ファジィ環境評価型強化学習 (FEERL) を用いた習得済ルールの類似環境への有効利用

Effective Reuse of Acquired Rules by Fuzzy Environment Evaluation Reinforcement Learning (FEERL)

立命館大学理工学部　○ 星野 孝総，亀井 且有
Yukinobu HOSHINO and Katsuari KAMEI
Computer Science Ritsumeikan University

Abstract Reinforcement learning is a powerful machine learning system, which is able to learn without giving training examples to a learning unit. But it is impossible for the reinforcement learning to support large environments because the number of *if-then* rules is a huge combination of a relationship between one environment and one action. We propose new reinforcement learning system ,FEERL ,for the large environment. In this paper, we have tried to use acquired rules on maze problem.

1 はじめに

熟練者の教師データがない場合，エージェントは試行錯誤によって学習を進めなければならない．未知環境に対する学習手法として，強化学習が提案されている．代表的な手法である Q-learning[4] を用いた場合，習得したルールを類似状態や異種類の行動有する環境に利用できない．さまざまな環境に対する場合，ルールを有効利用する手法を強化学習に組み込む必要がある．筆者らは，ファジィ環境評価型強化学習手法 (FEERL) を文献 [2][3] で提案し，チェスゲームを用いて，複雑で巨大な環境に対する有効性を示した．本論文では，迷路探索で習得したルールを，行動形態変更し類似迷路を有した迷路探索問題に有効利用することを提案する．

2 ファジィ環境評価型強化学習 (FEERL)

ファジィ環境評価型強化学習は，ファジィ推論を用いた強化学習 [1] の 1 手法である．この手法では，状態を評価するためのルールベースを過去の経験としてシステム内部に持ち，未知状態に対しファジィ推論によって状態評価を行う．この評価値と先読みアルゴリズムにより未来の状態評価を行いながら学習・探索を行うことができる．したがって，過去に経験した状態から未知状態に対する行動を決定でき，さらにリアルタイムで学習させることが可能な機械学習アルゴリズムである．

3 部分観測迷路問題への適用

図 1 に示すような環境を用い，環境評価ルール [2] を用いた強化学習を迷路探索問題に適用した．エージェントは，5×5 の視界を持っている．エージェントが観測した状態に対し，その評価値を出力するルールベースを用いる．実際に経験した状態，行動，遷移先状態の組み合わせを記憶させ，環境シミュレータを作成した．また，先読み深度は，3 ステップとした．エージェントは，スタート (S) を出発し，ゴール (G) に到達したときに報酬を与え，1 試行が終了する．25 ステップ毎に罰を与え，学習はステップ毎に強化した．

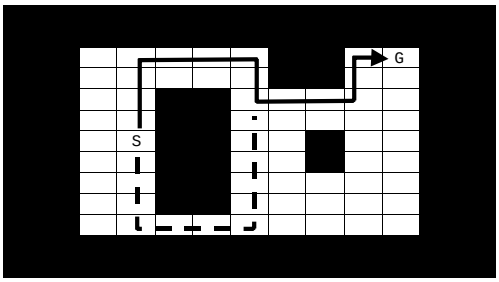

図 1: 習得した最短経路

3.1 迷路探索実験

実験は，66 試行おこなった，結果を図 2 に示す．グラフは，試行回数とゴールまでに要したステップ数を示している．試行回数に対して，ステップ数が減少している事が解る．また，図 1 に獲得した行動を実線矢印で示す．

4 FEERL による巨大環境への習得済ルール有効利用

迷路探索実験で取得したルールを，類似状態や異種類の行動を有する迷路問題に適用する．環境は，迷路探

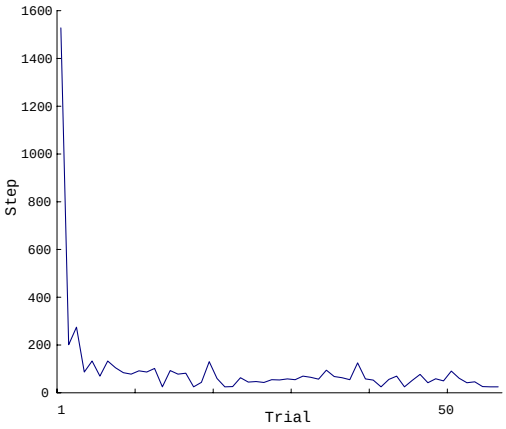

図 2: ゴール到達までのステップ数

索実験で学習した迷路の大きさを縦・横に 10 倍とし，図 3 のように複雑な形状にした迷路を用意する．迷路は 52×52 であり，エージェントの視界は 20×20 とする．先読み深度は，3 ステップとした．エージェントは，スタート (S) を出発し，ゴール (G) に到達したときに報酬を与え，1 試行が終了する．40 ステップ毎に罰を与えた．エージェントは，迷路内を (1) 式に示すような，運動方程式にしたがって移動するとし，行動は速さ v とハンドル角度 θ の操作量とする．ここで，x, y は縦・横位置，$\dot{x}$,$\dot{y}$ は縦・横移動量，ϕ は車体の角度，$\dot{\phi}$ は車体の角速度，L は車体のホイールベースである．迷路探索実験で習得したルールを縦・横に 4 倍に変更し，ルールベースの初期とする．

$$\begin{cases} L = 0.5 \\ \phi = \phi + \dot{\phi} \\ x = x + \dot{x} \\ y = y + \dot{y} \\ \dot{x} = cos(\phi) \times v \\ \dot{y} = sin(\phi) \times v \\ \dot{\phi} = tan(\phi) \times v \div L \end{cases} \tag{1}$$

エージェントは，視界情報のみを入力とし，速さ $v = (1.0, 0.8, 0.6, 0.2, -0.2)[m/s]$ とハンドルの操舵角度 $\phi = (-60, -30, 0, +30, +60)[deg]$ を出力とする．最初の格子空間で習得した環境評価ルールを初期値に用いる．エージェントの視界は 16 倍になっており，状態の組合せが大きくなる．

5 おわりに

本論文では，簡単な部分観測迷路探索問題を環境評価型強化学習に解かせ，習得したルールを 16 倍大きい類

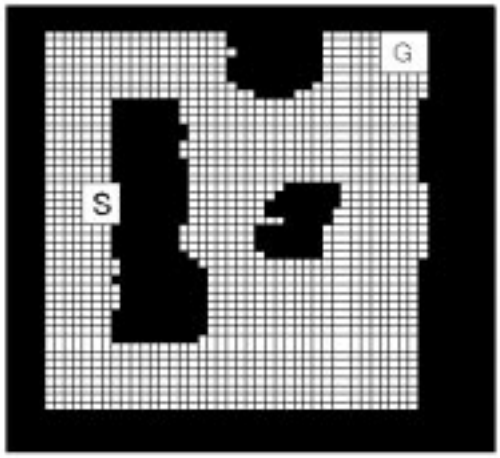

図 3: 実験に用いる巨大な類似環境

似環境に適用する事を提案した．今後は，提案手法を用いたルールの有効利用を確認する予定である．具体的には，先読み深度やファジィ推論のパラメータの調整を行う．さらに先読みアルゴリズムの高速化やモデル構築型の検討を行う予定である．

参考文献

[1] 堀内，藤野，片井，椹木: 連続入出力を扱うファジィ内挿型 Q-learning の提案; 計測自動制御学会論文集, Vol.35，No.2，pp.271–279 (1999)

[2] 星野，亀井: ファジィ環境評価ルールを用いた強化学習の提案とチェスへの応用; 日本ファジィ学会誌, Vol.13，No.6，pp.626–632 (2001)

[3] 星野，亀井: ファジィ環境評価型強化学習の LightsOut ゲームへの応用と探索における迂回行動の回避システム制御情報学会 論文誌，Vol.14，No.8，pp.395–401 (2001)

[4] Watkins.C.J.C.H: Learning from delayed rewards; Doctoral thesis, Cambridge University, Cambridge, England (1989)