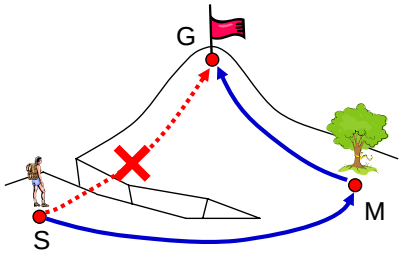
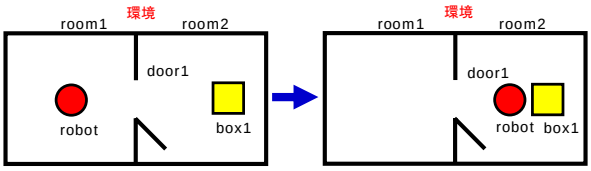
強化学習における中間目標の自動設定

- 強化学習[1]は，自律エージェントが環境から受ける時間遅れの報酬に基づき，自分自身のルールを強化することで行動最適化を行う機械学習の一種であり，近年特に注目されている．強化学習は扱う問題のクラス（マルコフ／部分観測マルコフ決定過程）と，アプローチの方向（環境同定型／経験強化型）によって特徴付けられる．
- 一方，強化学習に限らず機械学習全般において重要な課題として中間目標の設定がある．例えば，右のような登山の場合，出発点Sから目標地点Gを直接目指せないため，迂回ルートとしてMを当面の目標として目指す必要がある．このようなことは人間は普段から容易に行っているが，エージェントにとって自律的に決定することは非常に難しい．



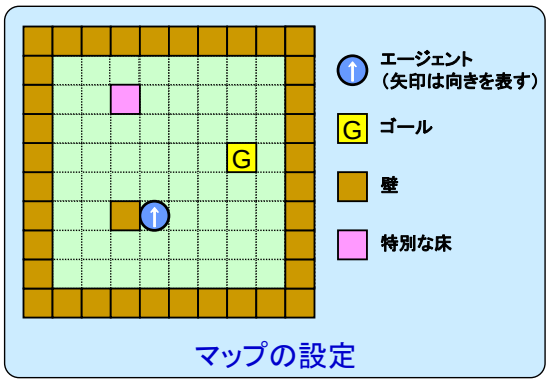
- 中間目標の自動設定の困難さは，当然，扱う問題の複雑さに依存する．例えば古典的な人工知能のプラニングでは，“ロボットが隣室の箱に近づく問題（下）”においてロボットが環境を完全に知覚できるなら，ドアに向かうという中間目標を獲得できるが，より複雑で抽象的な問題や，ロボットの知覚が部分的な場合などでは，一般に中間目標の自律設定は非常に難しいのが現状である．



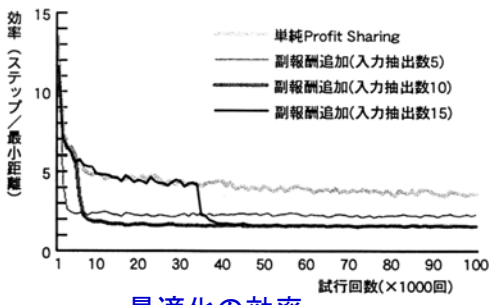
1) 詳細については，例えば，長尾智晴著「最適化アルゴリズム」昭晃堂 (2000) 他を参照のこと．

中間目標の自律設定の実験例

- ここでは長尾研@東工大の卒論として行われた，強化学習における中間目標の自動設定の実験例（大石・長尾 '99）を紹介する．
- 問題設定は次の通りである．
  - 右のようなマップ上でエージェントがゴールを目指す．周囲の床以外は試行毎に無作為に配置される．
  - エージェントはゴールに至ると正の報酬，障害物にぶつかると負の報酬を得る．特別な床はノイズとして作用する．
  - エージェントへの環境入力は{前，右，左，前方，右方，左方}で，前3つは1マスまで，後3つは遠方まで知覚できる．
  - Profit Sharing に基づくルールの強化を行いつつ，充分強化されたルールに関係した環境入力を複数個抽出し，それらに関与したルールに対して副報酬を与えた．
- 実験結果を右に示す．
  - 従来の強化学習手法である単純Profit Sharingに比較して，提案手法は試行回数が削減されるだけでなく，学習性能も2倍程度高いことを確認することができた．
  - 抽出数が性能に影響を与えるため，問題毎に学習中に自動設定できるようにした方が良いことがわかった．
- 今後，さらに複雑な問題に強化学習を適用する際の中間目標の自動設定について検討する．



マップの設定



最適化の効率