Frontiers in Robotics　グローサリー

Action【行動】
ここでは環境の状態の遷移を引き起こすロボットの運動のことをいう。

Actor-critic 【アクター・クリティック】
強化学習をベースにしたロボットを制御するためのアーキテクチャの一つで、価値関数を表すクリティックと方策を表現しているアクターから構成される。

AIBO【アイボ】
ソニーが発売していた子犬型などのペットロボット（エンターテイメントロボット）。

ASIMO【アシモ】
本田技研工業が開発し、ホンダエンジニアリング株式会社が製造している本格的な二足歩行ロボット。身長 130㎝。質量 52kg。歩行速度は最大 2.7㎞/h。現段階では市販されておらず、本田技研工業に問い合わせる事によって、賃貸することができる。

Average reward 【平均報酬】
方策を変化させないでロボットを制御し続けたときに得られる報酬の時間平均のことをいう。これは対応するマルコフ過程の定常分布のもとでの報酬の期待値に一致する。

Cognitive developmental robotics 【認知発達ロボティクス】
ヒトを理解するためのロボティクスをベースにした構成論的なアプローチのことをいう。理解の対象となる生物の発達・学習モデルをロボットに実装し、環境中で実際に動作させることによって得られた挙動を解析して、発達・学習モデルの設計論を構築することを目指す。

Coulped Neural Oscillator 【結合神経振動子】
ヤツメウナギなどの脊椎動物の神経生理学研究に始まり，運動パターンの生成機構は主に脊髄に存在し，その周期的で適応的な挙動が結合神経振動子系として数理的に理解できることが分かってきた．神経振動子とは，神経細胞を少数結合することで周期的な活動を達成でき，また外部から適切なレンジの周期入力を与えることでその周期に引き込まれるなどの性質を持っている．神経振動子を多数結合したものが結合神経振動子であり，例えば一列に結合しておくと，頭部から尾部へ波を送ることが容易に実現でき，魚の泳動運動を再現することができる．

HRP2【エイチ・アール・ピー・ツー】
経済産業省が実施するプロジェクト「人間協調・共存型ロボットシステム（HRP）」の一環として開発されたヒト型ロボット。身長 154㎝、体重 58kg。全身の自由度は 30 自由度あり、ホンダの ASIMO と比較した際のアドバンテージとしては、腰に新たに 2 自由度を持たせていることで、これによって転倒した状態から起きあがるのが容易になるだけでなく、離れたところに手を伸ばすといった単純な動作もよりスムーズになった。

Importance sampling【重点サンプリング】
モンテカルロ積分をする場合に、サンプリングする点を均等にばらまくのではなく、「重要そうな」ところに重点的にサンプリングする点をばらまいて積分することを重点サンプリングと呼ぶ。

Markov decision process【マルコフ決定過程】
マルコフ性を持つ確率過程のことをいう。

Markov property【マルコフ性】
状態の遷移が、そのときの状態と行動にのみ依存し、それ以前の状態や行動とは無関係

である性質をいう。多くの強化学習法がマルコフ性を仮定しており、ロボティクスではマルコフ性を満足するような状態表現を構築することが一つの鍵となる。

Particle filtering【粒子フィルタ】

カルマンフィルタなどのベイズフィルタを非線形確率過程にも応用できるようにしたもの。事後分布を重みつき粒子の集合で表現する。各粒子をシステムの時間発展にしたがい時間発展させ、観測過程に依存して重みを変更するという計算を繰り返すことで、逐次的に事後分布を近似することができる。

Policy 【方策】

学習ロボットの制御方法を定義するもので、各状態で行動を選択する確率を与える。とくに、もっとも確率の高い行動だけを選択する方策を greedy policy 【貪欲な方策】という。

Policy gradient 【方策勾配】

平均報酬を方策を表すパラメータで微分したもので、方策のパラメータを方策勾配の方向に修正するように学習は実施される。この方式による強化学習法は policy gradient based reinforcement learning 【方策勾配に基づく強化学習】と呼ばれる。

Q learning【Q 学習】

マルコフ決定過程において、最適な価値関数を学習するための強化学習のアルゴリズムの一つで、ロボティクスにおける行動学習法としてもっとも広く使用されている。しかし、学習に膨大な時間を必要とする、関数近似との親和性が悪い、離散行動しか扱えない、といったことから、ロボットの行動学習法として適していないとの指摘がなされている。

Reinforcement learning【強化学習】

機械学習の一手法で、事前に与えられた報酬からロボットの方策を学習するための枠組みである。ロボットの行動学習の手法としてだけではなく、動物における行動学習の計算モデルとして注目されている。

Reward【報酬】

強化学習における目的を定義し、状態と行動の組に対する即時的な望ましさをスカラー値で与える。目標状態にだけ正の値、それ以外の場合は 0 を割り当てることが多いが、これは疎な報酬関数と呼ばれ学習に膨大な時間を要する。

Self localization【自己位置同定】

ロボットに搭載されたセンサ情報と行動の系列から、ロボットの世界中心座標系における位置を推定することをいう。

State【状態】

ロボットの運動を記述するための変数であり、たとえば関節角度や関節角速度、環境中における位置や移動速度などである。

Support vector machine【サポートベクタマシン】

パターン識別器を構成する手法。カーネルトリックと呼ばれる方法を用いて、非線形の識別関数をも構成できる。マージン最大化と呼ばれる、未学習データに対して高い識別性能(汎化性能) を得るための工夫があるため、現在知られている多くの手法の中でも最も識別性能の優れた学習モデルの一つであると考えられている。

van der Pol Oscillator 【ファンデルポール振動子】

一つの閉曲線に軌道が収束してゆくリミットサイクルを持つ非線形力学系の一例.

数学的に研究が進んでおり，安定した振動や引き込みいった性質を持つため，
ロボティクスでは神経振動子や結合神経振動子の代わりに利用されることがある.

Value function 【価値関数】

一般には状態の関数または状態・行動対の関数であり、将来にわたって得られるであろう報酬の総量を表す。報酬とは異なり、状態と行動の組に対する長期的な望ましさを与える。value function based reinforcement learning 【価値関数に基づく強化学習】では、各状態において価値関数を最大にすることを目的とする。

View-based representation 【視点依存の表現】

ロボットにカメラなどのセンサが搭載されていたとき、環境の見え方そのものを用いる表現のことをいう。見え方とロボットの位置が一対一対応でない場合は視点依存の表現はうまく作用しない。

World coordinate representation 【世界座標系での表現】

環境中に固定された座標系でロボットや環境中の物体の位置などを記述する表現のことをいう。一般にロボットのセンサ情報だけから世界座標系での位置を得ることは困難であるため、自己位置同定などの方法によって推定しなければならない。