卒業論文要旨

階層型強化学習におけるサブゴールの自律的生成

(知能情報システム学)　名田　茂洋

1. 緒言

近年ロボット工学の分野の発展は著しく、ロボットが色々な分野において仕事をする必要性が大きくなってきている。そのために、様々な学習方法が確立され、使われるようになってきている。その中でより複雑な問題を解くためには、ロボットが自律的に学習することが、重要となってきている。さらに仕事を幾つかのサブタスクに分け、階層構造をもたらせることでより効率的に学習させる試みがなされている。本実験では、ロボットの学習方法の中でも階層型強化学習においてサブゴールを自律的に生成することを目指し、その効果を階層化を行わなかった場合と階層型強化学習においてサブゴールを見つける場合を比較しその効果を確かめるものである。

2. 強化学習について

強化学習とは、報酬を元にしてよりよい報酬が得られるように学習を進めていく学習方法である。その中で、今回主に使った Q 学習のアルゴリズムは簡単に示すと次のように示される。

① エージェントは環境の状態 s_t を観測する。
② エージェントは任意の行動選択方法(探査戦略)にしたがって行動 a を実行する。
③ 環境から報酬 r を受け取る。
④ 状態遷移後の状態 s_{t+1} を観測する。
⑤ $Q(s_t,a) \leftarrow Q(s_t,a) + \alpha [r + \gamma \max_a Q(s_{t+1},a)]$

ただし α は学習率$(0< \alpha \leqq 1)$, γ は割引率$(0 \leqq \gamma <1)$である。
⑥ 時間ステップ t を t+1 へ進めて①へ戻る。

また階層型強化学習とは上記の Q 学習などの学習を組み合わせ階層化したものである。本実験では、サブゴールができたときに、どのサブタスクを行うのかを選択する層とサブタスクを行う層に分かれている。

3. 実験環境と実験内容

開発環境は、コンピュータは Dell OPTIPLEX GX260 で CPU は Pentium4 の 2.53GHz を用い、OS は Microsoft 社の Windows2000 を用いた。開発言語としては、Microsoft 社の Visual C++6.0 を用いた。これらを用い、タクシー問題と部屋問題のシュミレータを作成し、シュミレーションによって学習を進めた。また、実験では、サブゴールを作って学習をするものと、作らずに報酬だけで学習するものの 2 つの学習方法についてデータを取った。

①タクシー問題

タクシー問題とはタクシーがスタート位置から出発し客を乗せに行き、客を乗せた後客の目的地に客を連れて行き、客を降ろすという問題である。この問題は階層型強化学習の効果を確かめる上で、学習させるのが難しい問題である。

タクシー問題は図 1 のように 5×5 のグリッドワールド上で実験を行った。

図 1.　タクシー問題のシュミレータ

この問題ではタクシーが行うことができる行動は上下左右に動く、その場に留まる、客を乗せる、降ろすの 7 種類である。また、学習を行うために与えた報酬は、１回行動すると-1 の報酬を与え、タクシーが客がいないところで客を乗せようとすると-10 の報酬を与えた。また、目的地でないところで客を降ろそうとすると-10 の報酬を与え、目的地で客を降ろしたときは+20 の報酬を与えた。以上の条件でタクシーに客を乗せ、目的地で客を降ろすことを学習させた。

②部屋問題

部屋問題とは壁によって隔てられた二つの部屋が一つ出入り口によってつながっていて、一方の部屋からスタートしたロボットがもう一方の部屋にあるゴールにたどりつくことを学習する問題である。この問題は階層型強化学習の効果を確かめる上で、タクシー問題と比べて学習が比較的簡単な問題である。

部屋問題は下図のように 10×6 のグリッドワールド上で実験を行った。

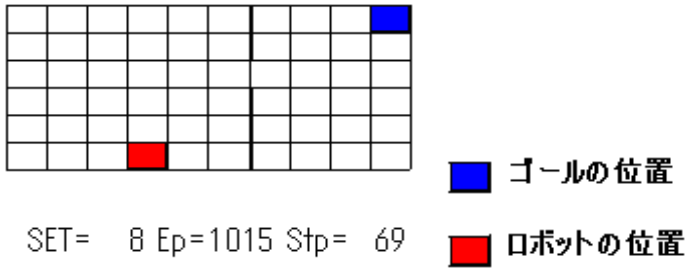

図 2.　部屋問題のシュミレータ

ロボットのスタート位置は x 座標は 0 に固定し、y 座標をランダムに動かした。ゴールの位置は(9,0)の座標に固定した。

また、この問題でロボットができる行動は上下左右とその場に留まるの 5 種類である。また学習を行うための報酬は、1 回の行動に付き-1 の報酬を与え、ゴールに着いた時は+20 の報酬を与えた。以上の条件で壁を通り抜けてゴールすることを学習した。

4.　結果

結果については、階層型強化学習でサブゴールを自律的に生成した場合と階層型強化学習を行わなかった場合との比較のグラフを卒業論文発表会の時に示す。