計測自動制御学会東北支部 第202回研究集会 (2002.7.2)
資料番号 202-5

状況認識に基づいた移動ロボットのナビゲーション

Mobile Robot Navigation Based on Situation Recognition

○釜谷博行*，阿部健一**

○Hiroyuki Kamaya*, Kenichi Abe**

*八戸工業高等専門学校, **東北大学大学院工学研究科

*Dept. Electrical Eng., Hachinohe National College of Technology,
*Dept. Electrical and Communication Eng., Tohoku University

キーワード： 移動ロボットナビゲーション (Mobile Robot Navigation), 強化学習 (Reinforcement Learning), 状況認識 (Situation Recognition), Sarsa(λ), RCE-Classifier

連絡先： 〒039-1192 八戸市田面木字上野平16-1 八戸工業高等専門学校 電気工学科
釜谷博行，Tel.: (0178)27-7283, Fax.: (0178)27-9379, E-mail: kamaya-e@hachinohe-ct.ac.jp

1. はじめに

建物内の未知環境中でロボットをある目的地（ゴール）へ移動させるナビゲーション問題において，ある任意の場所を原点としたグローバルな座標系を用いてゴールの位置を表現した場合，ロボットの正確な位置情報(x, y, θ)が必要となる。これには通常，車輪エンコーダが用いられ，この計測値からロボットの位置情報(x, y, θ)が逐一更新される。しかし，ロボットの初期位置を正確に設定したとしても，車輪のすべりや床の凸凹などの影響により，ロボットの移動とともに累積する位置誤差の問題は避けられない。

この問題に対処するため，ロボットの外界を計測する外界センサを用いて，カルマンフィルタによりロボットの位置情報を補正する方法[1)]や確率的な手法によりロボットの位置を推定する方法[2)]などが提案されている。しかし，これには事前に地図情報を用意しなければならず，しかも効率の良い地図の表現方法が要求される。同時に，ロボット移動系を正確にモデリングしなければならない。さらに後者の方法では，地図情報と外界センサ情報とのマッチングに膨大な計算が必要となる。

本研究では，移動ロボットの位置情報(x, y, θ)を用いず，外界センサ情報のみを利用したロボットナビゲーション手法について考える。ロボットの位置情報を利用しないため，グローバルな座標系に依存した環境地図を構築できない。このため，外界センサ情報の系列からゴールへ向かう行動を学習により獲得しなければならない。また，ゴール座標が未知であるため，実行した行動が良好であるかどうかをすぐには評価できない。このため，ロボットがゴールへ到達したときにはじめて与えられる評価値（報酬）に基づいて障害物回避行動とゴール探索行動を同時に学習しなければならない。以上の理由から，本研究では強化学習[3)4)]を用いて学習器を構成する。

高次元の外界センサ情報を直接強化学習器の状態として利用することも考えられるが，学習にかなりの時間を要するなどの問題がある。そこで，今回はセンサ情報の前処理としてRCE-Classifierによりロボット周囲の状況を分類し，その分類結果を強化学習器への状態入力とする。シミュレーション実験により提案する学習器を評価する。なお，実験はすべて移動ノイズ，センサノイズのある環境で行った。

2. 学習アルゴリズム

ノイズが含まれる高次元のセンサデータを直接強化学習の状態として利用すると状態数が膨大となるため，学習が極めて困難となる。そこで，「障害物が前方にある」，「障害物が両側にある」などのロボット周囲の状況をセンサ情報に基づいて自己組織的にうまく分類することで，状態数の削減を図る方法が考えられる。これには，Kohonenの自己組織化アルゴリズム(SOM)[5] が広く知られており，さまざまな分野で利用されている。

しかし，ロボットナビゲーションにSOMを用いる場合，事前に行動環境のセンサデータを取得し，状況の分類学習を行っておく必要がある。障害物回避行動とゴール探索行動と同時に状況の分類学習を行うため，本研究ではRCE-Classifier (Restricted Coulomb Energy Classifier)[6] を用いる。提案する学習器のシステム構成をFig. 1に示す。RCE-Classifierにおいてロボットの置かれた状況を分類し，分類結果をコード化することによって，強化学習器への状態入力として用いる。

2.1 RCE-Classifier

RCE-Classifierの各クラスはR-ベクトルと呼ばれる表現ベクトルで表される。RCE-Classifierに入力されたセンサベクトルは，すでに存在するR-ベクトルと比較される。入力パターンとR-ベクトル間

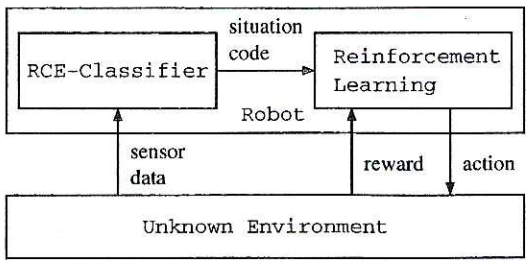

Fig. 1 学習器のシステム構成

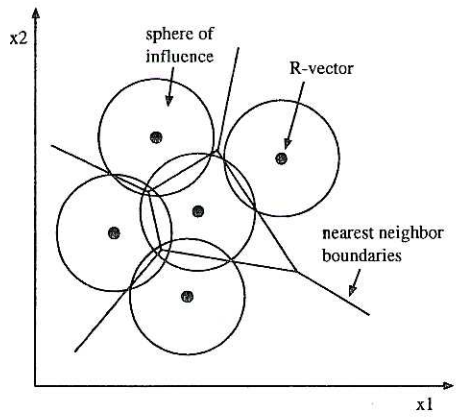

Fig. 2 RCE-Classifierの概念図

の類似度を計算し，その値があらかじめ設定されたしきい値内にあれば，入力パターンはそのR-ベクトルのクラスに属する。入力パターンが複数のクラスに属する場合には，nearest neighborルールにより一つのクラスを決定する。類似度がしきい値を越えた場合には，入力パターンは新たなR-ベクトルとなり，クラスが追加される。

Fig. 2は2次元の入力ベクトルをもつRCE-Classifierの概念図を示す。図中のそれぞれの黒丸がR-ベクトルを表す。R-ベクトルを中心とした円はクラスの境界を示す。

今回の実験では，RCE-Classifierへの入力ベクトルとして，ロボットに取り付けられた16個の超音波センサの計測値を用いる。類似度はユークリッド距離を用いて計算され，入力パターンが複数のクラスに属する場合には最も距離の小さなクラスに属するものとした。ロボットが環境中を移動するにつれて，16次元のセンサデータが「教師なし学習」により自動的に分類される。

2.2 強化学習

各時点$t \in \{0,1,2,\ldots\}$において，環境状態が$s_t \in S$のとき，その状態観測に基づき，エージェントが行動$a_t \in A$をとったとすると，報酬r_tを受け取り，環境状態は未知の遷移確率でつぎの状態s_{t+1}に遷移する。エージェントの目標は，報酬の割引期待利得$E\{\sum_{t=0}^{\infty}\gamma^t r_t\}$を最大にすることである。ここで，$\gamma(0 \leq \gamma \leq 1)$は割引率を表す。

強化学習では，ある状態 s において行動 a を選択するときの評価値をQ値と呼び，$Q(s,a)$で表わす。Q値の大きさ応じて，エージェントは状態sにおいて実行すべき行動aを決定する。環境と対峙したエージェントは試行錯誤を繰り返しながら，割引期待利得の最大化を目的として，各時点で得られる報酬r_tに基づいてQ値を更新していく。強化学習アルゴリズムとしてTD法[7]，Q-学習[8]，Sarsa[3]などが広く知られている。本研究では，Sarsa(λ)を用いる。

2.3 Sarsa(λ)

Sarsa(λ)では，遷移先の状態s_{t+1}の評価値として，s_{t+1}において選択した行動a_{t+1}に対応するQ値を用いる。$< s_t, a_t, r_t, s_{t+1}, a_{t+1} >$の5項組を用いて，Q値はつぎのように更新される。

$$\delta_t \leftarrow r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \quad (1)$$

$$Q(s,a) \leftarrow Q(s,a) + \alpha\delta_t e(s,a) \quad \text{for all } s \text{ and } a \quad (2)$$

ここで，αは学習率$(0 < \alpha \leq 1)$を，γは割引率を表わす。また，$e(s,a)$は*eligibility trace*と呼ばれ，

$$e(s,a) \leftarrow \begin{cases} 1 & \text{if } s = s_t \ \&\ a = a_t \\ 0 & \text{if } s = s_t \ \&\ a \neq a_t \\ \gamma\lambda e(s,a) & \text{if } s \neq s_t \end{cases} \quad \text{for all } s \text{ and } a \quad (3)$$

ここで，λは$0 \leq \lambda \leq 1$なる実数である。なお，Q値の更新に先立って*eligibility*の計算が行なわれる。

2.4 行動選択（Max-Boltzmann法）

さまざまな行動選択法が提案されているが，本研究ではMax-Boltzmann法を用いる[9]。これは，確率 p_{max} で最大のQ値をもつ*greedy*な行動を，確率 $1 - p_{max}$ でBoltzmann分布にしたがって行動を選択するものである。Boltzmann分布では，状態s_tにおいて行動a_iを選択する確率は次式で与えられる。

$$\Pr(a_i|s_t) = \frac{e^{\frac{Q(s_t,a_i)}{\tau}}}{\sum_k e^{\frac{Q(s_t,a_k)}{\tau}}} \quad (4)$$

ここで，τは行動選択のランダムさの度合いを決定するパラメータで温度係数と呼ばれる。なお，学習終了時に決定的な方策を得るために，p_{max}をある初期値から1まで徐々に増加させる。

3. シミュレーション実験

シミュレーション実験では，著者らが開発した移動ロボットシミュレータを用いる[10]。ここでは，直径46[cm]の円筒形の車輪型移動ロボットを考える。ロボットには，16個の超音波センサがリング上に配置される。このセンサにより22.5[deg]毎にロボット周囲の未知障害物までの距離を計測できる。今回の実験では，実機の特性にできるだけ近づけるため，超音波センサのモデリングにおいて，センサ主軸と壁面との角度が25[deg]を越えると壁を検出できないものとした。このため，Fig. 3に示すように近くに障害物があってもそれを検出できない場合がある。また，センサ計測値には10%のランダムなノイズが含まれるものとした。

移動ロボットのナビゲーションで利用する2つの実験環境をFig. 4とFig. 5に示す。大きさはそれぞれ5[m]×5[m]，13.5[m]×5.8[m]で，通路幅は1.8[m]である。黒い部分が壁や障害物を表わす。

ロボットの目的は，障害物回避行動と同時にスタート地点Sからゴール領域Gへ向かう行動を超音波センサ情報のみを用いて獲得することにある。

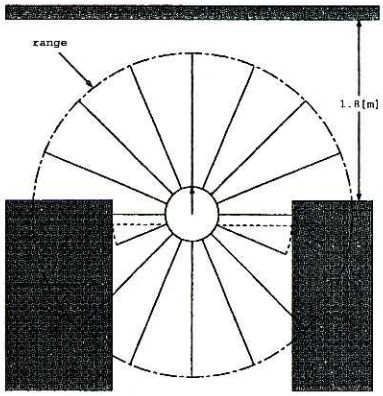

Fig. 3　超音波センサの計測データ

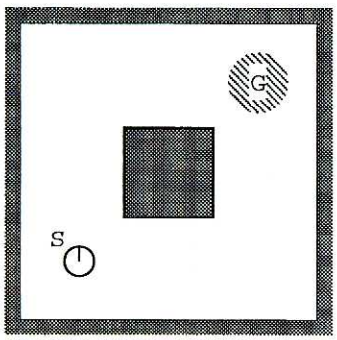

Fig. 4　環境1

なお，ロボットはその位置情報(x, y, θ)を認識できないものとした。

計測範囲を12～140[cm]に制限した16次元の超音波センサデータがRCE-Classifierに入力され，状況に分類される。RCE-Classifierから出力された状況コードが強化学習器の状態入力となる。強化学習器では，あらかじめ設定された行動の中からひとつの行動を選択・実行する。今回は0[deg]，±12[deg]，±45[deg]回転後に10[cm]直進するという5つの行動を設定した。このとき，移動量，回転量に対してそれぞれ10%のランダムなノイズが与えられる。回転角度が0[deg]の場合には±1[deg]のランダムなノイズとした。Fig. 6にロボットを10[m]直進させたときの位置(x, y, θ)の変化の様子を示す。ノイズの影響により，y方向では約16$[cm]$，角度方向では約3$[deg]$の誤差が生じている。

報酬はつぎのように定めた。

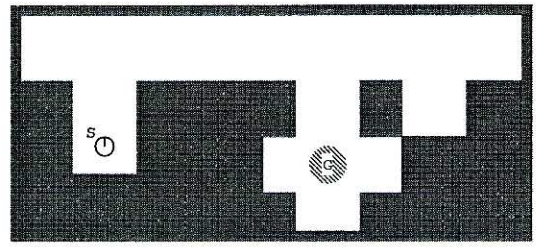

Fig. 5　環境2

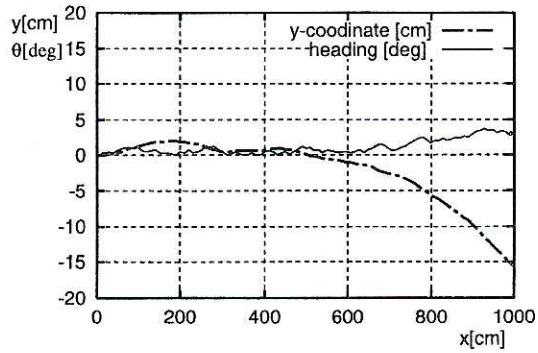

Fig. 6　移動ノイズの影響（10[m]直進時のロボットの位置）

状態	報酬
ゴール到達時	100
障害物への近接時	−1
同一地点での停留	−1
上記以外	0

障害物との接触を避けるために，超音波センサの計測値が20[cm]以下の場合には，ロボットを停止させ，近くの場所から再スタートさせる。このとき負の報酬を与える。なお，障害物の角など，超音波センサで検出できない障害物はロボットの周囲に取り付けられた接触センサで検出し，ロボットを緊急停止させる。このときも，同様な処理を行う。また，直進以外の同じ行動を連続して選択した場合には，ロボットが同一地点に停留していると考え，負の報酬を与えるようにした。

4.　実験結果

強化学習システムのパラメータは，$\gamma = 0.9$，$\alpha = 0.1$，$\lambda = 0.9$とした。行動選択においてp_{max}の値は，初期値を0.9とし，最後の試行で1.0になるよ

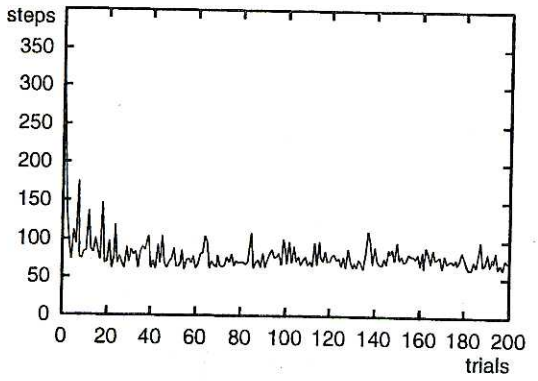

Fig. 7 ゴールまでのステップ数(環境1)

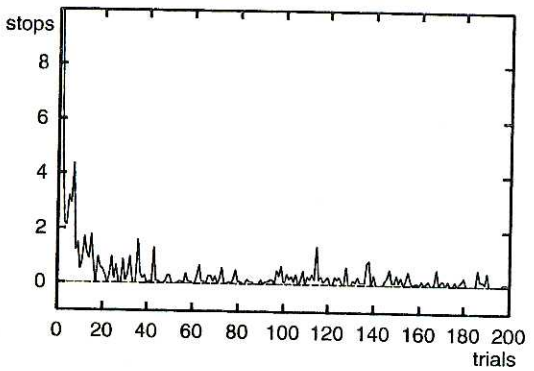

Fig. 8 障害物への最接近回数(環境1)

うに直線的に増加させる。最大ステップ数$T_{max} = 1{,}000$，試行回数200，Boltzmann分布の温度係数の初期値$\tau = 0.5$とし，試行回数の増加とともに徐々に減少させた。RCE-ClassifierのR-ベクトルの半径R_{max}を125$[cm]$に設定した。

1回の試行は，ロボットがゴールへ到達したときか，最大ステップ数の上限を越えたときのいずれかで終了する。つぎの試行は，ロボットをスタート地点に戻して行なわれる。ロボットが壁に接近しすぎた場合には，ロボットを壁から離れた場所に移動させて，試行を継続するようにした。

Fig. 7は環境1における試行回数に対するゴールまでのステップ数を表す。グラフは乱数の初期値を変えた10シミュレーションの平均値をプロットしたものである。結果をみると，試行回数の増加とともにゴールまでのステップ数が減少しており学習に成功していることがわかる。Fig. 8は試行回数に対する障害物への最接近回数を表す。試行とともに障害物への最接近回数が減少しており，障害物回避行動もうまく学習していることがわかる。

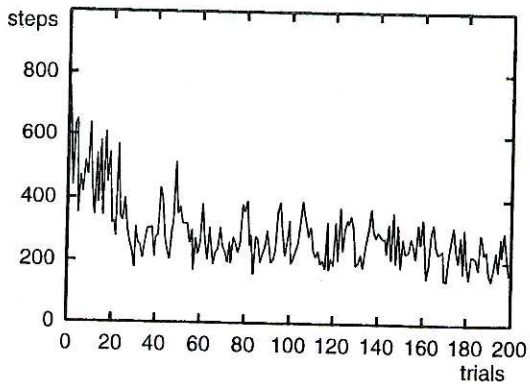

Fig. 9 ゴールまでのステップ数(環境2)

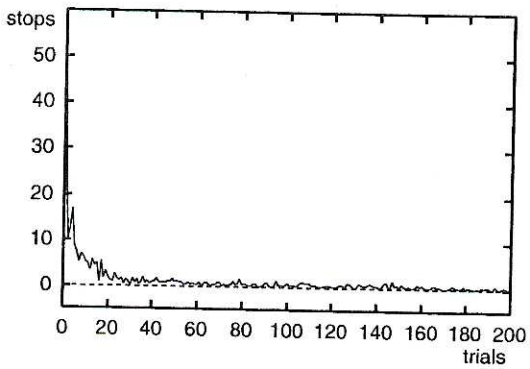

Fig. 10 障害物への最接近回数(環境2)

環境2における実験結果をFig. 9とFig. 10 に示す。環境2は環境1に比べてサイズが大きい。ゴール探索行動をうまく学習しているが，時折ノイズ等の影響で最大ステップ数内にゴールに到達できない場合もある。これが，Fig. 9のグラフの振動として表れている

なお，10回のシミュレーションにおいて学習後のRCE-Classifier（$R_{max} = 125[cm]$）のR-ベクトル数は，環境1では38～58，環境2では83～98の範囲であった。Fig. 11に分類されたセンサ情報の一例を示す。

5. おわりに

ロボットの位置情報を利用せず，外界センサ情報のみから障害物回避行動とゴール探索行動を同

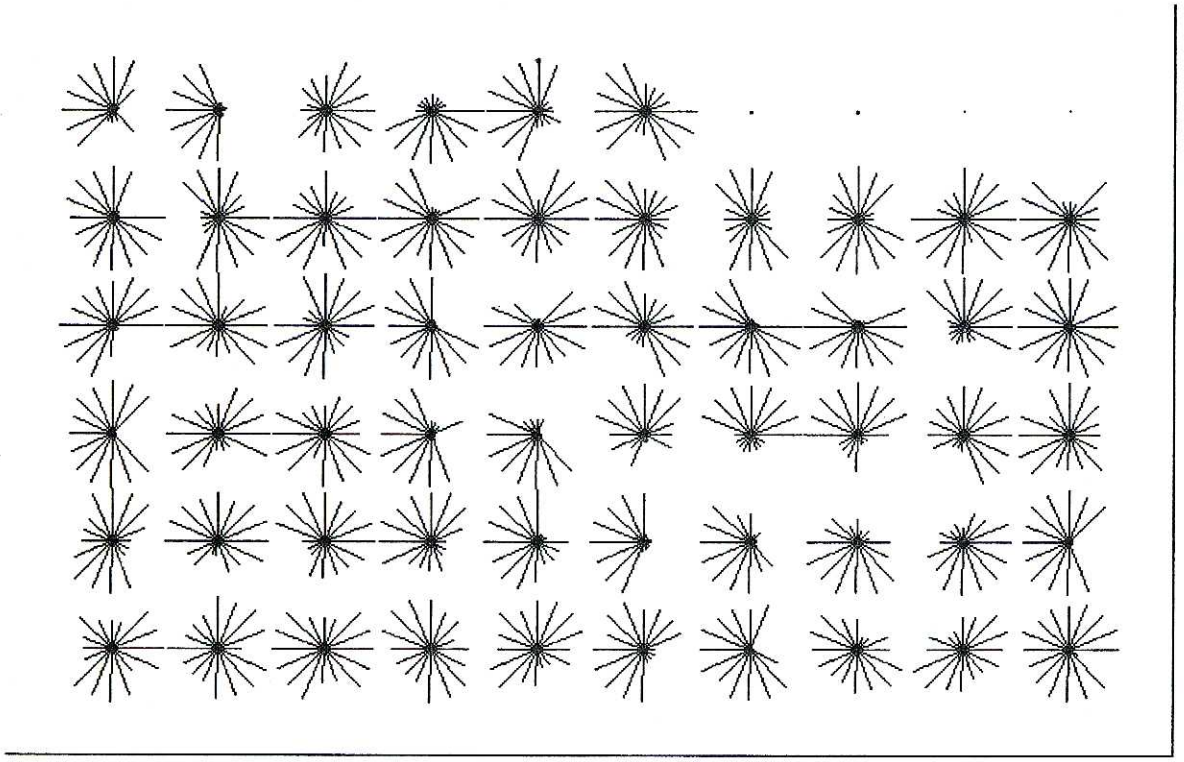

Fig. 11　分類されたセンサ情報の例

時に学習する学習器を提案した。単純な環境でのシミュレーション実験から良好な学習性能を得ることができた。

しかし，同じ状況に分類されるセンサパターンが環境中の至る場所に存在すると隠れ状態を生じ，これによってゴール探索行動の学習性能が著しく低下するという問題が生じると考えられる。このため，隠れ状態に対応できる強化学習アルゴリズムについて検討する必要がある。また，今回はシミュレーション実験の結果について報告した。今後は，実機を用いて提案する学習器について検討していきたい。

参考文献

1) D. Fox, W. Burgard, and S. Thrun: "Active markov localization for mobile robots," *Robots Autonomous Syst.*, **25**, no.3–4, pp.195–207, 1999.

2) P. Jensfelt and S. Kristensen: "Active Global Localization for a mobile robot using Multiple Hypothesis Tracking," *IEEE Trans. on Robot. Automat.*, **17**-5, pp.748–760, 2001.

3) R. S. Sutton and A. G. Barto: *Reinforcement Learning : An Introduction.* Cambridge, MA: MIT Press, 1998.

4) L. P. Kaelbling, M. L. Littman and A. W. Moore: "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, **4**, pp.237–285, 1996.

5) T. Kohonen: *Self-Organizing Maps 3rd ed.*, Springer-Verlag, 2001.

6) A. Kurz: "Constructing Maps for Mobile Robot Navigation Based on Ultrasonic Range Data," *IEEE Trans. on Systems, man, and cybernetics, Part B: Cybernetics*, **26**-2, pp.233–242, 1996.

7) R. S. Sutton: "Learning to predict by the methods of temporal differences," *Machine Learning*, **3**, pp.9–44, 1988.

8) C. J. C. H. Watkins and P. Dayan: "*Q*-learning," *Machine Learning*, **8**, pp.279–292, 1992.

9) M. Wiering and J. Schmidhuber: "*HQ*-learning," *Adaptive Behavior*, **6**-2, pp.219–246, 1997.

10) 釜谷博行,本間弘一,阿部健一: "オブジェクト指向設計に基づいた自律型移動ロボットの開発支援システム," 電気学会論文誌, **115**-C-6, pp.819–828, 1995.