

観測指向モデルによる自律エージェントの自発的な行動選択

向井 淳[†] 今井 倫太[†]

人間との対話などのインタラクションを行う自律エージェントやロボットでは、行動をどう決定するかという問題がある。対話などの複雑なタスクでは、どの行動を取ればよいかといった基準をあらかじめ決定できないからである。筆者らはエージェントが自発的な行動の選択を行うための観測指向モデルというモデルを提案している。本論文では、そのモデルの挙動についてシミュレーション環境で調査を行った。その結果、単なるランダムとは異なる複雑なふるまいをすることがわかった。

Spontaneous Behavior Selection of Autonomous Agents by Observation Oriented Model

JUN MUKAI[†] and MICHITA IMAI[†]

For the autonomous agents and robots which interact with humans, they have a problem to select their actions. In the case of complex tasks such like conversations, the rules to determine their behaviors for the situation cannot be defined a priori. For this problem, we proposed observation-oriented model for the spontaneous selection of agents' actions. In this paper, we investigate the behavior of our model by simulations. As the result, our model is proved to generate complex rules unlike simple random selections.

1. はじめに

近年、人間と対話などのインタラクションを行うエージェントやロボットの研究が進められている。特にロボットが日常生活の場で活動をするためには、人間と対話などのインタラクションが必要不可欠であると考えられる。本研究では、自律動作する対話エージェントの行動ルールをどう設計するか、という問題について扱う。

伝統的には、対話エージェントのルールは決められた if-then ルールによって構成される。つまり、現在の状況から条件部に合致した反応を返すことで対話を実現する。対話ロボットでも状況は大きくは変わらず、if-then ルールによってロボットの行動を決定するものも多い⁴⁾。

また、単純な if-then ルールではなく、単機能のモジュールと依存関係によって設計する行動ベースのアプローチ¹⁾によるロボットもある。たとえば Kismet²⁾は頭部のみコミュニケーションロボットであり、動く物体やカラフルな物体に注意を向けるといった行動

が組込まれている。この行動の組み合わせによって、幼児と世話焼きという社会関係を成立させている。

しかし、既存のアプローチだけでは対話のような複雑なタスクの実現は困難であると思われる。if-then ルールの場合、設計者はロボットが遭遇しうる状況をあらかじめすべて想定した上で、それに対する反応を組込まなければならない。また、状況によってロボットの反応は決められる。つまり、対話者は自分の働きかけに対するロボットの反応を容易に学習し、推測できるようにになってしまうため、対話が阻害されると考えられる。行動ベースのアプローチを取れば、ロボットの遭遇しうる状況を精密に想定しなくても適切な行動が実現できるが、ロボットの反応が状況に対して固定されるということは変わらない。また、行動ベースのアプローチでは、コミュニケーションを成立させるような行動モジュールを構成するのは極めて複雑である。

たとえば、エージェントと人間がちよつとした雑談を楽しむという状況を考える。エージェントが提示できる情報は複数あるが(たとえば天気予報やニュースの情報など)、どれを提示したらよいかということは事前には決められない。かといってその都度、どのトピックにするか人間に選択させるのであれば対話というインタフェースを用いる利点を失ってしまう。エー

[†] 慶應義塾大学理工学部情報工学科
横浜市港北区日吉 3-14
TEL: (045) 560-1070
{mukai, michita}@ayu.ics.keio.ac.jp

ジェントは自分でどれかを選ばなければならないが、どれを選ぶべきかは推論できない。

もちろん、どれかを選択させるのであれば単に乱数によって選択するという方法が考えられる。しかし、完全にランダムに行動を選択するとロボットの行動は無秩序になってしまう。また、完全にランダムであるということは人間がどんな働きかけをしてもロボットは影響を受けないため、コミュニケーションは阻害される。状況に応じて乱数の分布を変化させるようなモデルも考えられるが、結局は設計者がロボットの直面する状況をあらかじめ想定しておかなければ適切に乱数分布を設計できないため、同じ困難さがともなう。

以上をまとめると次のようになる。第一に、人間とのコミュニケーションというタスクの場合、行動ルールを完全に固定するのは人間がメニューを提示されてその中から選択するのと同じであり、望ましくない。しかし、完全にランダムに選ぶと、人間との文脈が共有できないためコミュニケーションが成立しない。つまり、基本的には何らかのルールに従っているが、何らかの乱数によってそのルール自体が変化し、行動に変化が生じることが望ましいと考えられる。

この問題に対し筆者らは、自律エージェントが行動を選択するために観測指向モデルを提案している⁵⁾。このモデルでは、エージェントは周囲の環境やエージェントの行動を観測し、その結果をもとに自分の行動を決定することで、安定した行動を実行する。ただし、その観測に意図的に誤りを混入させることで、行動を時間的に変化させることができる。

本論文では、この観測指向モデルを紹介し、そのふるまいについてシミュレーション環境で調べる。そして、観測指向モデルの行動ルールは単純なランダムとは異なる複雑さを持つことを示す。

本論文の構成は以下の通りである。まず2章で観測指向モデルについて概略を示す。次に3章で、この動作を確認するためのシミュレーションを説明する。4章で動作結果を示し、考察する。最後に、5章で結論を述べる。

2. 観測指向モデル

観測指向モデルは、郡司らの研究³⁾をもとにした行動選択モデルである。観測指向モデルでは、エージェントは他のエージェントや自己の行動をつねに観測しておき、その結果をコンテキストと呼ばれる領域に蓄積する。そして、コンテキストから行動を決定する。直前の自分の行動や周囲のエージェントの行動を観測することで過去の行動と類似した行動が出力され、さ

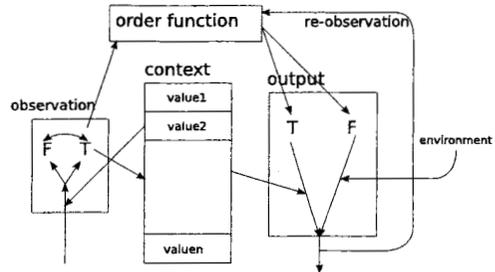


図1 観測指向モデルの概略

らにその行動結果がふたたび観測されコンテキストに蓄えられるというループを構成することで、安定した行動が生成できる。ただし、行動の観測には意図的な誤りを混入させる。そのため、コンテキストにはときどき、過去の行動とは異なるデータが入ることがある。また、コンテキストはある有限の数しか履歴を保存せず、古いデータは破棄されるため、異なるデータの混入が場合によって全体に波及することがあり、行動そのものを変化させる。

観測指向モデルの基本的な構成を図1に示す。

中央にあるのがコンテキスト (Context) であり、左側が行動の観測をする部分 (Observation) にあたる。上側にあるのが秩序関数 (Order Function) と呼ばれ、行動の大きな指針を与える。右側は行動の生成を行う部分であり (Output)、ここでは秩序関数の更新も行なわれる。以下に、観測と行動生成の具体的な方法について説明する。

まず最初に観測部の説明を行う。観測部では、センサ入力から存在するほかのエージェント2つと自分自身の行動を受け取り、観測する。観測のとき、コンテキストに蓄えられている過去の行動パラメータからランダムにひとつを選択し、観測した行動パラメータと比較し、一致するかしないかという真理値を計算する。また、一致するときはコンテキストに類似する行動をそのエージェントが取ったということだから、観測した行動パラメータをコンテキストに追加する。ただし、コンテキストは有限の長さしか持たないため、追加するときにはランダムに1つの要素を選択し、その要素を削除する。こうすることでコンテキストは類似した行動パラメータを保持する傾向を持つ。

観測部では周囲2つのエージェントと自分自身の行動を観測するから、観測結果としては3つの真理値が得られる。これは図で示されたように秩序関数 (Order Function) に渡される。秩序関数では、この3つの真理値から1つの真理値を取り、出力部に渡す。この真

理値は行動のおおまかな指針を与えるものである。なお、秩序関数は内部にテーブルを持っており、入力となる3つの真理値と出力となる1つの真理値の対応関係を保持しているが、後述するようにこのテーブルは更新されることがある。テーブルが更新されることで、同一の観測結果であっても出力が変化し、行動が複雑に変化する。

行動出力部では、秩序関数から1つの真理値を受け取る。この値は、真のときは「現在のコンテキストに従って行動する」を意味し、偽のときは「無秩序に行動する」を意味する。真のときはコンテキストからランダムに1つの要素を選択し、これを新しい行動パラメータとする。また偽のときは、周囲の環境からランダムな行動パラメータを生成し、出力する。出力された行動パラメータに従ってエージェントは行動をする。そして、この行動をふたたび観測部が観測することになる。

このように、行動を観測したら行動指針の決定し、その指針にもとづいて行動を決定し、行動したらまたそれを観測する、というサイクルを繰り返す。このサイクルのなかで観測によってコンテキストは次々と更新されていき、類似する行動パラメータを蓄積しつづける。秩序関数が真を出力するようになると、コンテキストから行動パラメータが出力され、その行動を行い、実行した行動を観測するというサイクルになるから、安定して類似する行動パラメータが維持されることになる。

以上は通常の動作である。ただし、このままではエージェントの行動パラメータはまったく変化しなくなってしまう。そこで観測に意図的に誤りを与え、行動に変化をもたらす。以下で、この仕組みを説明する。

まず観測部では、ある確率 p_a において観測結果を逆転させる。偽が真に変化することにより、コンテキストには類似しないデータが混入することになる。また真理値が逆転することにより、秩序関数の異なるルールを参照し、行動が変化する可能性がある。

また、出力部においても、出力しようとしている新しい行動パラメータを観測部と同じようにコンテキストを用いて観測しなおす。この観測結果は秩序関数のルールの更新に用いる。つまり、秩序関数の出力が偽なので周囲環境から行動パラメータを生成するとしても、この結果がたまたまコンテキストと一致したとすると、「今回の行動は真だった」とみなして、秩序関数の保持するテーブルのうち、該当するエントリを偽から真に変更する。逆に、秩序関数が真を出力したとしてもコンテキストが多種の値を保持していた場合に

は、出力する行動パラメータを観測しなおすと一致しないことがある。この場合も、今回の行動は結果的には偽だったとみなし、秩序関数の保持するテーブルを更新する。さらに、この出力部における観測でも、ある確率 p_a によって真理値を逆転させる。

以上のように誤りを導入し、秩序関数を動的に変化させることでエージェントの行動に変化を与える。具体的には次のような過程を経る。まず、周囲のエージェントの行動から、コンテキストに異種のデータが混入される。コンテキストは複数のデータを保持するから、そのような場合にもこの異種データはすぐ削除される可能性が高い。しかし、出力する行動パラメータの観測のときに、この異種データが選択される可能性がある。そうすると、秩序関数が真を返すのに結果が異なることがあり、秩序関数が偽を出力するように更新されることがある。あるいは、観測に導入される誤りによって秩序関数が更新されることもある。理由はともあれ秩序関数が更新されると、今度はエージェントは無秩序な行動を出力するようになる。自分の無秩序な行動は繰り返し観測される。こうした行動はコンテキストの中身とは一致しないために基本的には無視されるが、意図的に導入した誤りがあるために、結果として次々とコンテキストに入れられる。これが繰り返されるうちに、コンテキストの保持するデータと、出力部が生成する無秩序な行動パラメータがたまたま一致することが起こる。すると秩序関数はふたたび真を出力するようになる。コンテキストは異種のデータを保持するため、秩序関数が真でも行動はしばらく無秩序なままであるが、コンテキストは類似するデータを保持する傾向を持つために、しだいにコンテキストは同種のデータに統一されるようになり、出力する行動パラメータはふたたび安定する。ただしこのとき、コンテキストがその時点で保持している多種類のデータのうち、どれに統一されるかはわからない。このため、出力する行動パラメータは、はじめと異なるものになりうる。このように安定状態から無秩序な状態に移移し、それからまた安定状態に戻るという過程を経ることで、行動パラメータはしばらく安定しながらも、時間的に変化するというたふるまいが実現される。

3. 実験環境

観測指向モデルのふるまいを調べるため、シミュレーション環境で動作させた。シミュレーション環境を図2に示す。

この図において、小円はそれぞれエージェントを意味し、小円から出る線が、そのエージェントの「向い

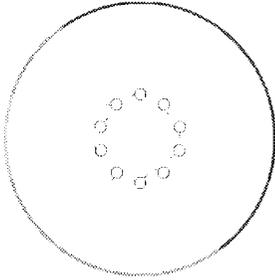


図2 シミュレーション環境

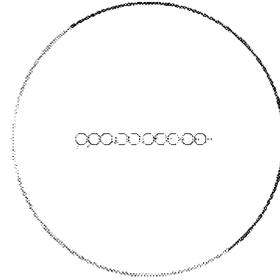


図3 直線配置パターン

ている方向」を意味する。各エージェントの取りうる行動は自分の向きを変更することだけである。

一方、大円は壁であり、壁には様々な色で塗られているという状況になっている。この色をエージェントの行動パラメータとする。つまり、コンテキストには色データが蓄積される。また、そしてシステムは色データを出力する。するとエージェントは周囲からその色の存在する方向を探索し、自分の向きをその方向へと変更する。ただし、エージェントは自分の向いている方向を連続的にしか変化させることができない。つまり、ぜんぜん違う方向に向きを変えたいとしても、現在の方向から少しずつ変化させていかなければならない。このため、あるエージェントがある時点である方向を向いていたとしても、それが移動途中であるのか、それとも目的地であるのかということとはわからない。

また、エージェント同士は、自分に最も近い2つのエージェントの行動を観測して行動パラメータを得る。互いに行動パラメータを直接交換するようなことは行なわない。そこでこの環境では、そのエージェントが現在向いている方向の色を現在のそのエージェントの行動パラメータであると推定する。先述したようにこの推定方式は、必ずしも現在の行動パラメータを正しく反映しているとは限らない。つまり、観測による行動パラメータの推定には誤りが存在しうるが、この誤りによって、より複雑な行動が生成される。

本シミュレーションでは、壁面は均等な幅の10色に塗り分けられているとし、ロボット10体を環境内に一定のパターンで配置した。図2は、円形に配置するパターンである。また、直線的に配置するパターンも実施した。直線配置パターンを図3に示す。

観測指向モデルでは、コンテキストの大きさと、意図的に導入された誤り率 p_o および p_a によって挙動は変化する。事前の予備実験から、コンテキストの大きさは8個、 $p_o = 20\%$ 、 $p_a = 2\%$ とした。また、1回の試行では10,000ステップ実行した。

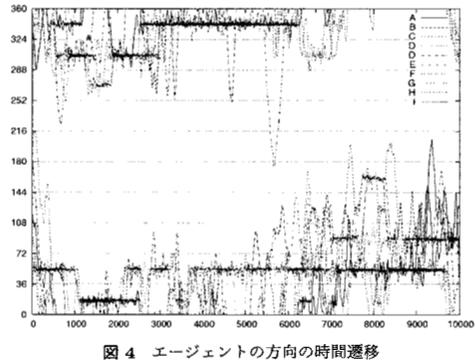


図4 エージェントの方向の時間遷移

4. 結果と考察

本節ではシミュレーション環境での動作した結果を示し、その結果から検討を行なう。まず、動作させたときにエージェントが実際にどのように自分の方向を変化させたかを調べる。図4は円形の配置のときの典型的な動作例である。この図において横軸は時間(ステップ)、縦軸は、その時点でそのエージェントが向いている方向を示している。

このように、エージェントは向いている方向を激しく変化させつつ、たとえば4000ステップ目付近で2つのグループに分かれ、この2つのグループの間を行き来するエージェントがいることがわかる。エージェントは位置的に近いエージェントの行動を観測するため、ある位置の近くのエージェントによってグループが形成される。ただし円形配置の場合はエージェントの配置は対称的なのでグループの境界に位置するエージェントは両方のグループの影響を受け、変化しがちである。6000ステップ以降では一方のグループはもう一方のグループに合流して消滅するが、7000ステップ目からまた別なグループが新たに形成され、というようにエージェントは自己の行動基準を変化させつつ

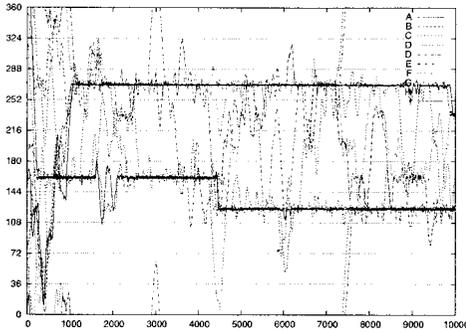


図5 直線配置でのエージェントの方向の時間遷移

相互に影響を与えることで、グループの構成を変化させている。

直線配置の典型的な例を図5に示す。

直線配置でも大きな差は見られなかった。ただし、配置によって生じる違いはある。直線配置の場合、端では相互に観測しあう関係に偏りがあるため、両端がそれぞれグループを構成しやすい。つまり、線の両端がそれぞれ異なるグループとなり、線の中央付近では両方のグループの影響を受けるという状況が多く発生する。この例でも、かなり初期から明確に2つのグループに分かれ、両者のあいだを行ったり来たりするエージェントの存在が見られる。

さて、我々に関心があるのは、このようなエージェントのふるまいにどのような特性が見られるかということである。そこで、実際にエージェントが向いていた方向の特性を調べた。

エージェントは、ある一定期間を特定の方向に向けるが、しばらくするとほかのエージェントの影響などで向きを変える。そこで、特定の方向にどれだけ連続して向いているかを調べた。100回の試行を行い、各期間がどれくらいの頻度で出現するかを調査した頻度分布が図6である。ただし横軸は行動基準が維持されるステップ数、縦軸はその出現頻度であり、見やすさのため両対数軸でプロットした。

まず目につくのは、期間の長さ30ステップ以下の場合に大きく上下に変動している部分である。しかし、短い期間の場合には観測指向モデルの特性よりはむしろ色領域の大きさやエージェントの回転速度といった環境の影響の方が強いと考えられるため、ひとまず無視する。この短い期間のものを除けば、頻度は30-40ステップあたりでピークを迎え、あとはおおむね直線状に頻度が減少している、いわゆるべき分布に類似している。

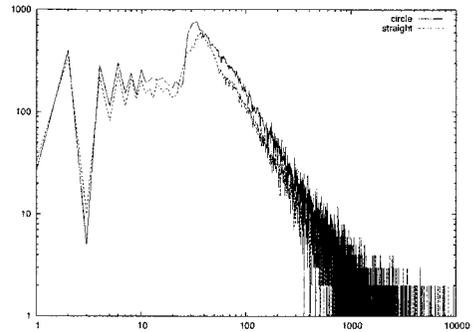


図6 特定の方向を向いていた期間の頻度分布

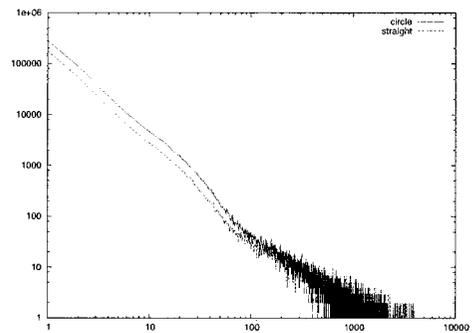


図7 行動基準が維持される時間の頻度分布

ただし、このプロットは実際にエージェントがどういう行動を取ったかということを知るには良いが、環境に依存する部分が多い。そこで、エージェントがコンテキストから出力した行動基準を調べることで、観測指向モデルではコンテキストがどのように変化するかを調べた。同じく100回の試行を行い、同一のものがどれくらい連続して出力されるかを調べ、回数ごとに頻度を計算する。この頻度分布を図7に示す。横軸と縦軸の意味は先程と同じであり、やはり両対数でプロットした。

こちらのグラフでは長さが1のものが最も大きく、ほぼ直線的に減少していることが見てとれる。そして実際に向いている場合の例と比較すると、期間の長さが30ステップを越える場合にはよく似た形状のグラフとなっており、コンテキストの内容とエージェントの行動が対応づけられる。つまり、コンテキストの中身も同じように、非常に長い安定した行動が生成されつつ、ときおり極めて短い不安定な行動が混ざるといことである。極めて短い期間の出力が頻繁に起こるのは、コンテキストに異種のデータが混ざっている場

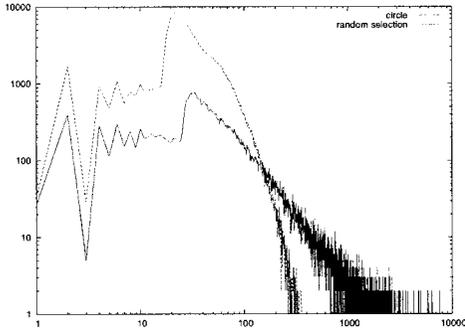


図8 ランダムとの比較

合を示している。この場合、コンテキストからは様々なデータが出力されるため、期間が短くなる。

また、単純にランダムに選択する方式を実装し、観測試行モデルとの比較を行った。このランダムモデルでは、ステップごとにランダムに行動基準となる色を選択する。ランダムモデルでエージェントが実際に向いた向きについて、同じように100回の試行を行い、頻度を計算した。エージェントを円状に配置した場合の観測指向モデルとともにプロットしたものを図8に示す。

やはり期間が30ステップ以下の場合と同じように上下しており、この部分は環境に依存するという仮定が確認された。また、ランダムモデルでも30-100ステップの頻度が極めて多いという分布は似ているが、より急激に減少する。図からはランダムモデルでも100ステップを越える期間で維持される行動も存在していることがわかる。これは単なる乱数分布よりは長い、通過するのにかかる時間がかかったり、途中で方向転換をする可能性があるためであると考えられる。そして、それより長いものについてはまったく存在しない。

観測指向モデルではベキ法則に従った頻度分布を構成し、単なるランダム選択とは異なるふるまいを示した。では、この分布にはどのような意味があるか。

ランダム選択のモデルを構築するとき、単に毎ステップでランダムに選択するのではなく、ランダムな一定期間、行動パラメータを維持するというモデルも考えられる。ただしその場合にも、その一定期間はどのような分布で決定するのか、という問題が残される。観測指向モデルの分布の場合、極めて長い期間維持されるものがある一方で、極めて短いものも存在している。つまり、ある程度の長さは維持しうるがいつかは変化する、という特性が実現されていると思われる。

5. 結 論

本研究では、筆者らが提案している観測指向モデルを紹介し、そのふるまいを調べた。

観測指向モデルは自律エージェントが勝手に行動基準を生成し、変化させることで多様な行動を出力するというモデルである。観測指向モデルの出力する行動はふだんは安定しているが、時間の経過により勝手に不安定化し、ふたたび安定するときに異なる行動に遷移する。このため、因果的に行動をしつつ、時にその予測が外れるような行動が実現されると考えられる。

ただし、観測指向モデルの出力する行動は確かに単なるランダムと異なるものの、どのような特性となっているかは調べられなかった。本論文では、行動基準が維持される期間の頻度分布によって、単なるランダムと異なり、観測指向モデルがベキ分布となっていることを確認した。このため、単なるランダムと比較して安定した行動が生成できることがわかった。また簡単な収束モデルを実装してふるまいを比較した。この結果、観測指向モデルでの行動基準は何か収束して変化しなくなるのではなく、しばらくすると変化しうることを示唆された。

本研究では、動作は単純なシミュレーション環境で確認した。今後はこの結果をもとに、実際のアプリケーションに応用する予定である。

参 考 文 献

- 1) Arkin, R. C.: *Behavior-Based Robotics*, MIT Press (1998).
- 2) Breazeal, C. and Scasselatti, B.: How to build robots that make friends and influence people, *Proc. of IROS99* (1999).
- 3) Gunji, Y.-P., Sadaoka, H. and Ito, K.: Bootstrapping System Defined by Inconsistent Relation between Boolean and Non-Boolean Algebra, *Applied Mathematics and Computation*, Vol. 79, pp. 43-97 (1996).
- 4) NEC: PaPeRo. NEC Personal Robot Center, <http://www.incx.nec.co.jp/robot/>.
- 5) 向井淳, 今井倫太: 属性変換を利用したコミュニケーションロボットの實現, SI 部門講演会 (SI2004) 予稿集, 計測自動制御学会 (2004).