科学研究費助成事業　　研究成果報告書

平成　２６　年　　６　月　１２　日現在

機関番号：３２６４４
研究種目：基盤研究(C)
研究期間：2009～2013
課題番号：２１５００２２２

研究課題名（和文）機械学習における学習の停滞現象と損失関数の極値の安定性

研究課題名（英文）Plateau Phenomena of the Learning Dynamics and Stabilities of the Local Minima of the Error Function in Machine Learning

研究代表者
尾関　智子（Ozeki, Tomoko）

東海大学・情報理工学部・教授

研究者番号：１０４０７９９２

交付決定額（研究期間全体）：（直接経費）　　3,400,000円、（間接経費）　　1,020,000円

研究成果の概要（和文）：　機械学習は，人間の脳のように外界から与えられたデータを学習することができるシステムを構築するための理論である．学習アルゴリズムは，教師あり学習，教師なし学習，強化学習の３つに大きく分類できる．本研究では，教師あり学習と強化学習のダイナミクスの解析と改善を行った．(1) 隠れマルコフモデルにおける特異構造と学習ダイナミクスの関係を数値実験により示した．（2）環境の変化に適応できる強化学習アルゴリズムの提案とそのダイナミクスの解析を行った．

研究成果の概要（英文）：Machine learning is one of the theories to construct the systems that can learn the data given from outside world like human brains. The algorithms of machine learning are divided into three categories such as supervised learning, unsupervised learning and reinforcement learning. In this research, we have investigated the dynamics of supervised learning and reinforcement learning and proposed some improvements. (1) We have investigated the relation between the singular structure of the parameter space of hidden Markov models and the trajectories of the learning dynamics. (2) We have proposed the reinforcement learning algorithm that can adapt to the changing environments and investigated the learning dynamics.

研究分野：総合領域

科研費の分科・細目：情報学・感性情報学・ソフトコンピューティング

キーワード：知能情報処理　機械学習　多層パーセプトロン　隠れマルコフモデル　強化学習

１．研究開始当初の背景

機械学習のアルゴリズムの一つである教師あり学習は，内部にパラメータをもつ学習モデルを仮定し，外部から与えられる多数の訓練データに潜む法則や数学的な構造を推論するものである．学習モデルが内部パラメータを変えることにより，望ましい出力を獲得していく過程を「学習」とよぶ．機械学習の中でもニューラルネットワークは生物の脳に学んだモデルであり，新しい情報処理様式として注目されてきた．工学的には，パターン認識やデータ解析などの幅広い分野に応用されている．しかし，ニューラルネットワークでは学習が途中で停滞してしまうという問題を抱えており，その原因は必ずしも明らかではなかった．

学習の停滞現象はニューラルネットワークに限らず，RBF（Radial basis function）ネットワークや多層パーセプトロン，混合ガウス分布などの学習（統計）モデルにも見られる．これは，学習モデルが階層構造をもつ場合に起こる，避けては通れない問題である．階層構造をもつ学習モデルでは，パラメータ数の少ない小さなモデルが大きなモデルのパラメータ空間に特異構造を生み出す．我々は，円錐モデルや多層パーセプトロン，混合ガウス分布において特異構造が統計的推測や学習ダイナミクスに与える影響を研究してきた．特異構造をもつ学習モデルでは，パラメータ推定や仮説検定，ベイズ推定，モデル選択，学習ダイナミクスに奇妙な振る舞いが見られることが明らかになってきた．

近年になって，萩原が多層パーセプトロンではモデル選択の基準である AIC があまり良い性能を与えないことを発見した[1]．甘利らは，特異モデルのパラメータ空間はリーマン空間であること，また，特異点上ではリーマン計量が縮退することなどを論じ，特異モデルにおける統計的推定において大きな成果をあげた[2]．また，学習データをもとに逐次的にパラメータを変えて学習していくときのダイナミクスの研究も行ってきた[3]．渡辺は特異モデルにおけるベイズ推論に代数幾何の方法を用いてさまざまな成果を上げてきた[4]．福水は接錐の方法を用いて，パーセプトロンの特異点における対数尤度の解析を行った[5]．特異点の研究は日本においてめざましい発展を遂げている．

機械学習のもう一つのアルゴリズムに分類される強化学習は，環境から与えられる報酬を手がかりに試行錯誤により最適な行動を獲得するものである．環境の変化に適応的に対応することができることが期待されているが，環境の急激な変化には対応できない．エージェントがゴールを探索する迷路問題に対してはいくつかの方法が提案されていた[6]．

２．研究の目的

(1) 特異点や学習の停滞現象は，多層パーセプトロンなどの学習モデルだけではなく，ARMA モデルや隠れマルコフモデルなどの時系列モデルにも見られる現象である．シミュレーション実験を通して，パラメータ空間の特異構造と学習の停滞現象の関係を明らかにしていくことを目的とする．

本研究では，特に隠れマルコフモデルを対象とする．隠れマルコフモデルは，音声認識や遺伝子解析，自然言語処理にもちいられ，実用化されている．これらのモデルが応用上利用される場合は，通常パラメータ数などの最適なモデルの大きさは事前にはわからず，大きなモデルから始めてモデルを小さくするということが頻繁に行われる．このとき，小さなモデルが大きなモデルの空間で特異構造をなし，推定に影響を与える．隠れマルコフモデルのもつ特異構造がもたらす影響を解明することは応用の面からも重要である．

(2) 環境の急激な変化に対応できる強化学習アルゴリズムを提案し，その学習ダイナミクスを解析する．

３．研究の方法

(1) 隠れマルコフモデルのダイナミクス

２状態隠れマルコフモデル（2-HMM）を考える．遷移確率行列を

$$A=\begin{pmatrix} 1-a_{12} & a_{12} \\ a_{12} & 1-a_{12} \end{pmatrix}$$

とし，隠れ状態からの出力確率をガウス分布

$$p(O\,|\,m)=\frac{1}{\sqrt{2p}}\exp\left[-\frac{1}{2}(O-m)^2\right]$$

とする．初期状態に関する確率をp_1, p_2とすると，学習モデルのパラメータは$(a_{12}, a_{21}, m_1, m_2, p_1, p_2)$となる．

まず始めに，状態２から状態１への遷移が起こらない left-to-right モデルを考える．Left-to-right モデルでは，図１のようにパラメータは(a_{12}, m_1, m_2)のみとなる[7]．

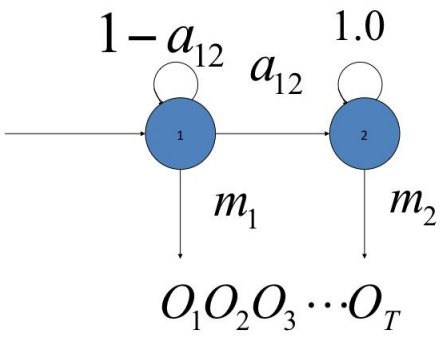

図 1：学習モデル（left-to-right）

2-HMM は１状態隠れマルコフモデル（1-HMM, 図 2）を含む．つまり，2-HMM のパラメータ

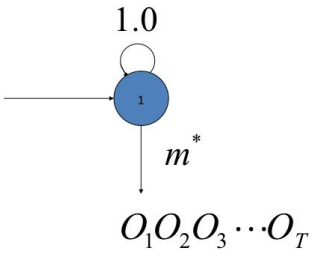

図 2：1状態隠れマルコフモデル

空間は1-HMMを実現するパラメータを無数に含んでいる．パラメータが以下の条件
1）a_{12}=0，m_1=m^*
2）m_1=m_2=m^*

を満たすとき，2-HMMは1-HMMとなる．1)の場合はm_2が同定不能となり，2)の場合はa_{12}が同定不能となる．

さらに，パラメータ空間上では，図3のように赤線上のパラメータ集合が同じ1つの1-HMMを表す．$(0, m^*, m^*)$は特異点となる．

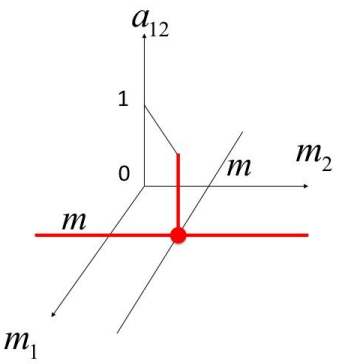

図 3：2-HMMのパラメータ空間

真のパラメータの位置を次の3つに分類する．1つ目は，真のパラメータが「特異点上」にあるとき，つまり真の確率モデルが1-HMMであるときである．2つ目は，真のパラメータが「特異点の近傍」にあるとき，すなわち，$m_1^* \approx m_2^*$となるときである．真の分布が「特異点上」,「特異点の近傍」,「特異点から遠いところ」にあるときの学習ダイナミクスの振る舞いをシミュレーションにより検証する．

(2) Concurrent Q-Learningのダイナミクス

Concurrent Q-Learning (CQL)とは，環境の動的な変化に対応できる強化学習法の一種である．CQLは，状態s*を目的地としたときの，状態sにおいて行動aをとる状態行動価値$Q^{S*}(s,a)$を同時に学習することで,すべてのゴールの可能性を同時に解決する学習手法である．CQLは，最適な経路を学習するためのRelaxationという

$$Q^{S_C}(s_A,a) \geq Q^{S_B}(s_A,a) \times \max_{a'} Q^{S_C}(s_B,a')$$

を満たさない場合は近道の価値を高くする処理と，最適な経路のみに適格度トレースを適用するNow Update Traceという処理を採用している．

しかし，図4に示すようにCQLを迷路問題に適用すると，環境変化の後にいったん学習した最短経路を保持せずゴールまでのステップ数が上昇してしまうという現象が見られる．Relaxationは，価値を高くするばかりで低くすることがないため，すべての価値が高い値で収束してしまい,正しい行動選択が行われなくなるためである(1).

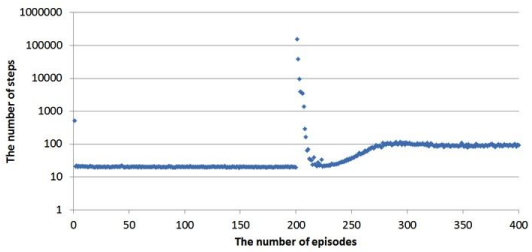

図 4：CQLのダイナミクス

そこで，Relaxationを適用する際に，同時に迂回路の価値を減少させる．以上のようにCQLを改良した手法を提案し，適格度トレースを合わせた場合の学習ダイナミクスを考察する．

4．研究成果

(1) 隠れマルコフモデルのダイナミクス

① 教師のパラメータが特異点上にあるとき（$a_{12}^* = 0, m^* = 3$）

真のパラメータが特異点上にある場合には，状態遷移パラメータa_{12}が素早く0に収束し（図5），m_1は素早く真のパラメータm^*に近づくが，同定不能なパラメータであるm_2もゆっくりとm^*に近づくことがわかる（図5）．この場合は深刻な学習の停滞現象は見られない．

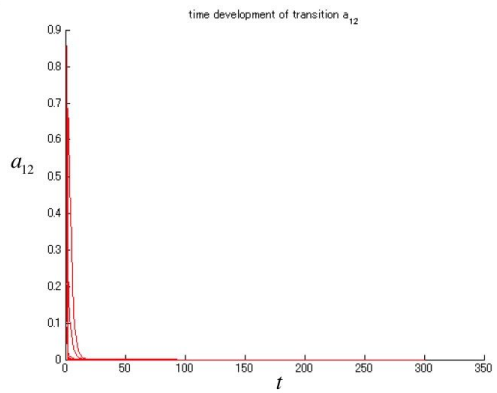

図 5：遷移確率の時間変化

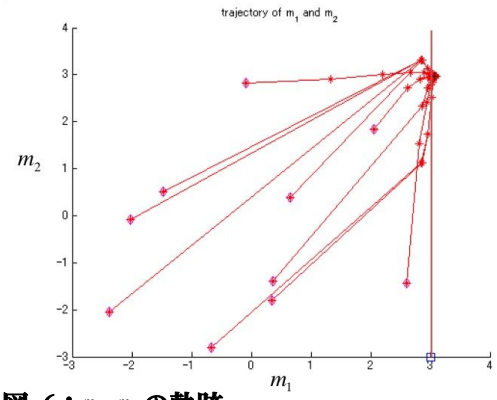

図 6：m_1, m_2 の軌跡

② 真のパラメータが特異点近傍にあるとき（$a_{12}^* = 0.4, m_1^* = 1, m_2^* = -1$）

図 7 と図 8 からわかるように，2-HMM はいったん $a_{12} = 0$ の 1-HMM(特異点)に近づいてから，2-HMM に戻っていくことがわかる．このとき，尤度が上昇しなくなり学習の停滞がおこる．

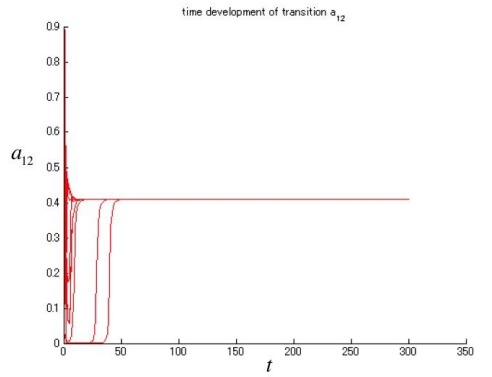

図 7： 遷移確率の推移

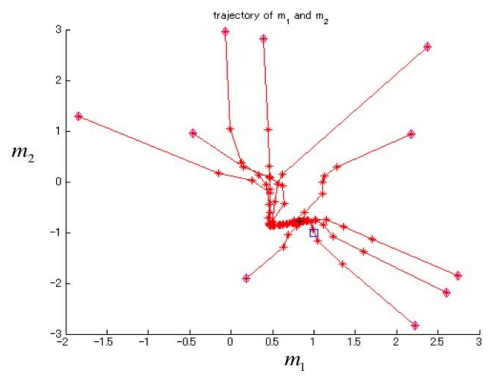

図 8：m_1, m_2 の軌跡

③ 真のパラメータが特異点の近傍にあるとき

$$\left(a_{12}^* = a_{21}^* = 0.7, m_1^* = 1, m_2^* = -1, p_1^* = p_2^* = 0.5\right)$$

図 9-11 は，状態 2 から状態 1 への遷移が起こりうる一般の 2-HMM モデルの学習ダイナミクスをシミュレーションしたものである．ある曲線に沿って学習の停滞が起こっている．

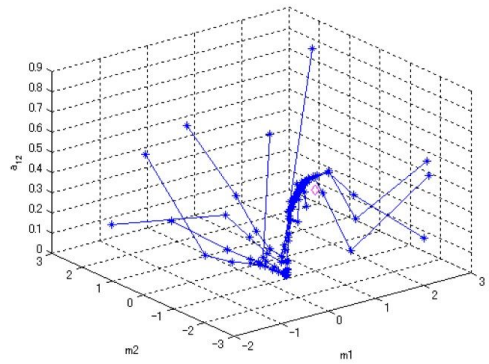

図 9：パラメータ空間上のダイナミクス

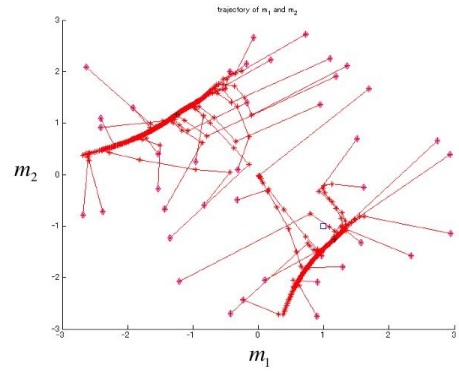

図 10：m_1, m_2 の軌跡

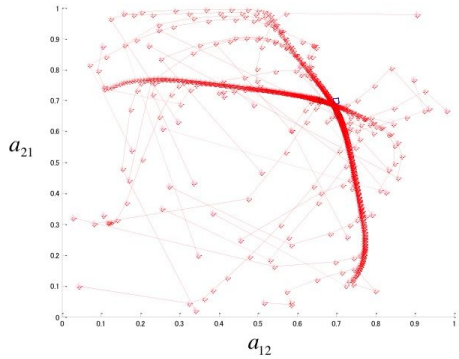

図 11：a_{12} と a_{21} の軌跡

(2) Concurrent Q-Learning のアルゴリズの改良とダイナミクス(1)-(4)

図 12 より，提案手法は環境変化後のステップ上昇をおさえることができていることがわかる．また，図 13 は relaxation と適格度トレースを併用した場合の学習ダイナミクスを示している．適格度トレースは，現在おこなった行動に対する価値の更新とともに，その行動に至るまでの行動の価値の更新を行うものである．したがって，学習がより早く進むことが期待される．しかし，Now Update Trace を用いた場合は，適格度トレースを用いない場合に比べ，ゴールへのステップ数が大きく，通常の適格度トレースを用いた場合は，適応がさらに遅くなることがわか

る．

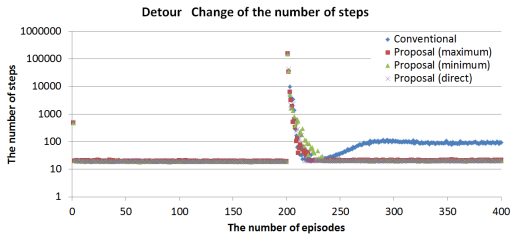

図 12：提案手法の学習ダイナミクス

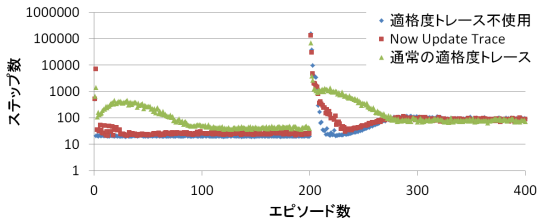

図 13：Relaxation と適格度トレース

参考文献

[1] Hagiwara, K., Neural Computation, 14, pp. 1979-2002, 2002.

[2] Amari, S. and Ozeki, T., IEICE Trans. Fundamentals, E84-A, 1, pp. 8, 2001.

[3] Amari, S., Park, H., Ozeki, T., Neural Computation, 18, pp.1007-1065, 2006.

[4] Watanabe, S., Neural Computation, 13, pp. 899-933, 2001.

[5] Fukumizu, K., The Annals of Statistics, 31(3), pp. 833-851, 2003.

[6] Ollington, R.B., Vamplew P.W.,Int. J. of Intelligent Systems, 20, pp. 1037-1052,2005.

[7] Yamazaki, K., Watanabe, S.,Neurocomputing,69, pp. 62-84, 2005.

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 3 件）

(1) 村上和謙，尾関智子，”Concurrent Q Learning における Relaxation の改良”, 東海大学紀要情報理工学部, 13, pp. 9-14, 2013.（査読あり）

(2) Kazunori Murakami, Tomoko Ozeki, “Improvement of the Relaxation Procedure in Concurrent Q-Learning”, Neural Information Processing Lecture Notes in Computer Science, 8227, pp. 84-91, 2013. （査読なし）

(3) 村上和謙，尾関智子，”Concurrent Q-Learning における Relaxation の改良”, 信学技報, 112(480), pp. 209-213, 2012.（査読なし）

〔学会発表〕（計 4 件）

(1) 村上和謙，尾関智子，“Concurrent Q-Learning における適格度トレースの影響”, 電子情報通信学会, 2014 年 3 月, 新潟大学.

(2) Kazunori Murakami, Tomoko Ozeki, “Improvement of the Relaxation Procedure in Concurrent Q-Learning”, ICONIP2013, 2013 年 11 月, Daegu, Korea.

(3) 村上和謙，尾関智子，“Concurrent Q-Learning と Sarsa, Q 学習の動的環境への適応”, IBIS2012, 2012 年 11 月, 筑波大学.

(4) 村上和謙，尾関智子，“動的環境における TD 誤差を用いた強化学習メタパラメータ学習法”, 電子情報通信学会, 2012 年 3 月, 岡山大学.

６．研究組織

(1)研究代表者

尾関　智子（OZEKI, Tomoko）

東海大学・情報理工学部・教授

研究者番号：１０４０７９９２

(2)研究分担者

該当なし

研究者番号：

(3)連携研究者

該当なし