ファジィ目標の強化学習による獲得と非ホロノミック移動体の移動制御への応用

Acquisition of Fuzzy Target by Reinforcement Learning and Its Application to Nonholonomic Vehicle Control

松原 智也
Tomoya Matsubara
筑波大学 理工学研究科
Master's Program in Science and Engineering, University of Tsukuba

安信 誠二
Seiji Yasunobu
筑波大学 機能工学系
Institute of Engneering Mechanics and Systems, University of Tsukuba

Abstruct: A "fuzzy target" is a control target value defined by "fuzzy set". In this paper, it is acquired by a reinforcement learning, which is a kind of machine learning. The membership value of fuzzy target is an evaluation value of target elements obtained by the reinforcement learning. By using this fuzzy target, a control system is constructed which is able to correspond to changes in the environment flexibly. It was applied to a driving control of the nonholonomic vehicle under the change of the situation. Simulation results show the effectiveness of this fuzzy target.

1. はじめに

人間は，様々な状況での経験を通して知識を獲得し，これを柔軟に利用することによって，状況が変化した場合でも適切に行動を決定している．このような過程で人間が行っている思考や行動，および人間が取り扱う知識などの情報は，状況に対して唯一のものではなく，あいまいさを含んでいる．

このような，あいまいさを含む情報を扱う方法として，ファジィ理論がある．ファジィ理論は，自然言語処理，意思決定などの，あいまいさを含む情報の処理が本質的な，複雑な問題に応用することができる [1].

我々はこのファジィ理論を用いて，熟練者が持つ知識や経験をメンバーシップ関数に記述し，制御に応用している [2]．しかし，知識や経験だけではなく制御を行う際の目標値も，現実世界においては，度合いを持った集合，すなわちファジィ集合であると考えることができる．ここでは，この制御目標値のファジィ集合を「ファジィ目標」[3] と呼ぶ.

本論文では，このファジィ目標を，機械学習の一種である「強化学習」により試行錯誤的に獲得し，学習の過程で得られる各目標要素に対する報酬（タスク達成に関する成功の度合い）を，ファジィ目標のメンバーシップ値とする．獲得したファジィ目標を，非ホロノミックな移動体である，四輪車の移動制御に応用し，新たな制約の付加によって環境の変化した場合でも，柔軟な対応が可能な知的制御システムを構築する．このファジィ目標を用いた知的制御システムの有効性を，実際の車両を模擬したシミュレーションによって評価する.

2. ファジィ目標の概要

「ファジィ目標」は，度合いを持った複数の目標値を，ファジィ集合として扱ったものである (図 1)．ここでは，学習の際に獲得した各目標要素に対する報酬を，ファジィ目標 "Target is Good" のメンバーシップ値とする.

制御を行う際にファジィ目標を用いる利点として,「目標の代替案」を含んでいる点がある．ファジィ目標は目標値となる複数の要素を含んでいる．そのため，新たに制約が加わった場合でも，適用可能な代替案を用いることで，制御目標を設定し，最適なタスク達成が可能である.

本論文では，四輪車の移動制御において，障害物の無い，十分に広い環境下で獲得したファジィ目標を，障害物を配置し，制約が付加されることによって環境が変化した状態で利用する．そして，適用可能な代替案を用いて，環境の変化に柔軟に対応する.

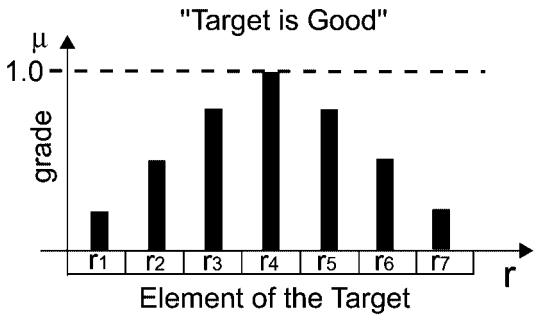

Fig. 1: The Fuzzy Target

3. ファジィ目標に基づいた知的制御システム

四輪自動車は，２入力 (ハンドルと速度の操作) で，３出力 (位置 (x, y) と車体の向き) を制御する，非ホロノミックな移動体である．そのため，現在状態から最終目標へ車両を移動させるためには，自車および周囲の状況に応じて，適切に途中経由地点 (サブゴール) を設定し，移動制御を行う必要がある．

本章では，四輪車の特性について触れた後，強化学習によるファジィ目標の獲得法と，ファジィ目標に基づいた制御システムについて述べる．

3.1. 四輪車の特性

前輪操舵の四輪自動車において，十分速度が遅いとき，タイヤの滑りや遠心力の発生を無視することができる．その際，四輪自動車は，図 2 に示すような幾何学的な動きをする．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，*Ackerman-Jeantaud* の操舵機構により構成されている．

この自動車の動きは，次に示す拘束条件式によって記述できる:

$$\frac{dx}{dt} = v\cos\phi\cos\theta,$$
$$\frac{dy}{dt} = v\cos\phi\sin\theta,$$
$$\frac{d\theta}{dt} = \frac{v}{L}\sin\phi$$

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の長さである．

四輪自動車は上記のような特性を持つので，現在状態 (x_0, y_0, θ_0) から，任意の状態 (x, y, θ) への移動は不可能である．例えば，車両を真横に移動させるためには，切返し地点を設定する必要がある．このような理由で，現在状態から最終目標 (x_G, y_G, θ_G) へ車両を移動させるためには，自車両および周囲の状況に応じた，適切なサブゴールを設定し，これに基づいて移動制御を行う必要がある．

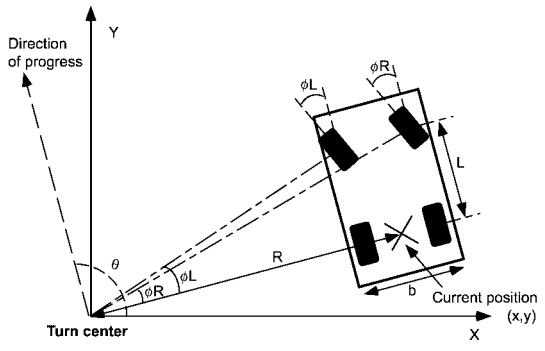

Fig. 2: Kinematics constraint in nonholonomic vehicle

3.2. ファジィ目標の獲得

サブゴールを適切に設定する際に必要な「目標設定知識」を，機械学習の一種である「強化学習」(Reinforcement learning) [4] によって獲得する．この学習法は，人間や動物が試行錯誤的に成功を学習する仕組みを模倣している．ある行動を選択したことによって，成功した場合には報酬が，失敗した場合には罰が与えられる．これを繰り返すことにより，タスクを達成するために必要な知識を獲得する．本研究では，強化学習を，その状況における制御指令の獲得のために使用するのではなく，サブゴールを設定する際に必要な知識を獲得するために使用する．

PSP (Profit Sharing Plan)-learning [5] は，強化学習の手法の一つとして提案されている．この学習法は，図 3 のように，初期状態から最終目標に至る，一つのエピソードにおいて，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-learning 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し，その報酬に個々の目標達成度をファジィ評価によって行う，fPSP (fuzzy Profit Sharing Plan)-learning [6] を用いる．fPSP-learning によって得られる知識は，"IF <condition:c_n>THEN<action:a_n>" なる形式の，状態と行動のペアと，その評価値からなる規則（S-table）である．

目標設定を獲得するためには，各状態 c_n において，目標 a_n をルーレット選択することで，未知の経験を試みる．実行した結果の報酬は，選択されたすべての 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，報酬は制限時間と所要時間の差で与える．例えば，制限

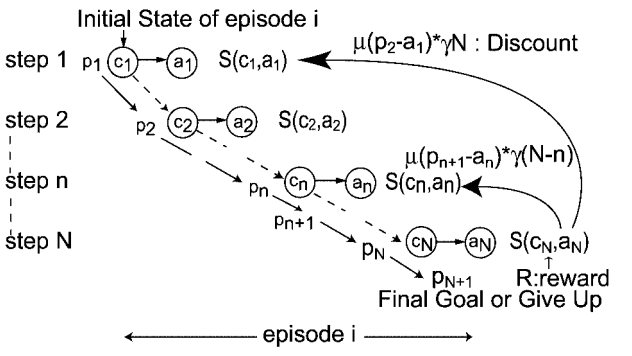

Fig. 3: PSP-learning deitributes reward of penalty to the previous fired rules

時間を 250 秒とし，最終的な目標に 80 秒で達したならば，出発地点で用いた知識に対する報酬は 170 に，経由した目標地点の数で割り引いた値である．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，罰を与え，用いた目標地点に対する評価を下げる．

学習後，得られた評価値をメンバーシップ値としたファジィ集合で，ファジィ目標を作成する．

3.3. 制御システムの概要

獲得したファジィ目標に基づいて四輪車の移動制御を行う，階層型知的制御システムの概要を図 4 に示す．

このシステムは，状況監視部，ファジィ目標設定部，自動運転部の 3 階層から成る．

状況監視部は，障害物への接触，最終目標およびサブゴールへの到達について評価し，新しい目標の必要性を判断する．何らかの理由で現在のサブゴールへの到達が困難となった場合は，現在のサブゴールをリセットし，ファジィ目標設定部に対し目標設定指令を出力する．

目標設定指令を受け，ファジィ目標設定部は，強化学習によって得られた S-table から，ファジィ目標を設定する．

自動運転部では，ファジィ目標を用いて制御指令を算出する．その算出過程を図 5 に示す．ファジィ目標では，個々の要素をサブゴールとして仮定し，図 6 に示すカスケードファジィ制御器 [6] によって操作指令候補を算出する．算出された操作指令候補を用いて車両を仮想的に動かし，車両の将来状態を予見する．予見した将来状態と，サブゴールの評価値 (メンバーシップ値)，サブゴールとの距離偏差，角度偏差，および障害物との距離を，総合的にファジィ評価し，操作指令候補の評価値を求める．これを，ファジィ目標の全ての要素に対して行い，最も評価値の高かった操作指令候補を選択し，これを制御指令として実車両に与える．

4. コンピュータシミュレーションによる評価

構築した知的制御システムの有効性を，実際の車両を想定したコンピュータシミュレーションにより検証する．今回は，初期位置 $(18m, 10m, 0.0\pi)$ から最終目標まで移動させる．

障害物の無い広い領域において，fPSP-learning を用いて，初期位置 $(18m, 10m, 0.0\pi)$ において獲得したファジイ目標のうち，評価値が 200 以上のものを次に示す．

Target position	Membersip value
(2m, 10m, 0.0π)	0.301
(12m, 6m, 0.0π)	0.367
(20m, 20m, 0.25π)	0.999

要素 $(20m, 20m, 0.25\pi)$ に対するメンバーシップ値が 0.999 と最も高く，この要素，すなわち最終目標を目標値とすることが，タスク達成に関し，最も価値があることを表している．

獲得したファジィ目標に基づいて，障害物が無い場合 (Case(i))，初期位置の前方に障害物を配置した場合 (Case(ii)) を想定し，シミュレーションを行った．その結果を図 7 に示す．

障害物が無い場合 (Case(i))，最終目標を目指し，到達することができた．最終目標到達までの経過時間は 27 秒であった．

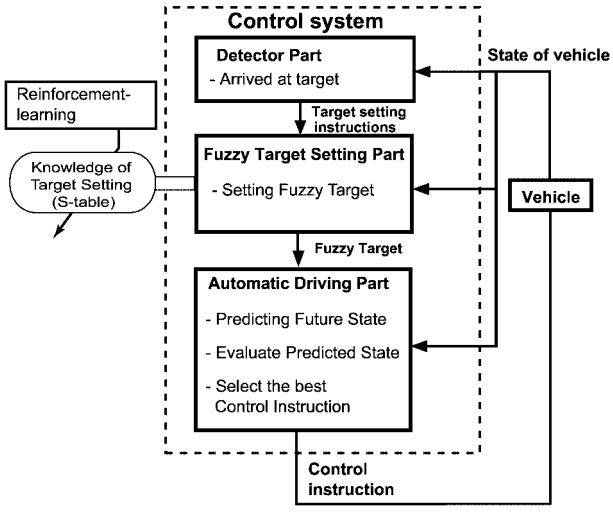

Fig. 4: Outline of the control system

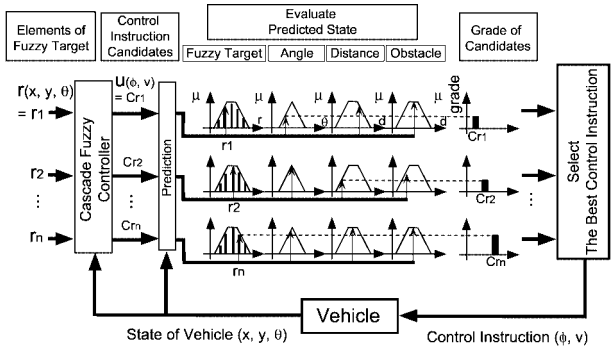

Fig. 5: Control system based on Fuzzy Target

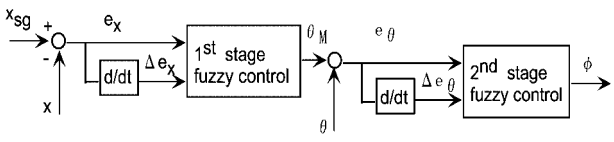

Fig. 6: Cascade fuzzy controller

障害物を初期位置の前方に配置した場合 (Case(ii)), 前方に移動することができず，また，四輪車の非ホロノミックな特性により，真横に平行移動して障害物を回避することもできない．そのため，獲得したファジィ目標の中から，要素 $(12m, 6m, 0.0\pi)$ をサブゴールとして選択し，後退で向かうよう制御指令を算出した．サブゴール $(12m, 6m, 0.0\pi)$ に向けて後退したところ，$(12m, 6m, 0.0\pi)$ 付近において，次のサブゴール $(6m, 2m, 0.0\pi)(\mu = 0.548)$ を選択し，この位置に向かうことで，障害物を避け，最終目標 $(20m, 20m, 0.25\pi)$ へ直接移動可能な状態となった．目標を最終目標 $(20m, 20m, 0.25\pi)$ とし，切返しを行うことによって，最終目標に到達した．最終目標到達までの経過時間は 100 秒であった．

このように，過去に獲得したファジィ目標に基づいて車両の移動制御を行うことで，障害物の配置によって変化した環境に，メンバーシップ値は低いが適用可能な代替案を用いることで適応し，最終目標に到達することができた．これにより，ファジィ目標の有効性を，シミュレーションによって確認した．

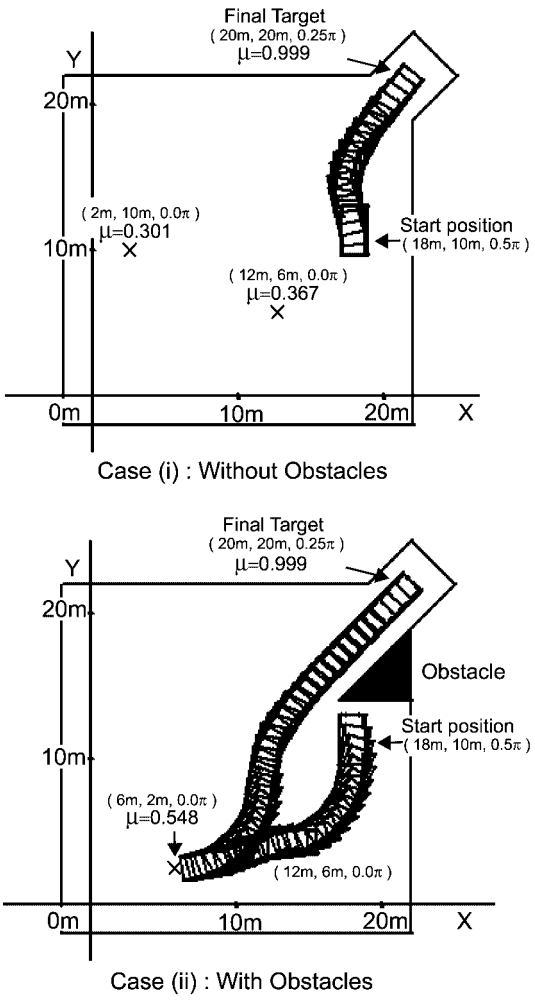

Fig. 7: The trajectory from (18m,10m,0.5π) to (20m,20m,0.25π)

5. おわりに

本論文では，制御を行う際の目標値をファジィ集合で捉えた「ファジィ目標」を，機械学習の一種である「強化学習」を用いて獲得し，非ホロノミックな移動体である，四輪車の移動制御システムに応用した．その結果，ファジィ目標を用いることで，新たに制約が付加されたことによって環境の変化した場合でも，過去に獲得した知識を活用し，車両の移動を行うことができた．これにより，ファジィ目標の有効性を，シミュレーションによって確認した．

参考文献

[1] L. A. Zadeh: “Outline of a New Approach to the Analysis of Complex Systems and Decision Processes,” *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-3, No. 1, pp. 28-44, 1973.

[2] S. Yasunobu and S. Miyamoto: “Automatic train operation by predictive fuzzy control, Industrial Application of Fuzzy Control (M. Sugeno, Ed.),” *North Holland*, pp. 1-18, 1985.

[3] S. Yasunobu and T. Matsubara: “Fuzzy Target Acquired by Reinforcement Learning for Parking Control,” *Proc. of SICE Annual Conference 2003 in Fukui (SICE2003)*, pp.1303-1308, 2003.

[4] Richard S. Sutton and Andrew G. Barto: *REINFORCEMENT LEARNING: An Introduction*, A Bradford Book, 1998.

[5] 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, pp. 645-653, 1999.

[6] S. Yasunobu: “Fuzzy-Reinforcment-learning Design of Intelligent Controller for Nonholonomic Vehicle:” *Proc. of The* 12^{th} *FAN Intelligent Systems Symposium*, pp. 105-108, 2002.

連絡先

〒 305-8573
茨城県つくば市天王台 1-1-1
筑波大学機能工学系 知的制御システム研究室
松原 智也
電話: 029-853-6186
E-mail: tomoya_m@edu.esys.tsukuba.ac.jp