3k-06

ロボカップサッカーエージェントの模倣による行動学習

林 崇博, 近山 隆
東京大学 工学系研究科*

1 はじめに

ロボットサッカーは、マルチエージェントシステムやその学習を、試してみる良い例題である。だが、教師付き学習 (supervised learning) を行なう以上、教師となる、効率的な学習のために適切な入力と出力の事例が必要となる。現在、サッカーエージェントの学習のための事例を集めるのは、難しく時間がかかり、個人の技術に依存する部分が多い。これは、人間の負担を減らすという機械学習の目的に反する。また、不必要な学習をエージェントがすることにもなりかねない。

こうした問題を解決するため、この研究では、既存のエージェントから得た入出力関係を、教師信号として使用する。システムとしては、階層型ニューラルネットワークを用い、学習法としては、バックプロパゲーションを使う。

2 サッカーエージェントの学習例と問題点

サッカーエージェントに学習をさせた論文は、[1] [4] などがあり、この二つに共通しているのは、エージェントをランダムに行動させて、成功した行動を教師信号にしていることである。成功したという判断は、人間がすることもあれば、ゴールにボールが入ったか否かという単純な物であったりする。

人間が判断を下せば、おかしな行動を教師信号として選んでしまうかも知れない。それを回避しようとすれば、細心の注意と熟練を要する。

ボールがゴールに入ったかを判断の基準にすれば、偶然、ボールが入ってしまった場合を教師信号として選んでしまうかも知れない。偶然のことを因果性があるとして学習を行えば、おかしな行動を行うエージェントが出来上がるだろう。それを回避するためには、やはり手間がかかる。

人間の手間を省くための機械学習のために、人間が一生懸命教師信号を集めるようでは機械学習の目的に反する。

3 教師信号を既存のエージェントから得る利点

教師信号を既存のエージェントから得れば、次のような利点がある。

- 得るのが簡単である: 学習させたい状況で、動かしていれば、教師信号は得られる。
- 矛盾が少ない: 既存のエージェントが乱数を使っていなければ、同じ入力に対して、異なる出力をすることは無い。
- 偶然が少ない: ランダムに動かしてうまくいった場合を教師信号とすると、偶然うまくいった場合を選んでしまうかもしれない。
- 起こる確率の高い場合について、学習を行える: 実際に起こり得る状況下で既存のエージェントを動かし、教師信号とするので、起こりやすい状況について、学習をすることができる。

4 柔軟性と計画性

柔軟性、反応性 (reactive) と計画性 (deliberative) のバランスをどうとるべきなのか。これは、なんらかの行動を行うエージェントを作る際の一番の課題と言える。

野田氏によれば [4]、サッカーシミュレーションは、以下のような特徴を持つ。

1. 緻密さよりもロバスト性が重要である。
2. リアルタイムに多様に状況が変化する。
3. チームプレイをした方が有利である。
4. 敵の作戦に応じて作戦を変更するほうが有利である。

Behavior learning of RoboCup agents by observing existing agents

Takahiro Hayashi, takahiro@logos.t.u-tokyo.ac.jp
School of Engineering, the University of Tokyo
Takashi Chikayama,chikayama@klic.org
Department of Frontier Informatics,
School of Frontier Sciences,
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113, Japan

3つ目を除いて、これらの特徴は、計画を優先させるのか、それとも柔軟性を優先させるのか、という問いに対して、柔軟性を優先すべきであることも示唆している。

[2] [3] では、ドリブルやパス、セットプレイなどの高次な行動概念を導入している。

しかし、これは反応性重視のシステムである。ただし、外部からの情報に直接反応するわけではない。外部からの情報を解釈して内部状態を作り、内部状態に応じて行動する。（図1参照）

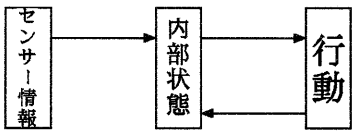

図 1: 計画性の実現

例えば、連絡を取り合い、オフサイドトラップをかけることになったとしよう。その時、クライアント内部ではオフサイドトラップを目指す内部状態になり、オフサイドになるか、もしくは予想外の状態になるまで、内部状態がオフサイドトラップを目指す状態であり続けるので、オフサイドトラップのための行動を続ける。

こういった方法により、一見、計画性重視の行動をとる、柔軟性重視のクライアントができる。

5 この方法の限界

内部状態を完全に模倣することができれば、既存のエージェントから、教師信号を得ることにより、行動を完全に模倣することができるはずである。

しかし、内部状態は、外部から知ることはできない。ただし図1のように、内部状態を決めるのは、外部からの与えられるセンサー情報と、行動からくるフィードバックである。センサー情報と行動から、内部状態を推測をすることは無理ではない。

例えば、ボールが見えていないとき、過去に見た情報から位置を予測するのは、簡単である。また、何らかの行動の結果、ボールの位置が相対的にどう変化したかを割出すのも、簡単である。

ただ、sayでの通信は、その内容が推測できない。

まとめれば、その内部状態と、推測した内部状態の誤差が、この実験の成否を決めることとなる。

6 ニューラルネットワークの構成

大きく分けて、二つのシステムを作った。行動の種類を決めるもの、パラメタを決めるものである。行動を決めるものは、行動の種類ごとに出力を持ち、最大値をとった行動をする。パラメタはそれぞれ学習はニューラルネットワークを用いて行った。

入力は、ボールの位置、ボールの速度、他のプレーヤの位置などである。

7 結果

図2は、自作した、ボールに向かって行って蹴るだけのクライアントから、学習を行った結果である。学習するクライアントの内部状態の生成機構は教師クライアントと同一である。また、教師信号は1000個、テストデータも同じもの使っている。そして、1000個の誤差の平均が図2の誤差である。

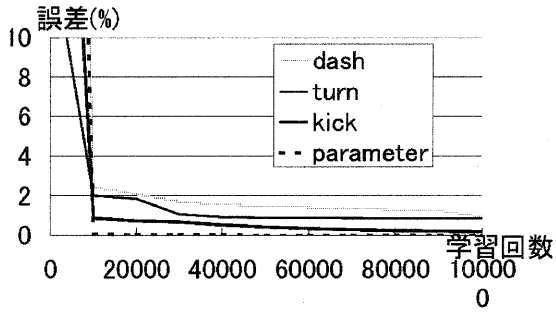

図 2: 学習結果

内部状態が完全に同一なので、うまく学習が行えている。

参考文献

[1] Peter Stone and Manuela Veloso. A layered approach to learning client behaviors in the robocup soccer server. *Applied Artificial Intelligence*, December 1998.

[2] Mannuela Veloso and Peter Stone. Task decomposition, dynamic role assignment, and low-bandwidth communication for real-time strategic teamwork. *Articial Intelligence 110(2)*, pp. 241–273, June 1999.

[3] Mannuela Veloso, Peter Stone, and Michael Bowling. Anticipation: A key for collaboration in a team of agents. *The 3rd International Conference on Autonomous Agents*, October 1998.

[4] 松原 仁 野田 五十樹 開一夫. サッカーにおける協調的な行動の学習. *Multi-Agent and Cooperative Computation*, December 1995.