ISS-P-133

カメラ画像入力に基づく高次元状態空間での強化学習

梶原　名月† 　上野　敦志† 　田窪　朋仁†
† 大阪市立大学大学院工学研究科電子情報系専攻

1. はじめに

強化学習では，画像入力などの高次元入力を扱う問題が長年課題となっている．過去の研究では人の手により特徴点が定義された手法や，ゲーム画面を入力とした手法は提案されているが，カメラ画像による人の手で特徴が定義されていない強化学習手法の成功例は少ない．よってカメラ画像入力に基づく人手による特徴のない高次元状態空間での強化学習手法を提案する．

2. Bag of Visual Words

画像の次元圧縮手法として Bag of Visual Words[1]を使用する．訓練画像の集合から局所特徴量を抽出し，教師なしクラスタリングにより visual vocabulary を生成する．クエリ画像を低次元のヒストグラムで表現するために，クエリ画像から得られた複数の局所特徴量をそれぞれ visual vocabulary の最も近いクラスタに分類する．分類された特徴量の個数がヒストグラムの各次元の値となり，特徴量の分類後，ヒストグラムは和が 1 となるように正規化される．

3. 経験強化型強化学習

Bag of Visual Words の高次元連続空間に対応するため，連続値入力に対応した経験強化型強化学習[2]に対し画像入力が扱えるよう改良を行う．[2]は連続値空間に対しガウス型関数と状態の価値を利用することで，連続空間の離散化を適応的に行う手法である．離散状態の認識には，入力値と離散状態の中心値とのユークリッド距離 $(u_i - d_i)^2$，及び離散状態の価値 $V(s_t)$ から算出される p 値を用いる．

$$f(\mathbf{d}) = \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(u_i - d_i)^2}{\sigma_i}\right)$$

$$p = f(\mathbf{d_t})\cdot V(s_t)$$

また，式(1)における分散 σ_i の値は離散状態生成時の入力状態 $\mathbf{i_t}$ と次の入力状態 $\mathbf{i_{t+1}}$ により決定される．

$$3\sigma_i = \begin{cases} |\mathbf{i_{t+1}} - \mathbf{i_t}| & (i=1) \\ \frac{1}{\sqrt{n}}|\mathbf{i_{t+1}} - \mathbf{i_t}| & (i\neq 1) \end{cases} \quad (1)$$

直交軸方向に $\sqrt{n}$ を除することで，主軸方向，つまり遷移方向への認識能力を向上させ，楕円型による適応的な分割を行っている．

初期エピソードでランダム政策により行動を行い，エピソード終了後，エピソード系列に従った連続値空間の離散化が行われ，次回以降のエピソードで使用される政策を生成する．

4. 提案手法

Bag of Visual Words は visual vocabulary を生成するための訓練画像として，[2]における初期エピソードでのランダム政策時に獲得した系列の画像を利用する．

式(1)において直交軸方向への補正は高次元では認識能力の低下が著しいため，$\sqrt{n}$ による除算を行わず，また分散は入力の差分ではなく実験前にあらかじめ固定させる．

5. 実験

シミュレーション環境で単眼カメラによるナビゲーションタスクを行い，ゴールまでの平均行動回数の収束値を調べた．行動は前後 0.1m，左右 45° 回転の4行動をとる．環境の大きさは 1.4m×1.4m で，壁付近を進入禁止領域とする．カメラの解像度は 512×512 で，画角は 90° である．また，四方と中央の壁にはすべて異なった画像が貼られている．

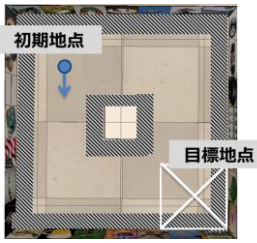

図1. シミュレーション環境

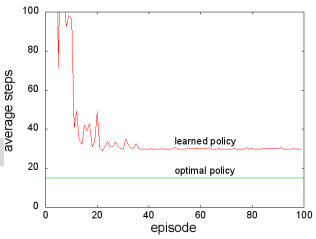

図 2.実験結果

経験強化型学習は経験に強く依存するため局所解に陥る傾向が強い．よって最適政策の獲得はできなかったが，20 エピソードまでに急速に収束し，40 エピソード以降はほぼランダムな行動をすることなくゴールまでの政策を獲得することができた．

6. まとめ

提案手法により，画像入力に対し事前にタスクに依った定義をすることなく政策を安定することが確認できた．今後の課題として，実環境での実験を行うこと，及び，他画像入力タスクにおいても有効かどうかを確めることがあげられる．

参考文献

[1] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, C. Bray, "Visual categorization with bags of keypoints," Workshop on statistical learning in computer vision, ECCV, vol.1, no.1-22, 2004.

[2] 藤井菜摘子，上野敦志，田窪朋仁，辰巳昭治，"連続値入力問題のためのガウス型状態表現を用いた TD 学習法，" 人工知能学会論文誌 29.1, pp.157-167, 2014.