

人間-エージェントの相互適応系における継続的 相互作用実現のための機能条件

Functional Requirements for Realizing Continuing Interaction in HAI (Human-Agent Interaction) Framework

近藤 敏之
Toshiyuki Kondo

東京工業大学大学院総合理工学研究科
Dept. of Computational Intelligence and Systems Science, Tokyo Institute of Technology
kon@dis.titech.ac.jp, <http://www.ito.dis.titech.ac.jp/~kon/>

若松 良久
Yoshihisa Wakamatsu

東京工業大学大学院総合理工学研究科
Dept. of Computational Intelligence and Systems Science, Tokyo Institute of Technology
ywaka@ito.dis.titech.ac.jp, <http://www.ito.dis.titech.ac.jp/~ywaka/>

伊藤 宏司
Koji Ito

東京工業大学大学院総合理工学研究科
Dept. of Computational Intelligence and Systems Science, Tokyo Institute of Technology
ito@dis.titech.ac.jp, <http://www.dis.titech.ac.jp/staff/itou.html>

keywords: human-agent interaction, continuing interaction, interaction design, personal robot assistant

Summary

In the field of Robotics and AI, there have been gradually increasing the works dealing with Human-Agent Interactions, e.g. pet/humanoid robots developments/applications, interactive teaching, affective computing and so on. However, most of these studies aimed at making robots intelligent, and/or the developed agents mostly depended on human's unavoidable nature, that is, we can implicitly adapt to artificial agents. In contrast, it recently becomes clarified that mutual interaction/adaptation is the critical factor for the interaction design between human and artificial agents. This framework is referred to as HAI (Human-Agent Interaction) design. In this study, we especially focused on a crucial problem in HAI design framework, that is, how to keep a continuing interaction between a human and an agent/robot, because human could be apt to lose interest for agent as time goes on. Due to actualize continuing interactions between agents and the users, it seems to be crucial that both systems possess intrinsic/epigenetic adaptation abilities, and the agent should have an evaluation criteria for estimating internal conditions of users (e.g. curiosity). Based on this consideration, we discuss key functions for realizing a continuing HAI which the agent should possess.

1. はじめに

ペットロボットや掃除ロボットなどの民生向けロボットが市販開始されるなど、日常生活において人間とロボットが相互作用する機会は増加しつつある [Arkin 01]。しかしながらこれらの応用例では、あらかじめ設計された行動戦略を有するロボットを用いてユーザや環境とインタラクションを行い、その様子を観察・評価して行動政策を修正するといった一連の構成論的な設計アプローチを採るものがほとんどである。たとえばペットロボットの設計では、人間の優れた適応・解釈能力を最大限に利用するため、実在する動物をモチーフとしてロボットの身体設計を行うことや、人間のコミュニケーションモダリティに則したインタラクションを考慮するといった経験的な設計指針が多く用いられている [Shibata 01]。

これに対し、人間と自律エージェントのパーソナルな

関係を実現するためには、その相互適応の過程を支援するメカニズムや、エージェントの側からも人間に対して能動的に適応することの重要性が認識されつつあり、認知心理、人工知能、ロボティクスなどの研究分野において、人間と自律エージェントの相互適応系設計に関する研究開発が活発に行われている。

たとえば [山田 03] では、人間とロボット (エージェント) がインタラクションを通して相互適応する過程を HAI (Human-Agent Interaction) と呼び、その適応過程を促進するインタラクションの設計指針の重要性が指摘されている。また [稲色 01] では、ユーザとのインタラクションの経験を蓄積・統計処理し、行動選択機構を確率表現することでロボットの環境適応が試みられている。[Katagami 02] では、ロボットの環境適応を促進するための教示 (インタラクション) のしかたが考察されている。一方 [Breazeal 02] のように実ロボット Kismet によ

るロボットの社交性デザインを対象とした研究もある。

このように現在人間との対話的インタラクションを通して適応する様々なエージェントシステムが提案されているが、エージェントが適応するためには長期間に渡って継続的にインタラクションを行うことが前提として必要である。しかしながら、一般に人間が人工エージェントに対して関心を保ち続けることは容易ではなく、現状では人間側の負荷が大きい。したがって人間-ロボット間の相互適応を促進するためには、ロボットは与えられた目的を達するのみではなく、インタラクションを通してユーザの関心を維持するような行動を起こすことも必要であると考えられる。

本研究では、通常は興味を消失したりマンネリ感を抱きがちな人間-エージェント間の継続的なインタラクションを恒常的に維持するためにエージェントが有すべき機能的条件を明らかにすることを目的とする。本稿では、エージェントがインタラクションを通してユーザの内部状態を推定するためのモデルを構築し、その有効性を実験的に検証する。

2. インタラクションの確率推論モデル

人間と自律エージェントが相互に適応する過程では、人間側の適応とエージェントの認識・行動系の適応が同時に関与するため、環境のダイナミクスはインタラクションを通して変化する。したがって、エージェントの適応を実現するためには、エージェントは環境の変化が相手(人間)の行動に起因するものであるか(受動的インタラクション)、あるいは自身の働きかけによるものであるか(能動的インタラクション)を推定できることが必要である。すなわちインタラクションにおける受動性とはエージェントが外部入力に起因した出力を生成する過程に対応し、また能動性はエージェントの内部状態の変化に起因して生成された行動出力に対するユーザからの応答を受け取るまでの過程に対応する。

本研究ではインタラクションの因果関係をベイジアンネットワーク(Bayesian network)を用いて確率的に表現し、これをエージェントはユーザの内部状態推定モデルとして利用する。ここでベイジアンネットワークの各ノードに記述される情報(確率変数)は、環境の状態やユーザからの働きかけに相当する自律エージェントの感覚情報、感情状態や欲動に相当する内部状態、およびエージェントの行動選択確率であり、詳細は4.3節にて説明する。

図1左下に示すように、受動的インタラクション(Passive interaction)を考慮したベイジアンネットワークでは、各ノードはそれぞれの確率計算の因果関係に応じて、順にエージェントの感覚情報ノード $S = \{s_1, \dots, s_l\}$ 、内部状態ノード $I = \{i_1, \dots, i_m\}$ および行動選択ノード $A = \{a_1, \dots, a_n\}$ に対応する。一方、同図右下に示すよ

うに、インタラクションの能動性を考慮したベイジアンネットワーク(Active interaction)では、エージェントの内部状態ノード I 、行動選択ノード A および感覚情報ノード S の順となる。

たとえば、エージェントがある行動 a_k を選択したときに感覚状態が s_i である場合、内部状態 i_j が作用素として関与し、インタラクションの受動性に関する各ノード間の相互関係は以下ようになる。

$$P(i_j | a_k \wedge s_i) = \frac{P(a_k | i_j) P(i_j \wedge s_i)}{P(a_k | s_i)} \quad (1)$$

これは感覚情報が原因、選択された行動がその結果となるインタラクション中の因果関係において、その関係に作用素として働くエージェントの内部状態の影響度を表す。

一方、エージェントの内部状態が i_j 、その時の感覚情報が s_i である場合に、選択された行動 a_k が作用素として関与する能動的インタラクションにおける、各ノード間の相互関係は以下ようになる。

$$P(a_k | s_i \wedge i_j) = \frac{P(s_i | a_k) P(a_k \wedge i_j)}{P(s_i | i_j)} \quad (2)$$

これは同様に内部状態が原因、フィードバックされたユーザからの入力とその結果となる因果関係において、その関係に作用素として影響を与えた行動の影響度を表す。

3. ユーザの内部状態推定機構を用いた相互適応システム

前節で述べた確率的状態推論モデルを用いて人間の内部状態を推定し、継続的な相互適応過程を支援するシステムを提案する。まずエージェントの基本構成について概説した後に、提案する相互適応システムについて説明する。

3.1 エージェント内部のシステム構成

図1に本研究で用いた自律エージェントのシステム構成を模式的に示す。エージェントは知覚・注意システム(Perception and Attention system)において、外界(すなわち人間や環境)からの入力刺激の識別を行い、対応する動機システム(Motivation system)内の欲動(興奮、緊張等)や感情(嬉しい、悲しい等)の状態を活性化させる。次いで行動決定システム(Behavior system)では、前段の動機システムにおける内部状態(欲動・感情状態)に基づいてエージェントの行動パターンを決定し、モータシステム(Motor system)に運動指令を伝達する。本来ならば人間の入力パターンから動機システムへのマッピングならびに内部状態から行動出力までのマッピングは、認知心理の知見を基に設計すべきところであるが、現段階ではこの部分は手設計している。

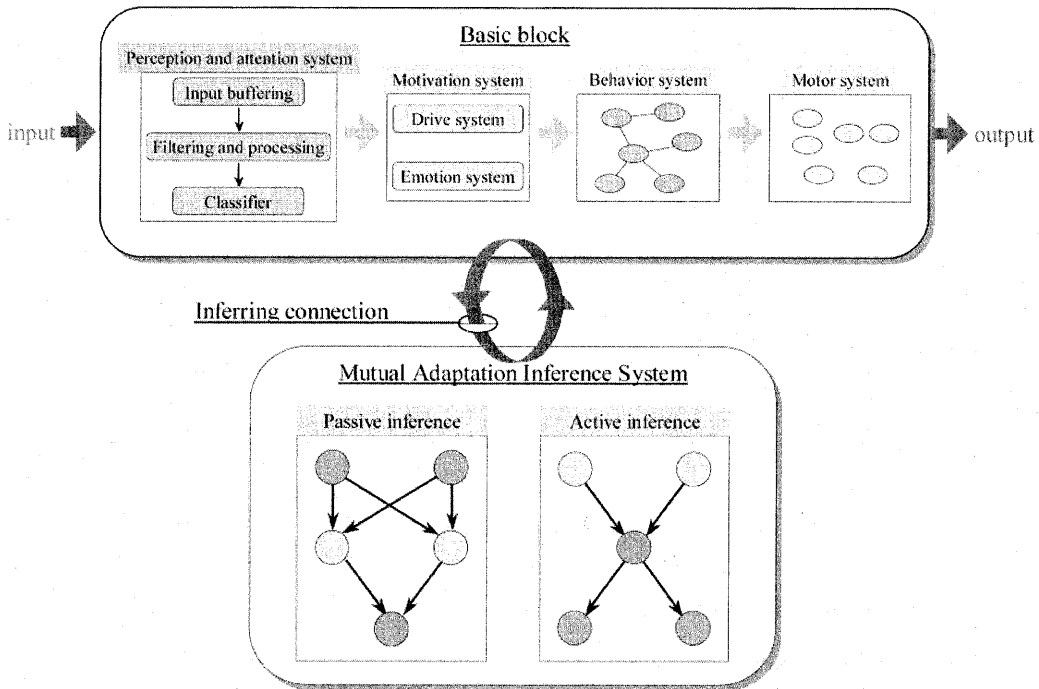


図 1 システム構成

3.2 状態推定機構を用いた相互適応

本研究ではユーザの能動的インタラクションが減少する状態、すなわちユーザがエージェントの行動戦略を概ね把握したために飽きてしまったと考えられる状態を定量的に推定し、その推定量に基づいてエージェントの側から能動的にインタラクションを促進することで相互適応過程を支援するシステムの実現を目的としている。これは図1に示すように、エージェントの行動戦略を決定する基本構成部が、ベイジアンネットワークで構成される状態推定器によって修飾を受けるという形でモデル化することが可能である。

4. 表情エージェントを用いたインタラクション実験

本研究では提案システムをユーザとのインタラクションを通して表情を変化させるアニメーション・エージェントに実装する。まず本実験において計測するデータや計測方法について説明し、次いで提案システムを実装した表情エージェントの外観や内部構成について述べ、最後に実験結果についての考察を行う。

4.1 実験課題

人間のエージェントに対する適応的な特性を観測するためにインタラクションの頻度と時間の関係に着目する。

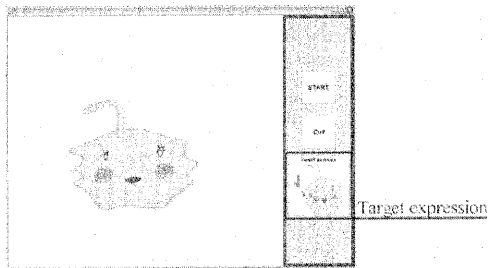
被験者には、コンピュータディスプレイ上にアニメーションとして表示されるエージェントとマウスを使った簡単なインタラクションを行わせ、その間のマウスの押下頻度とインタラクション時間（課題を達するのに要する時間）をデータとして計測した [Picard 97]。

ここで実験の課題は、図2(a)のようにディスプレイ上に表示されているエージェントの表情を、ランダムに提示される「初期表情」から、マウスを使ったインタラクション（「撫でる」または「叩く」の2種類）により、「目標表情」へと変化させることである。

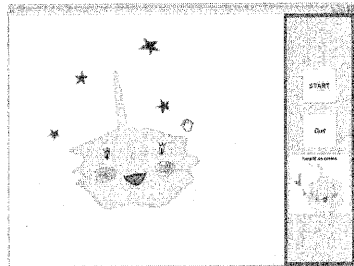
4.2 エージェントの構成

エージェントの行動出力は表情の変化としてユーザに提示される。図3(b)に示すように、本実験で用いたエージェントの表情は [Russell 97] に対応づけて、pleasure-arousal 空間上で定義される。ここで図中に斜体で記述されている *happy*, *elated*, *surprised*などは Motivation systemの感情状態と Behavior systemで生成されるエージェントの行動に対応し、非斜体の *excitement*, *calm*, *stress*などは Motivation systemの欲動に対応している。したがって、エージェントの内部状態となる欲動および感情状態がそれぞれ5種類および9種類あり、また選択される行動が9種類ある。

また、Perception and attention systemにおけるユーザの入力パターンの識別は、本実験ではマウスの操作を



(a) Initial state.



(b) After interaction.

図2 実験環境.

「撫でる」もしくは「叩く」というクラスに分類することに対応する。

4.3 実験結果

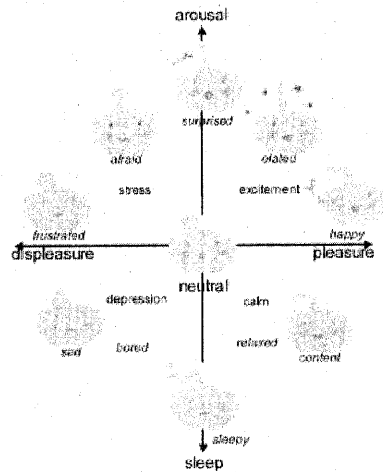
実験は研究室の学生の協力を得て行われた。被験者は3名の男子学生である。

実験は被験者それぞれに14試行回ずつ行った。実験データを被験者3名について平均した結果を図4に示す。

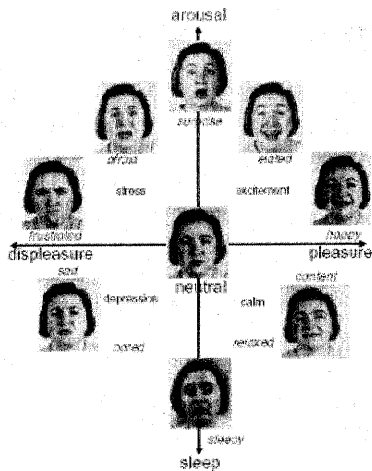
同図より、試行が進むにつれてインタラクション時間は徐々に減少し、インタラクションの頻度は試行の進行に対して上に凸の山型となることが見てとれる。つまりインタラクションの初期段階では被験者が試行錯誤する状態が長めのインタラクション時間および少ないマウスのクリック回数として現れていると考えられる。実験の中盤では初期段階に比べてインタラクション時間は短くなるが、マウスのクリック回数は増加していることから、被験者がインタラクション実験に対して徐々に慣れてきていることが伺える。そして実験試行の終盤ではインタラクション時間は短くまたマウスのクリック回数も少なくなることから、この段階で被験者はインタラクション実験に適応していると推測される。

また実験の後に被験者に対してアンケート調査を行った結果から以下のことが示唆された、

- 実験開始直後は、キャラクターの表情の変化を良く見て考えてから操作を行っていた。
- 次第に目標として指示される表情を直感的に導くこ



(a) Agent's expression



(b) Facial expression by Russell

図3 pleasure-arousal空間上で定義されるエージェントの表情.

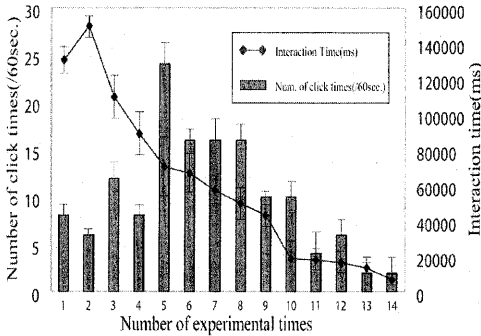


図4 実験の結果。

とができるようになった。

- 後半は飽きてしまった。

今回の実験では、エージェントに対する人間側の適応はインタラクションを通して段階的に進行し、大きく分けて3つの段階(試行錯誤→慣化→適応)を経て適応状態に至ることが示唆された。

また同実験においてエージェントがユーザとのインタラクションの末に獲得した確率的状態遷移モデルを表1に示す。この表は、能動的インタラクションの場合の結果を示している。ここではBehavior systemで選択された表情が *excitement*, Motivation systemの感情状態が *happy* かつ欲動が *excitement*, Perception and attention systemで知覚されるユーザからの入力が *stroke* (撫でる)である場合の各生起確率を示している。

ただし、式中の表記は以下にしたがう；

- Behavior systemを頭文字 *b* で表し、添え字が生じられる各振る舞いに次のように対応する：

b_{spr} : *surprise*,

b_{elt} : *elated*,

b_{hpy} : *happy*,

b_{cnt} : *content*,

b_{slp} : *sleepy*,

b_{sad} : *sad*,

b_{frt} : *frustrated*,

$b_{ afr}$: *afraid*,

b_{ntr} : *neutral*.

- Motivation systemのEmotion systemを頭文字 *e*, Drive systemを頭文字 *d* で表し、添え字によりそれぞれ

e_{hpy} : *happy*,

d_{exc} : *excitement*

と表す。

- Perception and attention systemを頭文字 *p* で表し、知覚されるユーザからの入力を添え字によりそ

表1 エージェントの状態遷移確率表の一例

$P(b_{spr} e_{hpy}, d_{exc}, p_{stk}, \bar{p}_{hit})$	0.0263
$P(b_{elt} e_{hpy}, d_{exc}, p_{stk}, \bar{p}_{hit})$	0.5157
$P(b_{hpy} e_{hpy}, d_{exc}, p_{stk}, \bar{p}_{hit})$	0.3896
$P(b_{cnt} e_{hpy}, d_{exc}, p_{stk}, \bar{p}_{hit})$	0.0251
$P(b_{slp} e_{hpy}, d_{exc}, p_{stk}, \bar{p}_{hit})$	0.0083
$P(b_{sad} e_{hpy}, d_{exc}, p_{stk}, \bar{p}_{hit})$	0.0000
$P(b_{frt} e_{hpy}, d_{exc}, p_{stk}, \bar{p}_{hit})$	0.0118
$P(b_{ afr} e_{hpy}, d_{exc}, p_{stk}, \bar{p}_{hit})$	0.0108
$P(b_{ntr} e_{hpy}, d_{exc}, p_{stk}, \bar{p}_{hit})$	0.0124

れぞれ

p_{stk} : *stroke*(撫でる),

p_{hit} : *hit*(叩く)

と表す。

この表から、エージェントにとって原因子となる内部状態、作用素となる行動の選択および結果子となるユーザからのフィードバックである知覚の3要素の関連に基づき、 b_{elt} : *elated* という行動がエージェントの *happy* という感情状態と *excitement* という欲動によって誘発され、またその行動がユーザの撫でるという行動を引き出しやすいということが示されている。つまり、インタラクションを通して要素間の関係性が形成されていく。

5. ま と め

本稿では、現在開発中のインタラクティブエージェントについて興味を消失したりマンネリ感を抱きがちな継続的なインタラクションを維持するための機能として従来の基本構成にベイジアンネットワークによるユーザの状態推定機構を導入したインタラクションシステムの提案を行った。提案システムは、これまでのHAIの枠組みで扱われているエージェントとは異なり、埋め込まれた内部ダイナミクスだけではなくインタラクションに則して獲得される確率的な原因子-作用素-結果子の因果関係を用いてエージェントの行動生成プロセスを修飾することが可能である。人間と表情エージェントのインタラクション実験では、インタラクションの頻度と時間の関係から、図4に示す結果を得た。同結果は、エージェントに対する人間の適応過程は大きく分けて3つの段階(試行錯誤→慣化→適応)を経て進行することを示唆している。また、同実験を通してエージェントの内部に表1に示すようなユーザの状態遷移モデルが獲得されることを確認した。

今後の課題としては、長期間に渡って継続的なインタラクションを維持するためには、推定したユーザの内部状態に基づいてエージェント側から能動的にインタラクションを働きかけることが有効であることを明らかにするため、エージェントの行動生成機構の改良と実験デー

タの収集を並行的に行っていく予定である。

謝 辞

本研究の一部は文科省科研費(課題番号:14350227, 14750362)および国立情報学研究所との共同研究経費によるものである。ここに謝意を表す。

◇ 参 考 文 献 ◇

- [Shibata 01] Takanori Shibata, Kazuo Tanie: Physical and Affective Interaction between Human and Mental Commit Robot, Proceedings of the IEEE International Conference of Robotics and Automation, pp.2572-2577, 2001
- [Arkin 01] Ronald C. Arkin, Masahiro Fujita, Tsuyoshi Takagi, Rika Hasegawa: An Ethological and Emotional Basis for Human-Robot Interaction, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2002), 2002
- [山田 03] 山田誠二, 各所 孝: ユーザ・システム間の適応のためのヒューマン-コンピュータインタラクションのデザイン, システム制御情報学会誌, Vol.47, No.4, pp.39-44, 2003
- [Breazeal 02] C. Breazeal: Designing Social Robots, The MIT Press, Cambridge, MA, 2002
- [稲邑 01] 稲邑哲也, 稲葉雅幸, 井上博允: PEXIS:統計的経験表現に基づくパーソナルロボットとの適応的インタラクションシステム, 電子情報通信学会誌, Vol.J84-D-I, No.6, pp.867-877, 2001
- [Russell 97] James A. Russell: Reading emotions from and into faces: Resurrecting a dimensional-contextual perspective, in James A. Russell and José-Miguel Fernández-Dols (eds.), The Psychology of Facial Expression, New York: Cambridge Univ., pp.295-320, 1997
- [Picard 97] Rosalind Picard: Affective Computing, The MIT Press, Cambridge, MA, 1997
- [Katagami 02] Daisuke Katagami, Seiji Yamada: Interactive Evolutionary Robotics from Different Viewpoints of Observation Proceedings of the IEEE/RSJ International Conf. on Intelligent Robots and Systems (IROS2002), 2002