第29回「知能システムシンポジウム」
（平成14年3月28，29日,学術総合センター）

知的自動車駐車システムの強化学習による運転知識獲得

筑波大学機能工学系　　安信誠二　ヤヤ・スルヤナ　末竹規哲

Reinforcment-learning Design of Intelligent Controller for Nonholonomic Vehicle

University of Tsukuba　　Seiji Yasunobu, Yaya Suryana and Noriaki Suetake

Abstract: In this paper, the profit sharing plan (PSP) is used to learn a knowlege of strategy target for achieving final target of nonholonomic vehicle moved in narrow parking area. The simulation result showed the effectiveness of the proposed method in generating control knowledge that could be time consuming and difficult to acquire by heuristic method.

1　はじめに

四輪自動車は，ハンドルと速度の 2 つの操作だけで，車体の位置と向きの 3 状態を制御する，ノンホロノミックな特性を持つ熟練と経験を要する運転である[1, 2]．特に，駐車場などの限られた領域への駐車を考えた場合，熟練した運転者は，駐車場の状況を把握し，適切な運転を行なえるが，非熟練者は，立ち往生することも多い．しかし，駐車場の誘導員の指示により，駐車目的の達成が可能である．

知的自動車駐車システムは，この熟練運転者や誘導員の車の特性を踏まえた運転知識を組み込み，人間に代わり予見ファジィ制御方式[3, 4]により滑らかに車を自動運転するすると共に，人間が運転する場合にも適切な支援情報を提供するシステムとして開発が進められている[5, 6]．しかし，車の駐車状況（空いてる場所には侵入可能など）により難易度の状況が異なる駐車知識の設定を，個別に人間が行うことは，手間を要し困難であった．

一方，最適解への勾配情報を用いずに，人間が試行錯誤的に行動を選択し，成功を学習している仕組みを取り入れている強化学習[7]の考え方が提案され，スケジューリングなどの問題に対して有効性が確認されている[8]．また，この強化学習の手法の一つとして，何段階かの行動の後の成功報酬を各段階にさかのぼって分配する PSP(Profit Sharing Plan)-学習の手法[9, 10, 11]が提案されている．

本論文では，ノンホロノミックな特性をもつ自動車の駐車制御を行う知的自動車駐車システム[5, 6]において，駐車場の形状や車体特性（最小旋回半径など）により異なり，設定に経験と時間を要する駐車知識の設定に，PSP-学習の手法を適用し，運転知識の獲得を試みる．

2　自動車駐車システムの問題

自動車は，Fig.1 に示すような幾何学的動きをし，左右前輪の角度 (ϕ_L,ϕ_R) は，四輪全てが旋回中心と直角に旋回半径 R で動くように，アッカーマン・ジャントウの操舵機構により構成されている．

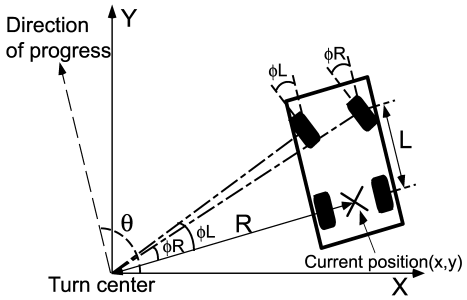

Fig.1 Kinematics constraint in nonholonomic vehicle

この自動車の動きは，以下の拘束条件式（1）によって記述できる[2]:

$$\begin{aligned}\frac{dx}{dt} &= v\cos\phi\cos\theta,\\ \frac{dy}{dt} &= v\cos\phi\sin\theta,\\ \frac{d\theta}{dt} &= \frac{v}{L}\sin\phi,\end{aligned} \tag{1}$$

ここで，x,y は自動車の位置，θ は x 軸と自動車の進行方向のなす角，v は車の速度，ϕ は旋回半径 R での動きを指令するハンドルの切れ角，L は車のホイールベース（長さ）である．

この自動車の駐車運転は，最終目標（所定の駐車スロット）に到達することが目的となる．しかし，上記の特性のため，その近くに居ても，車の向きや位置が異なると簡単には移動できず，単純に到達度を求めることができない．そのため，目標状態に到達するような制御指令を決定することが困難である．特に車の最小旋回半径が制御性能に直接関連するような，狭い領域での駐車は難しい．比較的広い駐車場を想定した場合は，追従すべき経路と走行パターンを余裕を持って設定可能であり，それに PID 制御などで追従することも可能であるが，この経路計画も難しい問題である．

これらを解決するため，到達可能な目標を与える部分と，その目標への自動車を操作する，階層型知的制御系を提案し駐車制御を実現している[5, 6]．

3　人間の駐車知識と PSP 学習

3.1　人間の駐車知識

ノンホロノミックな特性を持つ自動車をうまく駐車させている人間の運転方法を Fig.2 のような駐車場を例として考える．

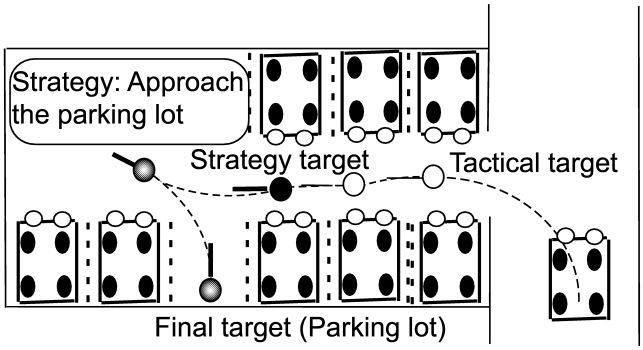

Fig.2 Human control strategy in parking vehicle

熟練者の駐車手順は，以下のように記述できる．

1. 駐車場の形状や，他のスペースの塞がり状況を把握し，駐車場と現在の車位置から,「車庫に寄せる」,「向きを変える」,「切り返し位置へ移動する」「車庫に入れる」「制御を終了する」など，大体実現可能な大局的な目標（戦略目標と呼ぶ）考える．

   さらに，そこに到達するために，ノンホロノミックな車両特性を考慮し，一回のハンドル操作で到達可能な，局所的な目標（戦術目標と呼ぶ）を決める．
2. 戦術目標に到達できるよう，壁や障害物を避けながら，ハンドルと速度を操作する．この操作は，基本的には，一定のハンドル切れ角，速度で実現できるものである．しかし，障害物や壁の状況などにより少しの調整や危険回避の行動をとる．
3. 目標に到達したか，到達が困難だと判断した時，現在の目標をリセットし考え直す．

この様な人間の運転方法に基づき，知的自動車駐車システムが構築されている．大局的な戦略目標が適切に設定されれば，局所的戦術目標は，車両特性に基づき比較的容易に設定できる．しかし，戦略目標の設定には，状況に応じた知識が必要であり，設定が難しい．本論文では，PSP-学習の手法をこの戦略知識の獲得に用いる．

3.2　*PSP*-学習

PSP-学習[9]は，離散的な状態遷移に対して適用されている場合が多いが，ここでは，連続値の状態の制御である車の運転に適用する．PSP-学習に関する基本概念を Fig.3 に示す．行動知識を，”IF <condition:c_n> THEN<action:a_n>” なる形式の状態と行動のペアからなる規則（S-table）で記述する．

ある初期値からの制御開始をエピソードと呼び，各ステップ n において，行動知識に基づいて，現在状態 c_n に最も適切な（または，ランダムな）行動 a_n を選択する．選択した行動を実行した結果の状態 c_{n+1} において，同様に次の行動 a_{n+1} を選択する．もし，最終目標に到達した時，報酬 R を γ で割引きながら，そのエピソードで使用した N 個の規則へ反映していく．途中で，実行不可能になった場合は，負の報酬（罰）を与える．

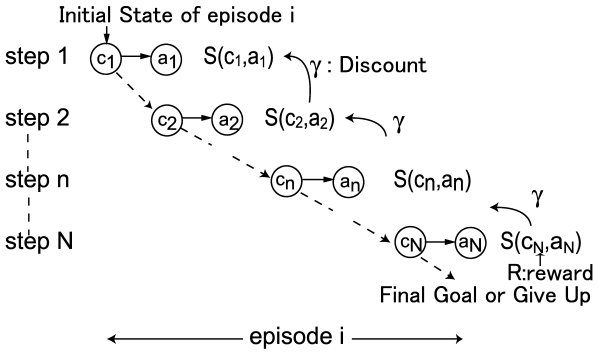

Fig.3 PSP-learning distributes reward or penalty to the previous fired rules

4　知的自動車駐車システムの概要

4.1　システムの概要

戦術目標への到達を判断する状況監視部，運転知識に基づき戦略目標と戦術目標を設定する目標設定部，戦術目標へ車を操縦する自動運転部からなる知的自動車駐車システム[5, 6]に，実行結果に基づき戦略目標の駐車知識を獲得する強化学習（PSP 学習）部を付加して今回構築したシステムの概要を Fig4 に示す．

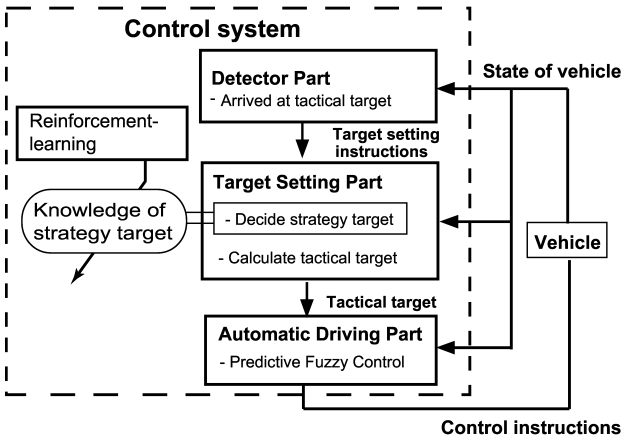

Fig.4 Outline of the hierarchical intelligent controller

4.2 状況監視部

車は，戦術目標へ到達するよう制御される．現在の戦術目標に到達した場合には，新たな戦術目標が必要である．ここで用いる戦術目標の決定では，車の特性は考慮しているが，壁などを考慮していないため，障害を回避し，現在の目標に到達できなかった場合には，現在の目標をあきらめ，新しい戦術目標を再度考えるように指示する．この状況監視部では Fig.5 に示すように，(a) 戦術目標と直角に引いた線より前方に出る，(b) 予定した距離を走行した，(c) 指定した方向へ（障害物に当たりそうで）進めない，場合に，現在の戦術目標をリセットする．また，最終目標への到達もここで検出する．

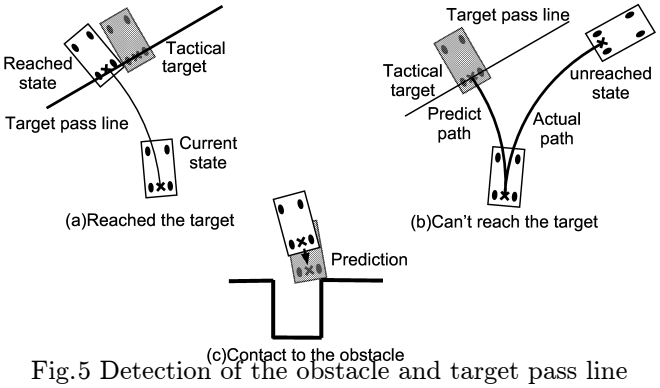

Fig.5 Detection of the obstacle and target pass line

4.3 目標設定部

目標設定部が状況監視部から目標を変えるよう指示を受けたとき，目標設定部では，車の現状が現在の戦略目標到達しているかを考え，未到達の場合は，新たに戦術目標を設定する．新しい戦略目標は，PSP-学習法を使用することで選択する．これまで[5, 6]は，以下のような駐車知識に基づき戦略目標（位置と方向）を記述している:

- 車庫に寄せる: 所定の駐車位置（車庫）に到達するために，現在位置に最も近くて簡単な目標を設定.
- 向きを変える: 車の向きを変えて車庫に入れやすくするため，一旦，本来進むべき方向とは異なる方向に目標を設定.
- 切り返し位置へ移動する: 次の目標設定で切り返し（逆の進行方向に移動）するため，その開始位置を目標に設定.
- 車庫に入れる: 車庫に入れるため，最終目標を目標に設定.
- 制御を終了する: 車が最終目標に到達し，車庫入れが完了.

4.3.1 *PSP* 学習による戦略目標設定

本論文で用いたシステムでは，PSP-学習で駐車知識 (S-table) を獲得し，その知識に基づき戦略目標を決める．

S-table は 2m 四方と 45 度刻みに仕切られたセルの中心値をラベルとする (c_n，a_n) の組み合わせである．獲得した運転知識の使用は，現在状態 c_n で最大の S-table 値を持つ目標 a_n が戦略目標として選択される．各初期位置から最終目標までの経路を，1 つのエピソードと呼ぶ.

駐車知識を獲得するためには，各状態において目標をルーレットによりランダムに選択することにより未知の経験を試みる．実行した結果の報酬は，エピソードの終わりに，選択された S-table 値に分配される．最終目標に達することができた場合，報酬は制限時間 (250 秒) と所要時間の差で与える．例えば，最終的な目標に 80 秒で達したならば，報酬 r_i は 170 である．制限時間内に到達できない時や，障害物などで動きが取れない状況に陥った場合，報酬（罰）r_i を-10 とする.

状態は三次元 x，y，および θ の仕切られた状態空間である．x，y に関しては，最終目標から 2m 毎の格子点を中心とし，各位置における車の向き θ は，0，0.25π，0.5π，0.75π，π，1.25π，1.5π，1.75π，の 8 方位に区分しラベル化している．例えば，5 章の Fig.8 で示す駐車場の状況では，78 個の格子点が置かれ，78×8=624 個のラベルがある．また，行動選択のための戦略目標 $a_n = (x_T, y_T, \theta_T)$ も 624 である．最終目標は (0m,0m,0.5π) である.

S-table 値を更新する PSP-学習のアルゴリズムを以下に示す..

1) 第 i エピソード (i=1,$Imax$) を， 2)-8) で実行.

2) 状態 c_1 にエピソード初期位置 (x_i,y_i,θ_i) をセット.

3) 第 n ステップ (n=1,$Nmax$) を, 4)-7) で実行.

4) もし，現在状態 c_n が最終目標のとき，報酬: R_i=$Tmax$-t_n, N=$n-1$ とし, エピソードを終了 8) へ.

5) もし，動けないまたは $Tmax$ を過ぎた時，報酬 (罰):R_i=-10, N=$n-1$ とし，エピソードを終了 8) へ.

6) 最大 S-table 値または，ルーレット選択により，戦略目標 a_n を選ぶ.

7) 予見ファジィ制御により，次の戦略目標 a_n へ移動を試みる．現在の戦略目標に到達した時，次のステップ実行のため 3）へ.

8) 次の式 (2) により，使用した N 個の S-table 値を更新した後，次のエピソードの実行のため 1) へ:

$$S(c_n, a_n) = (1-\alpha)\ S(c_n, a_n) + \alpha R_i \gamma^{N-n} \qquad (2)$$

ただし，R_i はエピソード i での報酬，$Tmax$ は最大時間，t_n は制御終了 (現在) 時間，α は学習率，γ は割引率，N は制御終了時のステップ数，$Nmax$ は最大ステップ数，$Imax$ は最大エピソード数である.

4.3.2 戦術目標設定

戦略目標は，車をどこに位置すれば上手く駐車できるかを指示するだけで，車のノンホロノミックな特性を積極的に意識していない．そのため，戦略目標に移動させるためには，車の特性に基づき Fig.6 のように最小旋回半径 R を用いて，戦術目標 (x_T,y_T,θ_T) を設定する．この戦術目標は，現在の状態からハンドル切れ角を一定にした場合に到達可能な地点と車体方向である．戦術目標の設定には，既報告[5, 6]の手法を用いている．

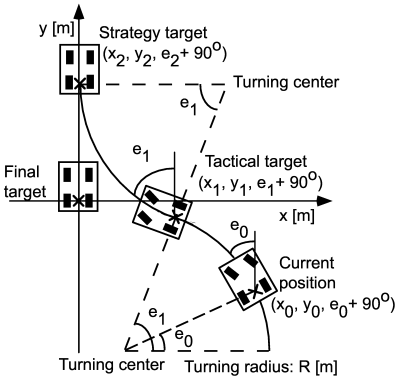

Fig.6 Tactical targets to attain strategy target

4.4 自動運転部

位置と車体方向で与えられる戦術目標 (x_T,y_T,θ_T) に達するように，車を運転するため，予見ファジィ制御[4][3]を用いている．予見ファジィ制御では，Fig.7 に示すように，ハンドルの操作角度の候補に対して，その動きを予見し，もっとも適切なハンドル操作角度をファジィ推論により選択する．このとき，他の車や壁までの接近の状況も評価し，状況に柔軟に対応した運転を行う．従って，外的要因が無い場合は目標に正確に到達できるが，状況が異なるとなるべく近くに（ファジィ評価）車を移動させる．この方式の詳細と有効性は，既に報告[5, 6]している．

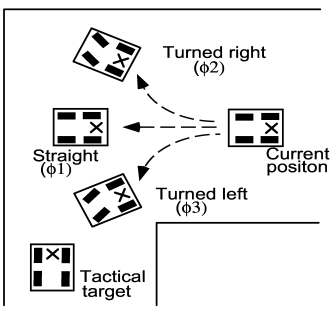

Fig. 7 Selection of control candidate in predictive fuzzy control

5 シミュレーション結果

シミュレーションに用いた自動車の特性を Table.1 に示す．また，駐車場の形状は，5m の道路から，6 台 ×2 列の駐車スペースを設定した（Fig.8）．

Table1 Characteristic of the vehicle

wheelbase	$2.6m$
distance between axis and bumper	$0.4m$
width	$1.7m$
smallest radius	$6m$
velocity	$0.4m/s$ (foward) $0.0m/s$ (stop) $-0.4m/s$ (backward)

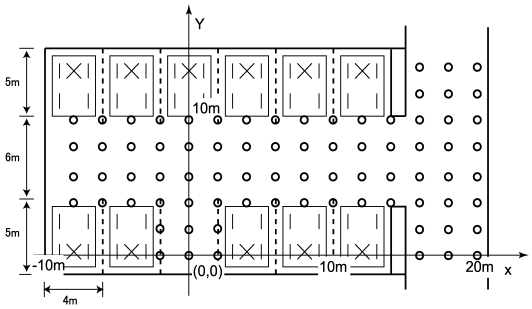

Fig.8 The dimension of parking lot

ここで，駐車知識を獲得するための初期状態は，Fig.8 に丸で示す，2m 刻みの 78 個の格子点と 45 度刻みの 8 方位であり，624 エピソード行う．この全エピソードについて，1000 回のトライアルを行い，駐車に成功したエピソード数の状況を Fig.9 に示す．各エピソードの初期状態は，車が障害物に接触している場合が多く，実際的に移動可能な殆どの場合の 160 のエピソードで成功している．このトライアルは，9 回のルーレット（ランダム）選択による冒険（Fig.9 の下部）と 1 回のグリーディ（最大値）選択（Fig.9 の上部）の繰り返しにより行った．また，学習率 α は 0.5，割引率 γ は 0.8 とした．所用時間は，Pentium4-2GHz で 1,116 分であった．ただし，今回は，自動運転部に人間と同様に柔軟な運転を行う予見ファジィ制御を用いたが，学習時に，より簡易な方法[11]を用いると 20 分程度でほぼ同様の知識を獲得でき，そこからさらに洗練することが可能である．Fig.10 に，獲得した運転知識（S-table : 状態 (c_n に対する最大値を持つ目標 a_n をプロット）を示す．

Fig.11 に，ここで獲得した運転知識を用いて，駐車場の中ほどの位置に 1 つの空きがある場合を想定し，獲得した運転知識を用いた走行（グリーディ選択）の様子を示す．初期状態は，駐車場の外の (18m,0m,0.5π) である．この状態から最初の戦略目標として駐車場の入り口の (16m,4m,0.75π) を設定し，その戦略目標に達するように，移動している．

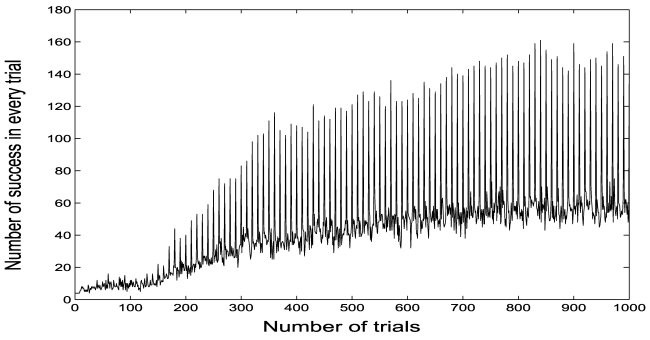

Fig.9 The performance of PSP-learning

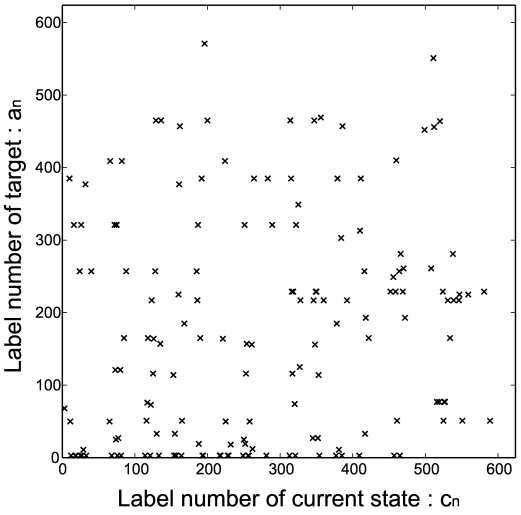

Fig.10 Obtained drive knowledge (S-Table)

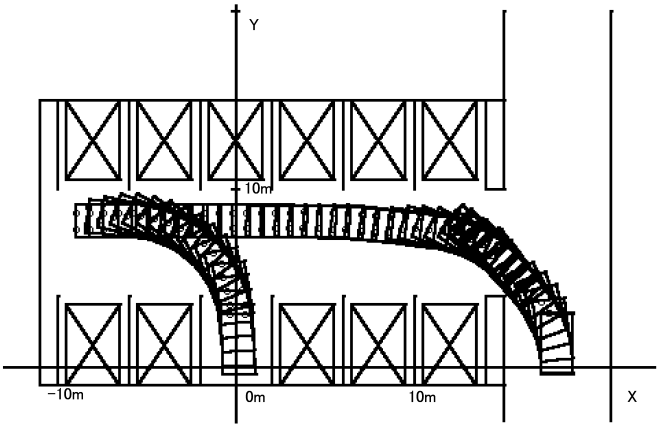

Fig.11 An example of parking trajectory from (18.0m, 0, 0.5π) to (0, 0, 0.5π)

その後，駐車スペースの奥の (-6.0m，8.0m，π) を戦略目標として設定し，最終目標 (0m,0m,0.5π) に後退により進入し終了している．これらの運転知識は,「車庫に寄せる (16m,4m,0.75π)」,「切り返し位置へ移動する (-6.0m，8.0m，π)」,「車庫に入れる (0m,0m,0.5π)」，といった知識に意味付けることができる．

Fig.12 は，上記シミュレーションよりさらに駐車知識の獲得が困難な例として，一番奥の駐車スペースであり，前進では駐車できず，入り口から後退で入る必要がある形状の場合を想定し，知識獲得を行った結果を示す．初期状態

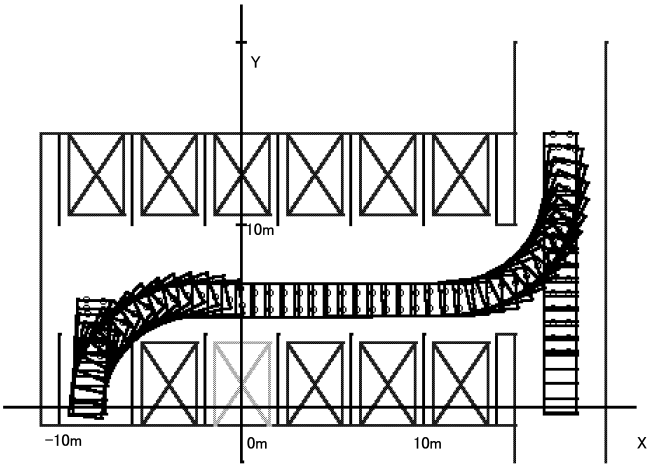

Fig.12 An example of parking trajectory from (18.0m, 0, 0.5π) to (-8.0m, 0, 0.5π)

(18m,0m,0.5π) から後退で駐車場へ進入するため入り口を通り過ぎて (18m,12m,0.5π) の位置に戦略目標を置き，移動させている．この地点から，車は後方に動くことになる．次に，駐車スペースの近くに寄せるため (-4m,6m,0) に戦略目標を設定し，その後最終目標 (0m,0m,0.5π) に移動させている．最後に駐車スペースに入れる時に，右隣の車との距離が少し近かったため（ハンドルを切りきれず）最終目標に一回で到達できず，幅寄せに相当する戦術目標の設定により入れ直している．

このシミュレーションによる知識獲得では，報酬として出発してから駐車までの所要時間を用い，9 回のランダム選択を実行し冒険を行うと共に，1 回の最大値選択（グリーディ選択）を用いて知識の強化を行うことにより，局所最適解からの脱出ができている．今回のシミュレーションでは，全く先見的知識が無いとし，単純な設定で行っているため知識獲得のための時間を要している．しかし，少しの共通的知識を用いることにより，早く良い解を得ることが可能である．

これらのシミュレーション結果により，強化学習（PSP-学習）を使用することで，従来，ノンホロノミックな車の特性を把握している熟練運転者が試行錯誤的に記述していた運転知識を獲得できることを確認した．

6　おわりに

本論文では，ノンホロノミックな特性をもち運転が難しい自動車の駐車制御に対して，予見ファジィ制御を用いた知的自動車制御システムの運転知識の獲得のため，最終目標への到達から途中の行動に報酬を与える PSP-学習法を適用した．具体的駐車場を想定し，シミュレーションを行った結果，大局的な運転知識を獲得できることを確認した．

ここでハンドル，速度操作に用いた予見ファジィ制御は，人間と同様に地点と方向で与えられた目標に障害物を回避

しながら車を移動させる．この機能は，人間の操作を代替するものであり，人間の柔軟な，言い換えると指示通りに運転するとは限らない運転を組み込んでいる．そのため，ここで用いた駐車システムは，人間を運転部分に組み込んで構成することが可能であり現在適用実験を進めている．ここで獲得した知識を用いることにより，駐車場の誘導員が行っているように，未熟な運転者などに適切な支援することが可能となる．

本研究の一部は，日本学術振興科学研究費補助金基盤研究 (C)「福祉車両操作の知的運転支援システムの研究」(課題番号 12650407) の支援によるものである．

参考文献

［1］ R.M. Murray, S.S. Sastry: Nonholonomic Motion Planning: Steering using Sinusoids, IEEE Transc. on Automatic Control, vol.38, no.5, 700/716 (1993).

［2］ J. Barraquand, and J.C. Latombe: Nonholonomic Multibody Mobile Robots - Controllability and Motion Planning in the Presence of Obstacle, Proceedings of the 1991 IEEE Conference in Robotics and Automation, California, 2328/2335 (1991).

［3］ 安信誠二: ファジィ工学, 1/177, 昭晃堂 (1991).

［4］ S. Yasunobu and S. Miyamoto: Automatic train operation by predictive fuzzy control, Industrial Application of Fuzzy Control (M. Sugeno, Ed.), North Holland, 1/18 (1985).

［5］ S. Yasunobu and N. Minamiyama: A Proposal of Intelligent Vehicle Control System by Predictive Fuzzy Control with Hierarchical Temporary Target Setting, Proc. of Fifth IEEE International Conference on Fuzzy Systems, New Orleans, 873/8781 (1996).

［6］ S. Yasunobu, S. Saitou and Y. Suryana: Intelligent Vehicle Control in Narrow Area based on Human Control Strategy, World Multiconference on Systemics, Cybernetics and Informatics (SICI 2000), Vol.VII, 309/314 (2000).

［7］ Richard S. Sutton and Andrew G. Barto (三上貞義，皆川雅章　訳): REINFORCEMENT LEARNING: An Introduction（強化学習）, 森北出版，1/351 (2000).

［8］ 卯木輝彦，末竹則哲: 強化学習による無人搬送車の分散型スケジューリング, 電気学会論文誌, Vol.117-C, No.10, 1513/1520 (1997).

［9］ J.J. Grefenstette: Credit Assignment in Rule Discovery System Based on Genetic Algorithms, Machine Learning 3, Kluwer, 225/245 (1988).

［10］ T. Horiuchi, A. Fujino, O. Katai and T. Sawaragi: Q-PSP learning: An Exploration-Oriented Q learning and Its Applications, The Society of Instument and Control Enggineers, Vol.35, No.5, 645/653 (1999).

［11］ Yaya Suryana and Seiji Yasunobu: Hierarchical Intelligent Control by PSP-learning and Cascade Fuzzy Control for Parking Vehicle in Narrow Area, T.IEE Japan, Vol.122-C, No.2, 315/316 (1997).