3E1-01

行為の選択に熟考と強化学習を併用するBDIアーキテクチャの実現について

Toward realization of BDI architecture that uses a combination of deliberation and reinforcement learning in selecting actions

新出尚之[*1] 高田司郎[*2] 山川 宏[*3] 宮崎和光[*4] 太田正幸[*5]
NIDE Naoyuki Shiro TAKATA Hiroshi YAMAKAWA Kazuteru MIYAZAKI Masayuki OHTA

[*1]奈良女子大学理学部
Faculty of Science, Nara Women's University

[*2]近畿大学理工学部
School of Science and Engineering, Kinki University

[*3]富士通研究所
Fujitsu Laboratories LTD.

[*4]大学評価・学位授与機構
National Institution for Academic Degrees and University Evaluation

[*5]産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

In this paper, we propose an extended BDI architecture which selects actions by a combination of deliberation and reinforcement learning. BDI architecture can act consistently by retaining intentions, but to select a suitable plan by only deliberation is often hard. Using reinforcement learning in combination with deliberation in BDI architecture, an agent can learn suitable decision while keeping consistent acts toward the agent's goal.

1. はじめに

人間が周囲の環境に依存して自分の目標を達成する行為を決めようとするとき、信念などを用いた推論と、試行の繰り返しによる学習を併用することが多い。例えば ①カヌーレーシング [高田 04a] のような課題では、スタート前にはコースの大まかなプランニング、レース中には学習で獲得した反射的な行動選択能力、の双方を要する ②複雑な交通機関網を用いてA 地点から B 地点に通おうとする時、実際の最適経路を推論によって絞り切れなければ、最適となりそうなものをいくつか試してみて、その繰り返しで適切な選択を求めていくことがある、などのような状況が考えられる。

そこで我々は、BDI アーキテクチャ[Singh 99, 高田 01] と強化学習 [Sutton 98] の結合により、熟考的な能力と反復による選択学習能力を共に持ち、柔軟な行動のできるエージェントの実現を目指している [高田 04a]。

BDI アーキテクチャは、「意図」の概念を明示的に持ち、目標達成のために選択した意図を保持して、その意図のもとに最適なプランを実践的推論による熟考で選択することで、目標達成に向けた一貫した動作を行うエージェントアーキテクチャである。強化学習との結合により、目標指向で、かつ状況に適した柔軟な行動選択を行うことが可能になると考えられる。

上記①については [高田 04b] で、強化学習によって獲得したスキルをサブプランとして BDI 側で用いる手法を提案した。本稿では②のようなケースにおいて、条件を満たすプランを推論と強化学習の併用において選択する方式について述べる。

2. 熟考と強化学習の併用

2.1 例題

非常に簡単な例として、奈良市街から京都に行くことを考える。可能なプランとしては、JR 利用、近鉄利用、車利用の 3 つがあるとする。いま、京都市内が渋滞しがちであるという理由で、選ぶプランを前 2 者のどちらかに絞ることまではできたとする。しかしこの 2 者は、本数や乗り換え駅での連絡の悪さなど一長一短があり、熟考のみでいずれかに決定することは難しい。

そこで、最終的な決定に何らかの手法による強化学習を用い、試行を繰り返すことで選択を改善していくことが考えられる (実際には学習効果が出るほど何度も両地点間を旅行するのは非現実的かも知れないが、何らかのシミュレーションは可能であるものと仮定する)。

2.2 BDI アーキテクチャ

BDI アーキテクチャは、動的に変化する環境を知覚し、合理的に問題解決を行うためにプランを選択しながら動作する、熟考型エージェントの内部アーキテクチャである。陽に表現されたエージェントの心的状態 (意図・目標・信念) やプランライブラリ・イベントキュー、およびそれらを参照・更新するインタプリタなどで構成される。

インタプリタは、環境知覚 (信念) と自らの目標から、プランライブラリの参照によって実行すべきプランを選択し、これを実行する意図を形成する。この意図は保持され、それを実行すべき時が来れば実行する。意図には一貫性や整合性が要請され、今持っている意図と矛盾する行動を取ったり、達成できないと信じていることを意図したりすることはない。また、もともとの目標を達成する必要がなくなったなどで、意図を持続する理由がなくなった場合はそれを破棄する (例えば、会合に出席すべく京都に向かっている途中で、その会合が取り止めになった旨の連絡が来たような場合)。

2.3 強化学習の導入

BDI アーキテクチャのインタプリタの主要部は、以下のようなループである。B, G, I はエージェントの信念・目標・意図を表す。

```
do
    options := option-generator(event-queue, B, G, I);
    selected-option := deliberate(options, B, G, I);
    update-intentions(selected-option, B, G, I);
```

連絡先: 新出尚之, 奈良女子大学, 〒630-8506, 奈良市北魚屋西町, Tel: 0742-20-3435, nide@ics.nara-wu.ac.jp

```
    execute(I);
    get-new-external-events();
    drop-successful-attitudes(B, G, I);
    drop-impossible-attitudes(B, G, I);
until quit.
```

すなわち、大まかには、外界からのイベントと心的状態 (およびプランライブラリ) から option-generator() によって、次に実行可能な意図の候補となるプランを選択し、deliberate() によって実際に実行する意図を選択する。そして update-intentions() によって、選択された意図となっているプランの本体 (実現手順基本行為およびサブゴールの列) のうち次に実行すべきものをポイントし、基本行為であれば execute() で実行、サブゴールであれば新たな目標に加える。次いで外部イベントの走査、心的状態の更新 (達成した、あるいは撤回したと信じる目標や意図の除去) を行う。これを繰り返す。

deliberate の過程は、信念・目標・意図を陽に記述できる時相論理体系であり BDI アーキテクチャの論理的基盤である、BDI logic で記述されたルールによる推論で行われる (現在完全には実現できておらず、Prolog による一階述語での推論の範囲で行っている)。そこで、今回はこの推論の過程に、強化学習による解選択の機能を組み込む。

すなわち、Prolog で実現された deliberate が、候補のうちルールで与えられた条件に合致するプランを全解探索で求める際に、(通常の Prolog のような先頭のルールから適用する方式でなく) 全解のうちから強化学習によって候補となるプランを選択するようにすればよい。今回の例では、

```
nominate(JR利用) :- bel(京都市内渋滞).
nominate(近鉄利用) :- bel(京都市内渋滞).
```

のようなルールを記述しておき、`bel(京都市内渋滞)` が成り立った場合の `nominate(X)` の解「JR 利用」「近鉄利用」を、この順に選ぶのでなく学習によって選択する。強化学習の方式は特定されない (Sarsa, Q 学習など、適切なものを選べばよい)。但し、そのプランを実行した場合の報酬がどこかから与えられる必要がある (例えば event-queue を介して外界から渡すなどが考えられる)。

2.4 結合の利点

BDI アーキテクチャと強化学習の結合の利点は、BDI 側が「意図」を持っており、これの保持による一貫した行動を取れる点、および必要な場合は意図の破棄によって別な目標への行動に移れる点である。

通常の強化学習では、環境モデルが一定であることが前提であり、周囲の状況に応じて目標が動的に変わるような場合には向かない。例えば、途中で京都に向かう必要がなくなった場合、意図による制御がないと、漫然と不要な行動を取り続けるかも知れない。京都に向かう意図を破棄して、次なる何か他の目標を選んでそのための意図を形成し、そのもとに新たな動作を起こせば、そのような不都合はなくせるだろう。

また、選択すべきプランの集合によって、政策の学習や利用を別々に行うこともできる。例えば、東京に行くプランがあり、その本体が ①京都駅へ行く ②(新幹線などで) 東京に向かう、というサブプランから構成されているとき、京都駅へ行く部分と東京に向かう部分とで別な学習を行うことができる。大規模な問題を扱う手法として多く研究されている階層的強化学習 [Parr 97] と同様なことができるが、この場合も目標の動的な変更に対応できることが BDI を用いる利点となる。

3. 結合のモデル化

BDI アーキテクチャの利点の 1 つは、BDI logic を用いて、エージェントの振る舞いや性質を形式的に議論できる点である。強化学習の機構と結合する場合、強化学習のモデルとして標準的に用いられる MDP での、確率的な手の選択や状態遷移が記述できる必要がある。CTL, CTL*を拡張して確率的な状態遷移を記述可能にし、モデルチェックアルゴリズムを与えた体系 PCTL, PCTL*が得られており [Hansson 94, Morioka 99]、例えば「次の時刻に 60%の確率で ϕ、40%の確率で ψ が成り立つ」は「$[X\phi]_{\geq.6} \wedge [X\psi]_{\geq.4}$」と書かれる。これを BDI logic に拡張して、例えば上記の信念を「$\mathsf{BEL}([X\phi]_{\geq.6} \wedge [X\psi]_{\geq.4})$」のように表現し、形式的仕様記述などに使うことが考えられる。

4. まとめ

本稿では、BDI アーキテクチャと強化学習 [Sutton 98] の結合によって、熟考的な能力と反復による選択学習能力を共に持ち、柔軟な行動のできるエージェントの実現を目指す一環として、プラン選択に熟考と学習を併用する方式について述べた。今後は、実装を通じて問題点の検討などを行っていきたい。

参考文献

[Hansson 94] Hansson, H. and Jonsson, B.: A Logic for Reasoning about Time and Reliability, *Formal Aspects of Computing*, Vol. 6, No. 5, pp. 512–535 (1994)

[Morioka 99] Morioka, T.: Automatic Verification of Probabilistic Systems, in *Collection of Reports for CSC2108 (Fall '99)* (1999), http://www.cs.toronto.edu/~chechik/courses99/csc2108/projects/

[Parr 97] Parr, R. and Russell, S.: Reinforcement Learning with Hierarchies of Machines, in Jordan, M. I., Kearns, M. J., and Solla, S. A. eds., *Advances in Neural Information Processing Systems*, Vol. 10, The MIT Press (1997)

[Singh 99] Singh, M. P., Rao, A. S., and Georgeff, M. P.: Formal Methods in DAI: Logic-Based Representation and Reasoning, in *Multiagent Systems*, pp. 331–376, The MIT Press (1999)

[Sutton 98] Sutton, R. S. and Barto, A. G.: *Reinforcement Learning: An Introduction*, The MIT Press (1998)

[高田 01] 高田 司郎, 五十嵐 新女, 新出 尚之, 榎本 美香, 間瀬 健二, 中津 良平：マルチエージェント環境において意図的に言語行為を遂行する合理的エージェントの基本設計, 電子情報通信学会論文誌, Vol. J84-D-I, No. 8, pp. 1191–1201 (2001)

[高田 04a] 高田 司郎, 山川 宏, 宮崎 和光, 新出 尚之, 長行 康男, 酒井 隆道：強化学習と BDI の統合について—カヌー・レーシングを例題とした統合手法の考察, 第 18 回人工知能学会全国大会論文誌 (2004), 1F1-02

[高田 04b] 高田 司郎, 新出 尚之, 山川 宏, 宮崎 和光, 太田 正幸：強化学習で獲得したスキルを実践的推論する BDI の 実現方式について, エージェント合同シンポジウム (JAWS2004) 講演論文集, pp. 517–524 (2004)