CIT自律移動_勉強会_2回目

～マルコフ決定過程と動的計画法～

千葉工業大学 未来ロボティクス学科

上田研　b3 池邉 龍宏

目次

- ナビゲーションについて
- 詳解確率ロボティクス10章の1部のマルコフ決定過程について
- 来週の内容

ナビゲーションについて

- 時間がなかった。

マルコフ決定過程について

- ロボットと環境の相互作用について
- 状態
  - 環境は、状態で特徴付けられる。
  - ロボットと環境の、将来に影響する全ての局面として状態を考えることが便利である。
  - あるいくつかの状態変数は，ロボットの近くにいる人々の位置のように、時間経過と共に変化する性質を持つ。
  - 他の状態変数は、（ほとんどの）建物の壁のように静的である。

マルコフ決定過程について

- 状態はxで表される。
- xに含まれる状態変数は問題に応じて変化する。
-時刻tの状態はx_tと表記される。

- 状態とは
  - ロボットの姿勢はグローバル座標系に対するロボットの位置、向きで構成される。
  - 環境中の物体の位置と特徴も、状態変数である。物体は、樹木であったり、壁であったり、ある壁面、床面に印された点だったりする。
  -ランドマークになり得る物体は、環境中で目立つ不変の特徴を持ち、確実に認識できるものである。
  -移動物体と人の位置や速度も、状態変数となり得る。

マルコフ決定過程について

- センサが壊れているかどうかも状態変数になり得る。
- 電池で動くロボットにとっては、電池の残量も状態変数になる。

- 観測や制御について
  - ロボットが過去の全てのセンサ計測値や制御動作を記録しておくことができると仮定したとすると、二つの異なるデータが定義できる。
  - 環境計測データは，各時刻の環境の状態に関する情報を与える。
  - 時刻t の計測データをz_t で表す。
  - 制御データは，環境の状態の変化に関する情報を与える。
  - 時刻t の制御データをu_tで表される。

マルコフ決定過程について

- ほとんどの場合，我々は小さな時間のずれを単純に無視する。(例えば、レーザセンサは高速に環境をスキャンするものの微妙に時間差が発生するが、我々は計測結果をある時刻に一瞬で得られたものと単純に仮定する。)

-ある時刻に複数の計測データが得られる場合にも簡単に拡張できる。

$z_{t_1:t_2} = z_{t_1}, z_{t_1+1}, z_{t_1+2}, \dots\dots, z_{t_2}$

時刻t_1からt_2までに得られた全ての計測値の集合

$u_{t_1:t_2} = u_{t_1}, u_{t_1+1}, u_{t_1+2}, \dots\dots, u_{t_2}$

時刻t_1からt_2までの状態の変化

- 確率的発生法則

- 条件付き確率は状態遷移確率や計測確率として用いられる。

- 状態遷移確率は、ロボットの制御utによって，どのように環境の状態が時間発展するかを規定するものである。

マルコフ決定過程について

- 計測確率は、環境の状態xからどの計測値zが得られるかということに関する確率法則を示す。

マルコフ決定過程について

- マルコフ決定過程について
  - マルコフ決定過程（MDP）を勉強する上で、１つずつマルコフ過程から少しずつ変数を増やして理解していくのが分かりやすいかもしれない。
    - マルコフ性
    - マルコフ過程
    - マルコフ報酬過程
    - マルコフ決定過程

マルコフ決定過程について

- マルコフ性

  - マルコフ性とは、現在の状態Xnが与えられた時、過去のいかなる情報(X0, X1, … , Xn−1)も、Xn+1を予測する際には無関係であるという性質。
  例えばサイコロを振る時、何回かサイコロを振っていたとしても、出る目は過去に依存しない。

  - 天気がマルコフ性を持つ場合を仮定すると。
  一般的に、翌日の天気は今日までの雲の動きを参考に予測される．つまり、翌日の天気を予測するためには、昨日以前の天気の情報も利用する必要がある。
  しかし、天気がマルコフ性を持つ場合、明日の天気は今日の天気にのみ左右される(昨日より前の天気は関係ない)ため、昨日以前の天気の情報を利用する必要がなくなる。

マルコフ決定過程について

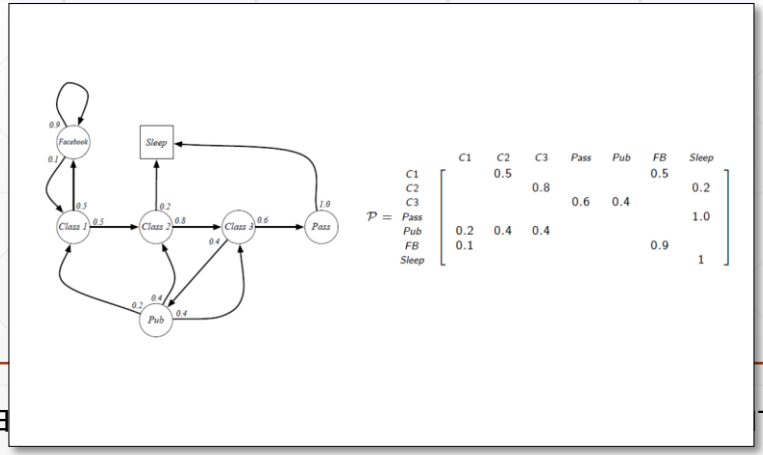
・マルコフ性に反する事象

- x_t に含まれない環境中の、モデル化されないダイナミクス（自己位置推定の例では、動いている人々やそれらの人々がセンサ計測値に与える影響）

マルコフ決定過程について

- マルコフ過程

  - マルコフ過程とは、マルコフ連鎖とも呼ばれ、タプル⟨S,P⟩で表現される。ここで、Sは状態の集合、Pは状態遷移確率の行列をそれぞれ表す。例えば下の図のように状態遷移表と状態遷移行列で表現できる。確率については、計算式で求められることもあるのでしょうが、実験的に求めることが多い。データを蓄積する過程で、状態遷移確率は適宜更新される。行成分の和が１になる。



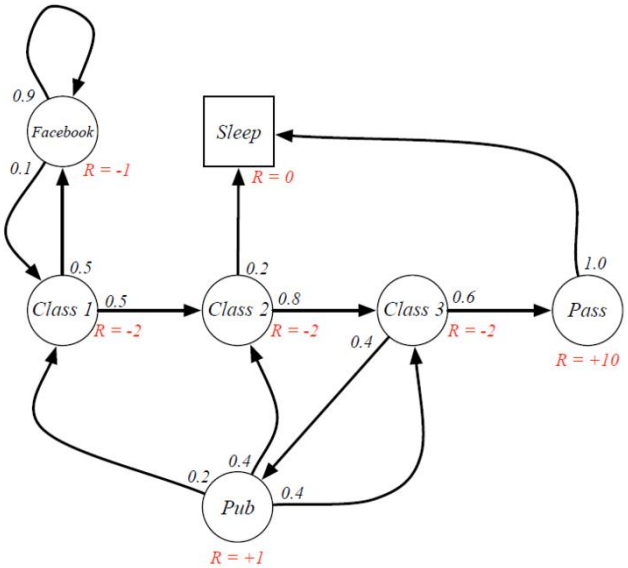
$$\mathcal{P} = \begin{array}{c} \\ C1 \\ C2 \\ C3 \\ Pass \\ Pub \\ FB \\ Sleep \end{array} \begin{array}{c} \begin{array}{ccccccc} C1 & C2 & C3 & Pass & Pub & FB & Sleep \end{array} \\ \begin{bmatrix} & 0.5 & & & & 0.5 & \\ & & 0.8 & & & & 0.2 \\ & & & 0.6 & 0.4 & & \\ & & & & & & 1.0 \\ 0.2 & 0.4 & 0.4 & & & & \\ 0.1 & & & & & 0.9 & \\ & & & & & & 1 \end{bmatrix} \end{array}$$

マルコフ決定過程について

- マルコフ報酬過程
  - マルコフ過程に報酬R　という概念を加えたもので、タプル⟨S,P,R,γ⟩で表現される。ここでRは報酬、γ∈[0,1]は割引率と呼ばれる。最終的に最も高いreturnを稼げる（最も褒められる）状態遷移シーケンスを見つけることが目的。将来に得られるだろう報酬に関しては、割引率が掛けられていることがポイント。未来の報酬というのは現時点では不確かなもので、得られない可能性がある。そのため確実に得られる直近の報酬はそのままで、将来的な報酬は先であればあるほど、その価値を低く（割り引いて）考える。

マルコフ決定過程について

- マルコフ報酬過程



報酬を考慮した時の状態遷移表

マルコフ決定過程について

- マルコフ決定過程

  - マルコフ報酬過程に対し、行動Aという概念を加えたもの。状態遷移という結果が勝手に起こる訳ないので、原因である行動を考慮することは大変合点がいく。どの行動をとるかについても、状態遷移と同様に確率で表現される。ただし行動は現在の状態を元に決定されるため、その行動に関する確率は状態を使った条件付き確率で表現される。この確率は方策と呼ばれます。

来週の内容

- 詳解確率ロボティクス10章の1部のマルコフ決定過程について

今週のナビゲーションの勉強会

- 内容
  - amclやslamについて話します