

ロジスティックマップを導入した強化学習の基礎研究

旭川高専 ○小平光夫, 渡辺美知子, 古川正志

要 旨

ロジスティックマップから生成されるカオスの基本的な性質を調べ, その性質を強化学習でよく知られたレーストラック問題へ適用する可能性を検討する. カオスのもつメモリ機能や埋め込みに注目すると, 初期値によりその後の時系列挙動がすべて定まり, メモリとして利用できる. レーストラックの各位置にメモリとしてこの初期値, 速度およびその経路評価ともに登録し, それらを利用した最短経路の探索をここでは試みる.

1. はじめに

カオスの埋め込み機能を利用した強化学習法がいくつか研究されている[1][2]. これらは時系列の中で過去の系列を保存しその近傍の望ましい状態遷移の埋め込みを行動の選択に利用する方法(三上, 和田, 嘉数)と, 望ましい状態遷移の埋め込みをコホーネンニューラルネットワークに保存しそれを行動の選択に利用する方法である(R.Der and M.Herrmann). しかしながら, これらの方法は予め望ましい状態がわかっており, そこへの近傍状態からの埋め込みによる遷移を利用している. 一方, レーストラック問題のような強化学習問題では次の時刻の望ましい状態を事前に行うことができないため, 上記のような方法を採用することが不可能である.

本研究ではカオスのもついくつかの性質を調べるとともに, 特に, カオスが初期値依存性をもつことと, 初期値を保存することで以下の時系列状態をいつも簡単に発生できることを適用し, これらをレーストラックのコース地点に登録し, それらを利用することで最短経路を求める方法を提案する. カオスの生成にはロジスティックマップを利用した. 数値計算実験の結果ではQ-Learningによる走行時間とほぼ同等の結果を得ることができ, また, 学習時間は試行錯誤を行うQ-Learningよりはるかに短い時間で最短経路を得ることができた.

2. ロジスティックマップによるカオスの生成と強化学習アルゴリズム

よく知られたロジスティックマップは以下の差分方程式から導かれる.

$$x_{n+1} = a_0 x_n (1 - x_n) \quad (1)$$

ここで, 式(1)は係数 a_0 によって状態の分岐が生じ(図1), $3.57 < a_0 \leq 4.0$ のときカオス状態となる. カオスは初期値に対する鋭敏な依存性を持ち, かつ, 埋め込み次元 d と遅れ時間 τ によって d 次元状態空間

を再構成でき, 元の決定論的法則を再現できる. $d=3$ の時の様子を図2に示す.

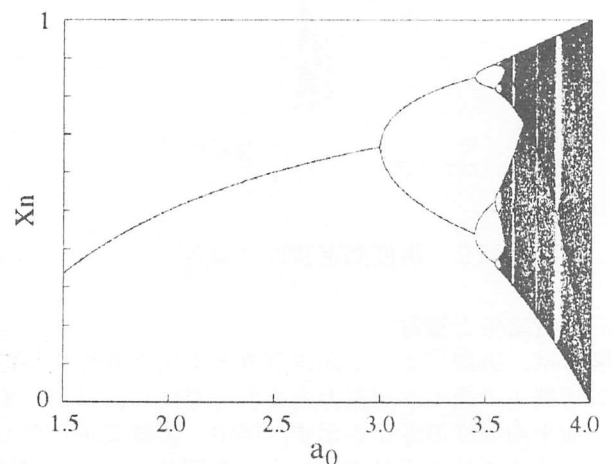


Fig.1 Branch of logistic map.

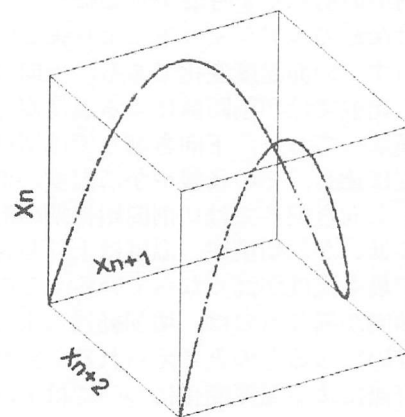


Fig.2 Embedding of chaos (a times delay is 3)

三上やR.Derによる方法はこうした埋め込みを利用した局所再構成による強化学習法を提案している. そのアルゴリズムは以下である.

- 1) 状態 x と係数 a を与える.
- 2) (x, a) におけるQ-値の繰り返し評価関数

$$\Delta Q(x,a) = \mathcal{E}_Q(r(t_n) + \gamma E(x') - Q(x,a)) \quad (2)$$

により更新する。ここで、 x, x' は時間 $tn-1, tn$ の状態であり、 a は

$$a = \arg \max_a Q(x,a) \quad (3)$$

である。また、

$$E(x') = \max_a Q(x,a) \quad (4)$$

である。

3) 行動 a_k は

$$P(a_k) = \frac{e^{Q(x,a_k)/T}}{\sum_k e^{Q(x,a_k)/T}} \quad (5)$$

で確率的に定める。

この手続きを採用するためには、状態 x' が近傍 x^* をもち、 x^* が望ましい状態 x にかわる局所再構成による埋め込みかなんらかの方法によるこのような状態遷移関数を必要とする。

3. ロジスティックマップによるレーストラックの経路探索

Q-Learningとして知られるレーストラック問題は学習が終了したと見なされる時点ではじめて望ましい状態が明らかとなる。従って、従来提案されているような方法を採用することができない。ここでカオスの初期値に依存してその後の時系列を再現できる特性をメモリーとして利用した方法を提案する。

3.1 レーストラック問題

図3に示したようなレーストラックを考える。ここで、車はスタート地点から出発しゴールに達することを目的とする。車の初期状態は速度 $v=(0,0)$ とする。制御量を加速度とし、 $a=(a_x, a_y; a_x, a_y \in \{-1, 0, 1\})$ とする。時刻 t の位置は

$$x(t) = x(t-1) + vx(t-1) + ax(t-1)$$

$$y(t) = y(t-1) + vy(t-1) + ay(t-1)$$

で与える。

3.2 カオスによる探索アルゴリズム

以下に今回用いた探索アルゴリズムを述べる。

1. 初期値 x_0 , 係数 a_0 をランダムに設定する。
2. 式(1)により x_1 を求め、 $0 \leq x_1 \leq 1/3$ ならば $a=-1$, $1/3 < x_1 \leq 2/3$ ならば $a=0$, $2/3 < x_1 \leq 1$ ならば $a=1$ とし、次の位置 $(x(1), y(1))$ に移動する。
3. 2を500回繰り返す、ゴールに到達したときの

み、1で設定した初期値で再度計算し、1ステップごとの位置に係数 a_0, x_n, g (ゴールまでの残りのステップ数)、速度 v を登録する。ただし、すでにデータが登録されている場合には、 g を比較して少ない方を残す。

4. 1~3を4000回行った後、スタート地点から各位置に登録されたデータに従って行動する。ただし、速度は $v=(v_x, v_y; v_x, v_y \in \{-3, -2, -1, 0, 1, 2, 3\})$ とし、壁に衝突した場合は $v=(0,0)$ とする。

4. 数値計算実験

3.2のアルゴリズムを用いて数値計算実験を行い、その結果を図3に示す。図3の細線は学習前、太線は学習後の車の経路である。学習前の時間ステップは132、学習後は18である。

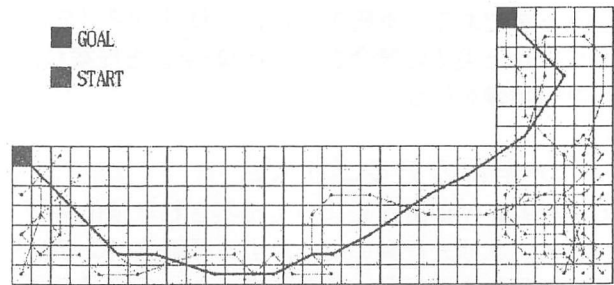


Fig.3 Experimental result.

5. おわりに

本研究でカオスのメモリー機能を、最短経路の探索に適用できることを実験を通じて示した。現段階ではこの方法は強化学習法とはいえないので、今後はカオスの埋め込みや本研究で利用したメモリー機能などの性質を適用した強化学習法を検討する必要がある。

参考文献

- [1] 三上 貞芳, 和田 充雄, 嘉数 侑昇: 群強調行動の獲得を目的としたカオティック強化学習の提案, ロボティクス・メカトロニクス講演会'96講演論文集(1996)
- [2] R.Der, M. Herrmann: Q-Learning Chaos Controller, IEEE(1994)
- [3] 廣田 薫: 知能工学概論, 昭晃堂(1996)
- [4] 合原 一幸: カオス-カオス理論の基礎と応用-, サイエンス社(1990)