モデル化すると共に，その設計に関する研究例について紹介する．

2 HAI 設計

2.1 インタラクション設計

HCI のような計算機とのインタラクションにおいても，HAI のようなエージェントとのインタラクションにおいても，人間の側には通常，それらとのインタラクションによって実現しようとする “タスク” が存在する．例えば，ワープロとインタラクションする場合のタスクは文書を作成することであり，案内ロボットとインタラクションする場合のタスクは目標場所に到達することである．このようなタスクを，人間が独力ではなく，計算機やエージェントといったシステムを利用して実現するわけであるから，人間の側からシステムへの何らかの働きかけによって，作成したい文書や目標場所等の目標タスクに関する情報を伝え，その結果として意図通りの文書ができあがったり，目標場所への経路が提示される，といったシステムの働きが得られることになる．

本稿では，後の議論において，インタラクションの対象となるシステムとして，特にロボットや CG キャラクタ等の擬人化エージェントを考えるため，上で述べたような人間からシステムへの働きかけや，そのフィードバックとしてのシステムの働きを，広く人間やシステ

ムの**行為**と呼ぶことにする．したがって，この行為は，ロボットのモータコマンドのような物理的行為から，擬人化エージェントの表情の生成のように何らかのモダリティや表現による外界への情報の表出までを含んだものである．また，このような行為を決定する規則に相当するものを**行動ルール**と呼ぶことにする．より具体的に言えば，ここでいう行動ルールとは，外界からの情報をセンシングして判定される条件の AND を条件部として持ち，その条件部が成り立てば後件部の行為が実行されるルールベースに相当する働きをもったものであり，その条件部は，それが成り立つか否かを判定する手続きも含んでいるものとする．ここで，人間やシステムの行為は，外界からの入力のみを用いて即応的に決まるとは限らず，自身の内部状態にも依存し得ることから，そのような内部状態と行動ルールの集合を合わせたものをさらに**行動**（behavior）と名付ける．このとき，人間とエージェント間のインタラクションは，お互いが自身の行動を実行することで生じることになる．したがって，システムの行動を設計することは，単に内部状態や外界からの情報とそれに対して実行すべき行為の対応関係を設計することにとどまらず，システムによる外界のセンシングの手続きや表情などの表現，つまり従来のユーザインタフェースの当たる部分も含むことになる．

人間がシステムとインタラクションする場合，人間は，システムの行動に関するモデル（以後，**行動モデル**と呼ぶ）を構築し，これに基づいて自分の行動を決めるから，人間とシステムとのインタラクションによって目標とするタスクが実現されるには，人間が持つシステムの行動モデルとシステムの実際の行動が一致している必要がある．これが一致していない場合，人間があるタスクの実現を目標として行動しても，それに対するシステムの行動は予想とは異なるものとなり，タスクの実現に失敗することになる．したがって，インタラクション設計では，この両者をどのようにして一致させるかが議論の中心となる．

2.2 行動の親和性

従来の HCI におけるインタラクション設計では，上のようにシステムの行動モデルと実際の行動を一致させるための方策として，人間が最初から持っている初期的なシステムの行動モデルとできるだけ一致するように，システムの行動を設計するというアプローチが採られてきた．このように，人間とシステムとのインタラクションにおいて，システムの行動が，人間の持つ初期的なシステムの行動モデルに近いことを，**行動の親和性**と呼ぶことにする．

HCI における行動の親和性を実現するための代表的なインタラクション設計の手法が，GUI(Graphical User Interface) で用いられるモデル世界メタファや直接操作 (Direct Manipulation) 等であり，ポインタやアイコンによって，ファイルの操作を現実の机上での書類操作と類似した形で実現することにより，システムによるファイル操作の実現に，現実の机上での書類操作のためのメンタルモデルがそのまま活用できるように工夫されている．

2.3 適応過程の活用

人間の持つシステムの行動モデルと実際の行動を一致させるための方策には，2.2 で述べたような行動の親和性を利用することに加えて，もう一つ，人間およびシステムによる適応を利用するという方法が考えられる．

HCI や HAI において，システムの行動モデルと実際の行動が異なるために目標とするタスクの実現に失敗するような状況が生じた場合，人間は通常，システムへの適応を試み，システムとのインタラクションを通じて，システムの正しい行動モデルを学習しようとする．この行動モデルの学習が難しい場合，人間からシステムへの適応は困難なものとなるが，学習が容易である場合には，人間はほとんど意識することなく適応を完了する．したがって，もし人間が元々持っている適応能力をうまく活用し，人間が正しい行動モデルを速やかに学習できるように，適応を促進すれば，人間に負担を感じさせることなく，人間の持つシステムの行動モデルと実際の行動を一致させることができる．このような過程を実現することが，適応のためのインタラクション設計の問題である．

上のような適応過程の利用の仕方としては，以下の２段階が考えられる．

(1) 人間による適応のみを利用する．

(2) (1) に加えてシステムからも人間に適応する．

ここで，システムから人間への適応のみを考えないのは，人間によるシステムへの適応が不可避であり，人間による適応の生じないシステムの適応過程というのは，実際には起こり得ないためである．

2.4 適応のためのインタラクション

上で述べた２段階における人間とエージェントによる**適応**とは，相手の行動モデルを学習し，そのモデルに基づいて自身の行動を変化させることで実現される．

ここで，2段階のうちの (1) を促進するためには，システムから人間に対して，人間がシステムの正しい行動モデルを効率的に学習するための情報を伝達する必要がある．また (2) では，これに加えて，システムが人間の正しい行動モデルを効率的に学習するための情報を，人間からシステムに対して伝達する必要がある．このような情報は，いずれもタスクの実現のために直接利用される情報ではなく，適応を促進するために必要となる情報であり，そのような情報伝達が，元々タスクの実現に直接利用される情報を伝達するために行われていたインタラクションの過程で，相手への適応のために二次的に派生することになる．そこで本稿では以後，これら二種類の情報伝達過程を区別し，タスクの実現に直接利用される情報を伝達するための元々のインタラクションを，**タスク実現のためのインタラクション**，相手による適応を促進するための情報を伝達するために二次的に派生するインタラクションを，**適応のためのインタラクション**と呼ぶことにする．ただし，この分類は，二種類のインタラクションで利用される情報が必ず異なることを含意するものではなく，同一の情報が両方のインタラクションで利用される状況ももちろん存在する．

2.5 学習の親和性

適応のためのインタラクションにおいて，相手による行動モデルの学習を促進するための情報を提供するには，“相手は自分の行動モデルをどのように学習するのか” を知っている必要がある．また，人間やシステムの行動は，相手の行動モデルに依存するから，(2) のように人間とシステムの両者がお互いの行動モデルを学習しつつある状況では，それぞれの行動が，相手側の行動モデルの学習によって変化することになる．したがって，このような行動の変化が行動モデルの学習に悪影響を与えないようにするためには，やはり相手の学習について知っていることが重要となる．すなわち，適応のためのインタラクションでは，相手の適応を促進するために，相手が持っている自分の学習のモデル（以後，**学習モデル**と呼ぶ）と実際の自分の学習が一致している必要がある．

この要求は，2.1 で述べたタスク実現のためのインタラクションにおける行動モデルへの要求と同一である．このための方策として，2.2 で述べた行動モデルに対する親和性と同様，学習モデルに対する親和性が有効となる．すなわち，人間が最初から持っている初期的なシステムの学習モデルとできるだけ一致するように，システムの学習を設計する一方，システムが適応の際に利用する人間の学習モデルとして，人間の学習に関する知識を導入することが考えられる．

2.6 学習の汎用性

一般に，どのような条件のときにどのような行為をとるかという行動は，個人やタスクによって様々に異なることから，このような行動に関して，人間の持つシステムの行動モデルに一致するようにシステムの行動を設計することは必ずしも容易ではない．これに対して，“成功体験の繰り返しはその行動を活性化し，失敗体験の繰り返しはその行動を抑制する” といったことは，個人やタスクに関わらず，人間やその他の生き物の学習に広く共通して見られる性質である．このことを本稿では，**学習の汎用性**と呼ぶ．このような学習の汎用性を利用すれば，上のような性質を持った学習モデルを，相手の学習モデルとして人間やシステムに共有させることにより，個人やタスクによらず，学習モデルの親和性を実現することが可能になる．

2.7 擬人化

HCI に対して HAI では，人間がインタラクションするシステムが，ロボットや CG キャラクタ等の擬人化エージェントとなる．ただしここでいう擬人化とは，人間だけではなく，動物などの人間以外の生き物に似せる場合も含める．我々の日常生活において，クッションに話しかける人は少ないが，ぬいぐるみに話しかける人は珍しくないことからもわかるように，人間は，物質としては同じであっても，擬人化された対象とはインタラクションしやすい．本稿では，擬人化の持つこのような特長を，**インタラクションの誘発性**と呼ぶ．従来の HAI では主に，HCI にこのインタラクションの誘発性を導入し，ロボットを擬人化することによって，人間からのインタラクションを容易なものにしたり [10]，擬人化エージェントのふるまいをより人間らしくすることによって，人間から必要な情報を得やすくする [2] といった試みがなされてきた．

上のようなタスク実現のためのインタラクションに加えて，適応のためのインタラクションを考える場合においても，このようなインタラクションの誘発性は，適応のための情報伝達を活性化できるために活用できる．また，擬人化は一種のメタファであり，人間は，擬人化されたシステムに対しては，擬人化されている人間等とのインタラクションで用いているものと同様のモデルを利用してインタラクションができることを期待するから，2.2 や 2.6 で述べた行動の親和性と学習

の汎用性の両者を同時に実現するための方策としても，擬人化は非常に有効な手段となる．

2.8 IDEA

以上の議論に基づき，IDEA では，人間とシステムとのインタラクションにおいて，人間によるシステムへの適応が不可避的に生じることを考慮し，人間－システム間で互いの行動モデルと実際の行動とを一致させるために，行動の親和性だけではなく，人間の適応能力も積極的に活用しようとの立場から，擬人化の利点であるインタラクションの誘発性，行動の親和性，学習の汎用性をうまく利用することよって，人間やシステムによる適応を促進するような適応のためのインタラクションを設計することを目指す．

すなわち，従来の HCI が，2.2 で述べたような行動の親和性に基づくタスク実現のためのインタラクション設計を，また HAI が，それに加えて 2.7 で述べたようなインタラクションの誘発性も導入したタスク実現のためのインタラクション設計を議論していたのに対し，IDEA では，人間とエージェントによる適応に焦点を当て，両者による適応のためのインタラクション設計を考えることになる．このときの課題は，以下の２段階となる．

1. 人間の適応を促進するインタラクション設計
2. 相互適応を促進するインタラクション設計

なお，本稿では，“適応” は，対象とのインタラクションをある評価レベルに維持することであり，“学習” とは，エージェント（あるいは，人間）が，自身の行動を対象に合わせて変化させることにより，適応を実現すること，つまり適応の一実現方法であると解釈する．

上のように設計の焦点がタスク実現のためのインタラクションから適応のためのインタラクションへと推移することに伴い，HCI や HAI では行動モデルのみが設計対象であったのに対して，IDEA では，行動モデルの学習を目指して学習モデルを設計することになり，HCI や HAI に比べて設計対象が拡張される．

ここで注意してほしいのは，IDEA では，相手の行動モデルが学習対象であり，学習モデルについては，設計対象とはなるものの，学習対象にはしない点である．また，学習モデルの設計においては，人間は元々，エージェントの学習モデルを持っており，それを変化させること，つまりエージェントの学習モデルを学習することはしないと仮定する一方，エージェントには人間の学習モデルを設計して組み込んでおき，以後変化させないものとする．これは，“学習モデルを学習するためのメタ学習モデルを考え，さらにそのメタ学習モデルを学習するためのメタメタ学習モデルを考える...” というような無限後退を，現実的に妥当なレベルで回避する意味がある．

さらに，人間側の行動や学習を設計することは現実的でなく，またユーザの自由度を妨げるため，これから述べるすべてのモデルにおけるインタラクション設計の設計対象は，人間を含まない．つまり，IDEA における適応の促進は，エージェントの設計で実現される．

次章では，このような IDEA の概念をより明確，具体的にするために，IDEA の目指す人間とエージェントの適応について，適応の観点からモデル化を行い，従来研究との比較や IDEA の扱う問題の分類，そして，IDEA の設計対象と考えられる方法論について議論する．

3 適応のための HAI モデル

3.1 従来の HCI

人間とエージェント（あるいは，システム）とのインタラクションを，適応の観点からとらえた場合，これまでの HCI は，図 1 のように表される．モデルの構成要素を以下に説明する．

- **人間の行動** B_i^H：i ターンのインタラクションにおいて，適用，実行されるエージェントのもつ行動．B_0^H は初期行動であり，エージェントとのインタラクションが始まる前から持っている行動である．
- **エージェントの行動** B_i^A：人間の行動と同様のエージェントの行動．B_0^A は，設計者により設計された初期行動を意味する．
- **行動モデル** $M(B)$：相手の行動 B のモデル．
- **インタラクション** I_i：人間，あるいはエージェントの学習– 実行の周期により決まるターンの i ターン目における，人間とエージェント間の情報（物理的な行為を含む）のやり取り．人間の行動 B^H とエージェントの行動 B^A により決定される．

また，行動モデル $M(B)$ と行動 B の半円が接触している状態は，行動 B が $M(B)$ の影響を受けることを意味する．

図からわかるように，従来の HCI では，人間の行動 B_0^H は元々人間がもっている行動であり，エージェントに適応して変化することはなく，固定されていると仮定している場合がほとんどである．また，ごく一部の適応インタフェースを除く，多くの HCI の研究では，

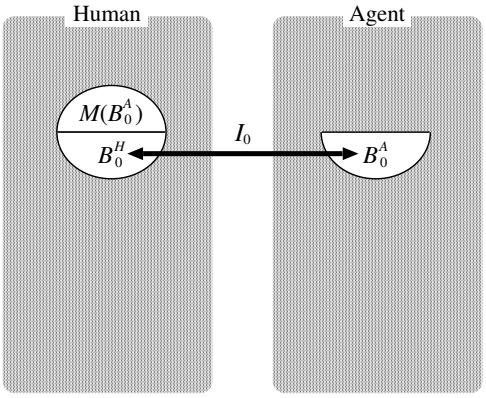

図 1　従来の HCI

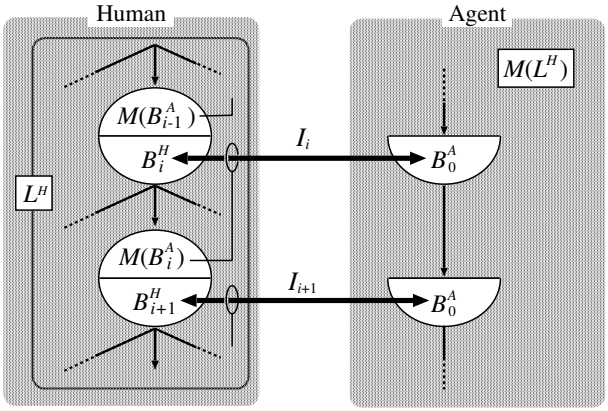

図 2　人間からの適応

エージェント側もその行動 B_0^H を変化させず，適応を行わない．

よって，人間の要素は設計対象にならないという指針，およびインタラクション I は，B_0^H と B_0^A により決定されることから，このモデルにおけるインタラクション設計の設計対象は，エージェントの行動 B_0^A となる．なお以降では，B_0^A，$M(B)$，そして後述する学習 L の設計を **HAI 設計**，あるいは単に**インタラクション設計**と呼ぶ．

従来の HCI では，この B_0^A を，できるかぎり $M(B_0^A) = B_0^A$ となるように設計してきた．つまり，人間がもともと持っている固定的なエージェントの行動モデルとエージェントの行動ができるだけ一致するような方針で，エージェントの行動を設計してきた．これが，2.2 において議論した行動の親和性である．

機能が理解しやすいアイコンの設計や直接操作（direct manipulation）などの理解しやすい行動の設計など，多くの HCI の研究は，適応の観点から図 1 のモデルで説明できる．次節以降では，IDEA の目標とする適応のためのインタラクション設計のモデルについて述べる．

3.2　人間の適応を促進するインタラクション設計

2.8 でまとめたように，HCI とは大きく異なり，IDEA では，人間からエージェントへの適応と，エージェントから人間への適応を最重要に考える．このうち，人間からエージェントへの適応は不可避的におこるものなので，IDEA が対象とするのは，人間からエージェントへの適応のみの場合と，両者が互いに相手に適応する相互適応の場合の 2 つである．

本節では，まずエージェントが適応をしない場合，つまり人間からエージェントへの適応のみの場合を扱う．その HAI モデルは，先の図 1 に，人間の適応を追加した図 2 のようになる．$M(B_i^A)$ は，人間が学習するエージェントの行動モデルであり，任意の過去の人間とエージェント間のインタラクション I_{i-1} を基に学習され，その $M(B_i^A)$ が現在の B_i^H に影響を与える．そして，人間の行動 B_i^H とエージェントの行動 B_i^A とが実行されることにより，インタラクション I_i が生じる．

一般には，行動 B_i は相手のモデル $M(B_{i-1})$ に影響を受けること，さらに，そのモデル $M(B_{i-1})$ は，過去のインタラクションから学習されたモデルであることに注意して欲しい．このモデルで，“適応” とは，相手の行動モデルと自身の行動のペアを探索しながら学習していくことであり，そのような学習，あるいはその探索戦略（学習戦略）を L^H（人間の学習），L^A（エージェントの学習）と呼ぶ．ここでは，エージェントは人間の行動モデルを学習しないので，その行動モデルは省略してあり，エージェントの行動は，設計された初期行動 B_0^A で固定されることになる．

このモデルにおける，インタラクション設計の対象は，エージェントの行動 B_0^A である．

この B_0^A をどのような指針の基に設計するかが問題であるが，IDEA では，**人間の適応を促進するように**設計する．この指針に基づいたインタラクション設計は，さらに “擬人化による親和性の向上” と “探索の効率化” の 2 つの設計指針に分類される．

3.2.1　擬人化による親和性の向上

エージェントの初期行動 B_0^A を人間のもつ初期モデル $M(B_0^A)$ にできるだけ一致するように設計する．この設計も，2.2 において議論した行動の親和性に基づいており，さらに，IDEA では，2.7 で議論した擬人化を最大限に用いて，親和性を高めることが重要である．

また，学習を探索として見た場合，この指針は，探索空間自身をコンパクトにすることに対応する．この設計指針は，3.1 で述べた HCI と類似しているが，後述するように，IDEA では，人間がエージェントを擬人化するバイアスを最大限利用するところが，HCI と異なる．

親和性の向上の設計指針による具体的なインタラクション設計では，エージェントの入力（センシング），出力（行為），内部状態，つまりエージェントの行動が人間に直観的に把握しやすいことが必須であり，これらの親和性を高める方法は，擬人化を実現することで達成される．具体的には，少なくとも以下のことが考慮されるべきである．

- **内部状態の表出：**一般には，推定しにくいエージェントの内部状態が，エージェントの表情，単純な信号などの直感的に把握できる自然な表現で表出されることが望ましい．
- **決定論的な行動：**確率的な行動は，人間にはモデル化しにくいので，行動 B_0^A は決定論的であることが望ましい．
- **行為の文節化：**エージェントの実行した行為が，人間にとって文節化しやすい方がよい．
- **明示的なセンシング：**エージェントがどのセンサを使って，どのような情報をセンシングしているかが明示的にわかる方がよい．

研究例としては，エージェントの状態モデルとして，人間が飼育成長させたソフトエージェントをロボットに憑依させることにより，ロボットと人間の親和性を高める研究 [10]，人間と擬人化エージェントがお互いに相手の表情から内部状態を推定しあうマインドマッピングの相互適応の研究 [16] がある．また，擬人化ならぬ擬犬化を利用して，AIBO の古典的条件付けのために，行動を犬に類似したものを設計することにより，人間の適応を促進する研究 [17] などがある．

3.2.2 探索の効率化

人間の学習 L^H のモデル $M(L^H)$ を基に，人間がエージェントの行動モデルを効率よく学習できるように，エージェントの行動 B_0^A を設計する．このためには，人間の認知モデルである $M(L^H)$ が必要である．これにより，前述の親和性のように，探索空間そのものをコンパクトにすることとは独立に，学習における探索が効率化される．

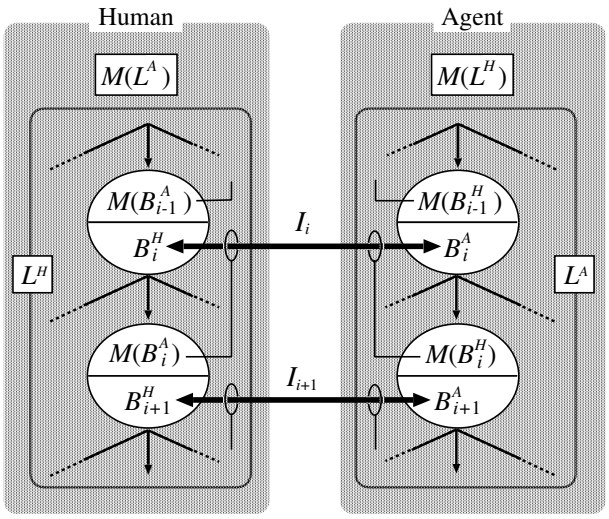

図 3　相互適応

具体的な方法として，人間への適切な教示がある．人間の学習モデル $M(L^H)$ に基づいて，エージェントが人間に適切なタイミング，表現で情報を教示することにより，人間の学習を促進することができる．このとき，2.6 の学習の汎用性を利用して，エージェントの行動を設計することにより，タスクに依存しない促進を実現できる．

しかし，残念ながら，このインタラクション設計は，ほとんど研究例がない．人間の学習の認知モデル $M(L^H)$ の構築を行い，それに基づき探索の効率化を図るような研究が今後期待される．

3.3 相互適応を促進するインタラクション設計

HAI において，人間は対象であるエージェントに不可避的に適応しようとするし，エージェントはユーザ適応するために，人間への適応を試みる．このような人間とエージェントが，同時平行的に相互に適応しあう状態を，**相互適応**（mutual adaptation）と呼ぶ．広義の“相互適応”[16] は，このような意味であるが，IDEA の観点，つまり適応の観点から相互適応を見た場合，お互いが相手の行動モデルを学習しあうことが本質であり，そこに IDEA が解決すべき様々な課題が生じる．このような考えから，IDEA で対象とする相互適応のモデルは，図 2 にエージェントの適応を追加した，図 3 で表される．

この相互適応のモデルにおける設計対象は，エージェントの行動 B^A，エージェントの学習 L^A である．

このモデルにおいて，人間の適応の促進，エージェントの適応の促進，相互適応の促進の 3 つのインタラ

クション設計の目的が考えられる．ただし，人間の適応の促進については，3.2 で既に議論した内容がそのまま適用できる．よって，本節では，エージェントの適応の促進，相互適応の促進を目指したインタラクション設計について議論する．

3.3.1 エージェントの適応の促進

相互適応そのものを促進する前に，まずは，図 3 の相互適応において，エージェントの適応を促進することが考えられる．具体的な方法は，設計対象であるエージェントの学習 L^A，つまり機械学習アルゴリズムの開発と，人間の教示を動機付けるような行動 B^A の設計である．

人間を教師とする機械学習　HAI における，人間に対するエージェントの適応は，人間の行動モデルの学習により実現される．過去の個々のインタラクションから人間の行動モデルを帰納的に学習することは，機械学習における帰納学習の一つであり，エージェントの適応の促進には，人工知能の機械学習が知見を与えれくれる．

ただし，従来の人工知能におけるほとんどの機械学習は，分類されるべきクラスのわかっている訓練例を多量に与えられ，そこからクラス分けのための判別関数（規則）を帰納的に学習する分類学習（classification learning）[8] であり，学習対象を明示的に人間の行動モデルに絞って研究された例は，筆者らの知るところでは見当たらない．

人間の行動モデルの学習として，人間のモデルを学習するには，人間と訓練例であるインタラクションをもつコストが非常にかかる．よって，人間の行動モデルの学習を促進するための機械学習アルゴリズムに求められるのは，単に情報利得の期待値や判別関数と訓練例のマージンなどから判別関数を学習するのではなく，人間の教師であることを考慮していることであり，その観点から以下の研究が重要である．

- **コスト依存学習**（cost-sensitive learning）：通常の分類学習に必要な訓練例は，コストがかからずに得られると仮定されているが，実際は訓練例を得るにはコストがかかる場合がほとんどである．よって，訓練例を得るコストも考慮した分類学習であるコスト依存学習 [14][5] が重要である．特に，IDEA の場合，本当の人間が教師であるので，人間の認知的付加を考慮した機械学習が必要になる．しかし，今のところ，人間の教師に対し，教示のコストを考えた機械学習の研究は見当たらない．また，教示に対する人間の認知的負荷をモデル化する研究もほとんどないようである．
- **能動的学習**（active learning）：単に訓練例が与えるのを待つのではなく，エージェント自身が自分にとって必要な訓練例を進んで獲得する学習が能動的学習である [4][15][11]．計算論的学習理論に基づいて，有効な訓練例を推定する研究が多いが，IDEA では，人間の教示コストを考慮した能動学習が必要である．
- **トランスダクティブ学習**（transductive learning）：通常の分類学習では，大量の訓練例から判別関数を学習するが，少数の訓練例と正負のラベルのない大量の試験例から，通常の帰納学習よりも精度の良い分類子を学習するのがトランスダクティブ学習である [6]．人間の教師から得られる訓練例は非常に少数の場合が多く，もしラベルなし試験例が大量にあれば，トランスダクティブ学習により，学習性能を向上できる．

人間を教師とする高速な機械学習アルゴリズムの 3 つの要素技術については，研究が進みつつある．しかしながら，人間の教師を想定した，人間の行動モデル学習へ応用した研究例はほとんどない．今後の発展が期待される．

人間の教示の動機付け　エージェントの学習 L^A の設計によりエージェントの適応を促進する一方，エージェントの行動 B^A の設計により，エージェントの適応を促進するために，人間の教示への動機付けが考えられる．エージェントの適応の訓練例である，人間とのインタラクションには人間に負荷がかかるため，なんとかして人間に快くエージェントとインタラクションをもってもらい，できるだけ多くの訓練例を得ることが必要である．そのためには，人間がエージェントに楽しく教示ができるような枠組みが必要である．

このような人間の教示への動機付けは，例えば，人間とエージェントの学習速度が同じぐらいでないと，人間はやる気を失う [7] などの報告はあるが，システマティックな評価実験等は研究例がない．しかし，いくつかのケーススタディはある．山田らは，人間とエージェントのマインドマッピングの相互適応 [16] を実現し，そこで，人間とエージェントがお互いの表情からマインド（内部状態）を当て合う相互読心ゲームを設定することにより，ユーザがエージェントとゲームを楽しみながら教示を行うことを可能にしている．

また，人間の教示への動機付けを考慮したもう一つの研究として，人間の表情を擬人化エージェントで再

現する際に，人間のどのような表情を擬人化エージェントのどのような表情に対応させるべきかという表情マッピングをシステムに学習させることを目指した研究がある [18]．この研究では，人間の教示への動機付けとして，(1) 教示内容が以降の処理に速やかに反映されなければ，教示の意欲が失われる，(2) 教示の負担を減らすためには，訓練例はなるべく少ない方が望ましい，(3) どこかの時点で教示が不要にならなければ，安心して利用できない，の 3 点を考慮して，RBF(Radial Basis Function) ネットワークを利用した学習によって，人間の期待するものと近いバイアスを持った追加学習を実現することにより，(1)，(2) を実現すると共に，正例・負例のクラスタリングによって人間の要求する学習精度を推定し，それを達成するように RBF を修正することにより，(3) を実現する手法を提案している．

3.3.2 相互適応の促進

図 3 のモデルにおいて，相互適応そのものを促進することが考えられる．相互適応は，人工知能のマルチエージェントシステムにおける同時学習 [19] と類似したコンセプトであり，同様の課題をもっているが，HAI における相互適応を促進するために，人工的なマルチエージェントシステムにはなかった新たな課題を解決する必要がある．相互適応の促進とは何かについては議論の余地はあるが，3.2，3.3.1 で議論した適応の促進以外の相互適応における課題，そして関連する相互適応の特性として考えられるものを以下に挙げる．

- **適応干渉：**相手のモデルの学習をすることにより，相互適応が起る場合，お互いの学習対象である相手のモデル自身が，相手が適応することにより変化してしまい，追従できない状態を**適応干渉**（adaptation interference）と呼ぶ．
  図 3 の相互適応のモデルで適応干渉を表すと，人間がエージェントの行動モデルの学習が収束したと思ったとき (i ターンのインタラクション) に，エージェントも人間の行動モデルを学習して，$M(B^H)$ および B^A を変更してしまい，その結果，$B^A_i \neq B^A_{i-1}$ となり，人間が学習したエージェントの行動モデルと現在のエージェントの行動が一致しない，つまり $M(B^A_{i-1}) \neq B^A_i$ となった場合が，適応干渉である．また，逆に，エージェントが学習が収束したと判断した時に，$M(B^H_{i-1}) \neq B^H_i$ となる場合もある．マルチエージェントシステムの研究では，主に適応干渉を回避する方法が開発されている．これに対し，IDEA では，単に適応干渉を回避するのではなく，次に述べる適応の非対称性を利用した解決方法の可能性がある．
- **適応の非対称性：**一般に人間の適応能力は，エージェントのそれをはるかに凌駕している．つまり，IDEA では人間とエージェント間に適応能力の非対称性があり，適応の観点からは極端に異種のマルチエージェントシステムである．このことが，IDEA における（相互）適応をより興味深い問題にしている．この非対称性をうまく利用するのが，IDEA の有力な方法論と考えられる．具体的には，相互適応においても，人間の適応を最大限に引き出すことで，学習の速い人間と学習の遅いエージェントからなる異種マルチエージェントシステムが構成され，その学習能力の差により，例えば，多少の適応干渉を回避ではなく，許容することができると考えられる．ここでも，人間の適応を最大限に促進する必要があるため，2.7 の擬人化を利用できる．
- **学習モデルの親和性：**人間のもっているエージェントの学習モデル $M(L^A)$ と実際のエージェントの学習 L^A との間の親和性を高めるように，L^A を設計する．つまり，できるだけ人間がもつエージェントの学習モデルに類似しているように，エージェントの学習アルゴリズムを設計する．そうすることで，人間のエージェントに対する適切な教示を引き出すことができ，相互適応が促進される．
  エージェントの学習速度，時間変化，学習途中の行動の一貫性などが人間に直観的に理解しやすいものであることが，人間の適応を促進するためには望ましい．このような性質をもった機械学習アルゴリズムは，これまで研究されておらず，新しくかつ重要な研究分野であり，IDEA としての重要性のみならず，機械学習への貢献も大きい．
- **エージェントに対する飽き：**通常，人間は，道具に飽きることはない．鉛筆や冷蔵庫に飽きることは，ごく稀な場合である．しかし，対象が，インタラクションを持つこと自体が楽しみであるエンタテイメントロボットやゲーム中のキャラクタの場合は，子供がオモチャにあきるように，ユーザもエージェントに飽きることが多い．相互適応モデルでは，“飽き” は，お互いの適応が完全に収束してしまい，行動モデルにより，相手の行動が全てよそくできて，意外性がなくなった状態と考えられる．この状況に陥らないためには，相互適応が継続していく必要があり，特に人間が完全にはモデル化できないようなエージェントの設計が重

要であろう．

残念ながら人間とエージェントの相互適応をメインに扱った研究はまだ少ない．行動と行動モデルが独立である場合における，人間と擬人化エージェントの相互適応を実現した研究 [16] はあるが，残念ながら，上記の課題が解決されるには至っていない．課題も多いが，今後この相互適応の促進の研究が重要である．

4 まとめ

人間とエージェント間のインタラクション HAI の重要性とそのインタラクション設計の目的，方法論について議論した．特に，HAI では，人間はエージェントに対して適応し，エージェントも人間に対して適応することでユーザ適応を行うというように，“適応” を中心にインタラクション設計を眺めることが重要であり，その場合，人間からエージェント，エージェントから人間への適応，そして相互適応の 3 つの軸について，方法論，従来研究の紹介，今後進むべき方向性について考察した．

謝辞

本解説は,「適応のためのインタラクション設計」研究グループ IΔEA （イデア）のメンバーとの有益な議論が基になっています．記して，IΔEA のメンバーに感謝致します．

参考文献

[1] AIBO Official Site. http://www.jp.aibo.com/.

[2] T. Bickmore and J. Cassell. Relational agents: A model and implementation of building user trust. In *Proc. of Conference on Human Factors in Computing Systems (CHI2001)*, pp. 396–403, 2001.

[3] J. Cassell. Embodied conversational agents: Representation and intelligence in user interface. *AI Magazine*, Vol. 22, No. 4, pp. 67–83, 2001.

[4] D. A. Cohn, Z. Ghahramani, and M. I. Jordan. Active learning with statistical models. *Journal of Artificial Intelligence Research*, Vol. 4, , 1996.

[5] C. Elkan. The foundations of cost-sensitive learning. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, pp. 973–978, 2001.

[6] T. Joachims. Transductive learning via spectral graph partitioning. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pp. 143–151, 2003.

[7] P. Maes. Agents that reduce work and information overload. *Communications of the ACM*, Vol. 37, No. 7, pp. 30–40, July 1994.

[8] T. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[9] MS agent Web site. http://msdn.microsoft.com/msagent/.

[10] T. Ono and M. Imai. Reading a robot's mind: A model of utterance understanding based on the theory of mind mechanism. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*, pp. 142–148, 2000.

[11] T. Onoda and S. Yamada. Relevance feedback with active learning for document retrieval. In *Proceedings of International Joint Conference on Neural Networks*, pp. 1757–1762, 2003.

[12] Roomba Web site. http://www.irobot.com/.

[13] 田村博（編）．ヒューマンインタフェース．オーム社, 1998.

[14] M. Tan. Cost-sensitive learning of classification knowledge and its applications in robotics. *Machine Learning*, Vol. 13, pp. 7–33, 1993.

[15] S. Tong and D. Koller. Support vector machine active learning with applications to text classification. In *Journal of Machine Learning Research*, Vol. 2, pp. 45–66, 2001.

[16] 山田誠二, 山口智浩. 人間と擬人化エージェントによるマインドマッピングの相互適応. 日本知能情報ファジィ学会誌, Vol. 17, No. 3, pp.??–??, 2005.

[17] S. Yamada and T. Yamaguchi. Training aibo like a dog. In *The 13th IEEE International Workshop on Robot-Human Interaction*, pp. 431–436, 2004.

[18] 角所考, 李立群, 伊藤淳子, 美濃導彦. エージェント媒介型表情コミュニケーションにおける利用者の主観に基づく表情マッピングの獲得. 日本知能情報ファジィ学会誌, Vol. 17, No. 3, pp.??–??, 2005.

[19] 荒井幸代, 宮崎和光, 小林重信. マルチエージェント強化学習の方法論 – Q-learning と Profit Sharing による接近 –. 人工知能学会誌, Vol. 13, No. 5, pp. 609–618, 1998.

[20] 山田誠二, 角所考. 適応としての HAI. 人工知能学会誌, Vol. 17, No. 6, 2002.

[21] 山田誠二, 角所考, 新田克己. 特集：HAI ヒューマンエージェントインタラクション. 人工知能学会誌, Vol. 17, No. 6, 2002.

筆者紹介

山田 誠二

1984 年大阪大学基礎工学部卒業．1989 年同大学院博士課程修了．同年大阪大学基礎工学部助手．1991 年同大学産業科学研究所講師．1996 年東京工業大学大学院総合理工学研究科助教授．2002 年国立情報学研究所教授，現在にいたる．工学博士．人工知能，特に，ヒューマンエージェントインタラクション，知的 Web に興味をもつ．

角所 考

1988 年名古屋大学工学部電気学科卒業．1993 年大阪大学大学院工学研究科通信工学専攻博士課程修了．1992〜94 年日本学術振興会特別研究員，1993〜94 年スタンフォード大学客員研究員，1994 年大阪大学産業科学研究所助手，1997 年京都大学総合情報メディアセンター（現 学術情報メディアセンター）助教授．博士（工学）．視覚情報メディア，コミュニケーションに関する研究に従事．

連絡先

山田誠二
国立情報学研究所 知能システム研究系
〒 101-8430 東京都千代田区一ツ橋 2-1-2
Tel&Fax: 03(4212)2562
E-mail: seiji@nii.ac.jp