

コミュニケーションの強化学習における ノイズ付加による連続値信号の離散化

柴田 克成[†]

[†] 大分大学工学部電気電子工学科 〒 870-1192 大分市大字旦野原 700 番地

E-mail: shibata@cc.oita-u.ac.jp

あらまし 筆者は、シンボルグラウンディング問題を解決するためには、シンボル処理とパターン処理を分けずに行う必要があり、そのためには、ニューラルネットを用いることが最良であると考えている。しかしながら、ニューラルネットはシンボル処理が苦手であると一般的に考えられている。本研究では、ニューラルネットを強化学習で学習させることによって必要に応じてシンボルが創発する可能性を探ることを目的とする。そこで、2 エージェント間の一方方向コミュニケーションを強化学習で学習する非常に単純なタスクを考え、発信側でのアナログコミュニケーション信号の生成と、受信側での受け取った信号からの行動生成をともにニューラルネットで行った。そして、信号伝達時にノイズを付加すると、単に強化学習に基づいてそのニューラルネットを学習させるだけで、学習可能なノイズレベルの範囲内において、ノイズレベルの増大とともに発信側では信号が離散化され、受信側でも受け取った信号から行動を離散化する傾向があり、結果的にノイズ付加による性能低下が抑えられることを示した。また、ニューラルネットをリカレントとすることで、信号の離散化がさらに促進されることを示した。

キーワード 強化学習, ニューラルネット, コミュニケーション学習, 信号の離散化, シンボルの創発

Discretization of analog communication signals by noise addition in reinforcement learning of communication

Katsunari SHIBATA[†]

[†] Dept. of Electrical & Electronic Engineering, Oita University, 700 Dannoharu, Oita 870-1192

E-mail: shibata@cc.oita-u.ac.jp

Abstract Towards the unified processing of symbols and patterns by neural networks, it was examined that symbols emerge using neural networks that is trained only by reinforcement learning. A very simple communication-learning task was assumed, and some noise is added to the communication signals. After learning, as the noise level during learning became larger, the communication signals were binarized more, and the system became more tolerant of noise unless the noise level was too large. The receiver was also trying to interpret the signals as binarized value. Furthermore, it was examined that recurrent neural networks promote the discretization.

Key words reinforcement learning, neural network, communication learning, discretization of signals, symbol emergence

1. ま え が き

シンボルは、われわれ人間の hochi 機能において重要な役割を果たしている。特に、コミュニケーションや論理的思考においては、不可欠なものである。このような高度な知能の実現においては、シンボル処理とパターン処理を分けて行うことが効率的であると考えられている。また、われわれの脳においても、シンボル処理は左脳で、パターン処理は右脳で行われていると

言われている [1]。

そして、この右脳と左脳での機能的な分化とのアナロジーから、デジタルコンピュータがシンボル処理を、ニューラルネットがパターン処理を行うという役割分担が一般的に受け入れられてきた。しかしながら、現在のところ、両者をつなぐ有望な手法はなく、「シンボルグラウンディング問題 [2]」は依然として重大な問題として残っている。

右脳と左脳は、前述のような機能的な差異が強調されている

が、にもかかわらず、外観を見ると、非常に良く似ている。そこで、筆者は、シンボル処理とパターン処理は別々に行うのではなく、両者を区別なく、一つの処理として扱うことが必要ではないかと考えてきた。そして、それを実現するためには、両者をニューラルネットで行うことが最良であると考えてきた。しかしながら、一般的に、ニューラルネットは、パターン処理は得意とするが、シンボル処理は苦手であると言われている。

Elman は、リカレントネットを用いて、文法的知識を獲得できることを示している [3]。この研究は、ニューラルネットがシンボルを扱うことができるという意味で重要な研究であるが、パターン情報との統合については言及されていない。

一方、大森らは、ニューラルネットベースの“PATON” というシステムを提案している [4]。ここでは、シンボルとパターンの情報を、連想記憶の機能を用いてニューラルネットで処理し、注意の機構が両者間の信号の流れを制御している。しかしながら、予め、シンボルを処理するところとパターンを処理するところに分かれており、両者は別のものであるとの前提がある。

本研究では、両者の処理を区別することなく、アーキテクチャの助けを借りつつ、学習を通して、必要に応じてシンボルが創発することを考える。ただし、ここでは、第一段階として、簡単のため、離散化された情報表現をシンボルと考える。世界にはさまざまな言語が存在することを考えると、シンボルは学習によって獲得するものと考えられる。また、強化学習は、生物の学習にヒントを得た自律的で合目的な学習手法である。そこで、本研究では、出力を離散化した方が有利な環境を想定し、単に強化学習を適用するだけで、ニューラルネットの出力が離散化するかどうかを検証する。本研究は、シンボルとパターンの一体化した処理、シンボルの創発への道を開くために、科学的、工学的の両面から意味があると考えている。

2. シンボル創発の要因

ここで、なぜシンボルが創発するのかを考えてみる。必要性という観点から考えると、「論理的思考」「ノイズ耐性の向上」ということが考えられる。「論理的思考」は、情報量の減少による効率的な思考や、離散的な意志決定を可能とする。一方、われわれがノイズの多い環境でコミュニケーションをする際には、離散化は、ノイズの影響を減少させる効果を持つ。これは、まさに現在デジタルコンピュータが広く使われている理由の一つでもある。強化学習は、報酬という形で「必要」を表現し、それを実現するための学習であるため、これらの要因の存在が、強化学習で信号を離散化させることができる可能性を示すことになる。と考える。

次に、アーキテクチャという観点から考えると、「リカレントネット」の連想記憶、つまり、複数のアトラクタを有する固定点収束のダイナミクスが離散化を促進する一つの解として考えられる。リカレントネットで単に記憶を必要とするタスクの学習をさせるだけで、必要に応じて連想記憶の機能が形成されることは、すでに示されている [5] [6]。

本研究では、シンボル創発の研究における最初のステップとして「ノイズ耐性の向上」の必要性と「リカレントネット」の

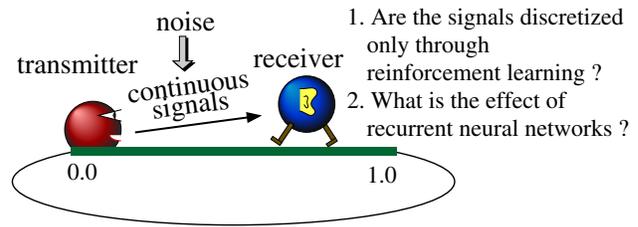


図 1 簡単な一方向コミュニケーションタスク

Fig. 1 A simple one-way communication task.

アーキテクチャに注目する。そして、単にノイズがある環境でコミュニケーションを学習するだけで、信号が離散化されるかどうか [7]、また、リカレントネットを導入して、離散化が促進されるかどうかを検証する。

3. コミュニケーションの学習とタスク

簡単なコミュニケーションを行う環境として、図 1 のように一次元の空間に 2 エージェントが置かれているとする。文献 [8] [9] を参照し、片方のエージェント (以下、受信者とする) は、相手の状態を観察し、アナログのコミュニケーション信号を送信するが、動くことはできないとし、もう片方のエージェント (以下、発信者とする) は、相手の状態を観察することはできないが、信号を受信して、動くことができるとする。そして、両者が接触した際に、両者とも同時に同じ報酬を得るとする。空間は一次元のリング状になっており、受信者は右に行っても左に行っても発信者と接触することができる。便宜上、発信者がいる位置を 0 とし、図のように、受信者の位置は、右に行くほど大きくなるように設定した。受信者の位置が 0.11 以下または 0.89 以上になったときに接触するとした。また、受信者の動作は、動作信号が 1 ステップでの移動距離を表すとし、絶対値の最大が 0.4 で、正の時に右に、負の時に左に行くとした。これによって、受信者は、真ん中より右にいる時は右に、左にいるときは左に行く最適な行動を取れば、1 ステップで発信者と接触することができる。しかし、両者の距離に対して動作信号が大き過ぎる場合には、行き過ぎてしまい、接触できないとした。

図 2 に、各エージェントの構成と信号の流れを示す。各エージェントは、それぞれ actor-critic アーキテクチャ [10] を一つのニューラルネットで構成し、通常の TD (Temporal Difference) 型の強化学習で別々に学習する。ニューラルネットとしては、前半のシミュレーションでは、通常の 3 層の階層型ニューラルネット、後半のシミュレーションでは、Elman 型のリカレントネットを用いた。発信者は、受信者の位置の情報を、まず

$$GS_i(dist) = \exp\left(-\frac{1}{2}(29dist - i)^2\right) \quad (1)$$

($dist$ は受信者の位置、 i は gaussian の番号で $i = 0, 1, \dots, 29$) の式で表される 30 個の gaussian で局所化してから、その 30 個の出力をニューラルネットに入力した。これによって、ニューラルネットでの計算で求められる非線形性を少なくし、受信者の位置に対するニューラルネットの出力の急激な変化を容易に

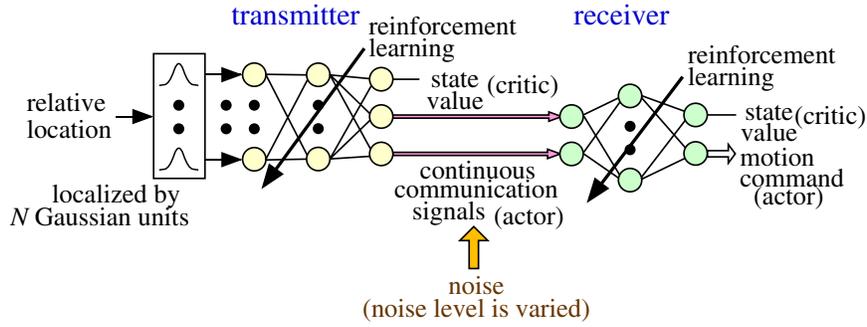


図 2 各エージェントの構成と信号の流れ

Fig.2 The architecture of each agent and the signal flow.

学習できるようにした。

学習は、まず TD 誤差を

$$\hat{r}_t = r_t + \gamma P_t - P_{t-1} \quad (2)$$

(r_t は報酬, P_t は critic の出力で, ニューラルネットの出力の一つを割り当てる. また, γ は割引率である.) と計算する. critic の出力に対しては, 教師信号を

$$P_{s,t-1} = P_{t-1} + \hat{r}_t = r_t + \gamma P_t. \quad (3)$$

と計算し, バックプロパゲーション法で教師あり学習をさせる. 出力ベクトル M は,

$$M_t = \alpha(2.5A_t + \text{rnd}_t + n_t) \quad (4)$$

(A は actor の出力ベクトルで, やはりニューラルネットの出力のうちのいくつかを割り当てる. rnd は, 強化学習で用いられる試行錯誤のための乱数ベクトルであり, n は, ノイズベクトルである. また, α は定数であり, 発信者は 1.0, 受信者は 0.4 とした.) と計算する. この出力は, 発信者にとっては, コミュニケーション信号であり, 受信者にとっては, 動作信号である. ここでは, コミュニケーション信号の数は 2 個とし, 動作信号の数は 1 個である. ただし, 動作信号にはノイズは乗せなかった. actor の出力に対しては, 教師信号を

$$A_{s,t-1} = A_{t-1} + 0.5\hat{r}_t \text{rnd}_{t-1} \quad (5)$$

と計算し, やはり教師あり学習をさせた.

中間層および出力層の各ニューロンの出力関数は, -0.5 から 0.5 の値域のシグモイド関数で, 教師信号は, 最大で 0.4, 最小で -0.4 とした. したがって, コミュニケーション信号は, 式 (4) より, ほぼ最大で 1.0, 最小で -1.0 となる. 試行およびノイズ成分をのせても 1.0 または -1.0 を越えないとした. また, 式 (2) で TD 誤差を求める際の critic の値は, 実際のニューラルネットの出力に 0.5 を足してから用いた. 接触した場合の報酬 r は, 0.9 とした. 中間層ニューロンの数は, 発信者, 受信者ともに 10 個とした. 試行ベクトル rnd の各成分は, 正負対称の一樣乱数を 3 乗した値を用い, 最大値を 0.4 とした. 一方, コミュニケーション信号に付加するノイズ n の各成分は一樣乱数とし, その大きさ (ノイズレベル) は 0.0 から 1.6 の間で変化させた. ただし, 試行成分も, 実際には受信者にとってはノイズ成分として働く.

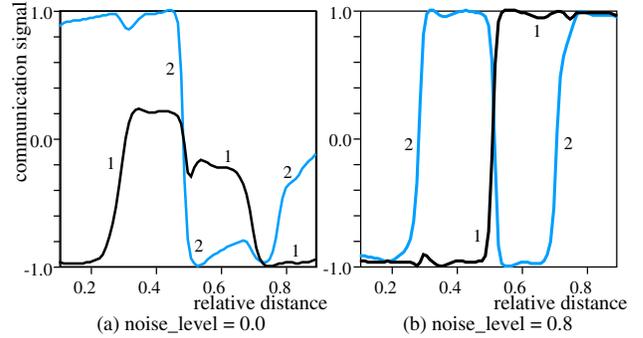


図 3 学習後に発信者が生成した受信者の位置に対するコミュニケーション信号

Fig.3 The communication signals generated by the transmitter as a function of the receiver's location.

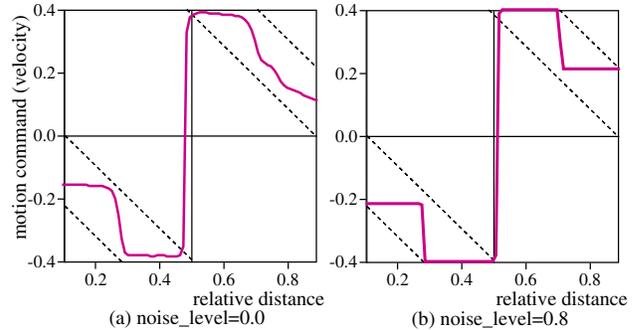


図 4 学習後に受信者が図 3 のコミュニケーション信号から生成した動作信号

Fig.4 The receiver's motion command generated from the communication signals as shown in Fig. 3 by the receiver as a function of the relative distance.

4. シミュレーション結果

図 3 に, 学習後の, 受信者の位置に対する発信者のコミュニケーション信号の値を, ノイズがない場合とある場合 (ノイズレベル 0.8) について示した. このとき, 学習係数は 5.0 で慣性項なしで 400000 試行の学習を行った. また, 同様に, 図 4 に, 受信者が図 3 のコミュニケーション信号から生成した動作信号を, やはりノイズがない場合とある場合について示した. 図中の斜めの破線は, その位置での最適動作信号の上限と下限を示

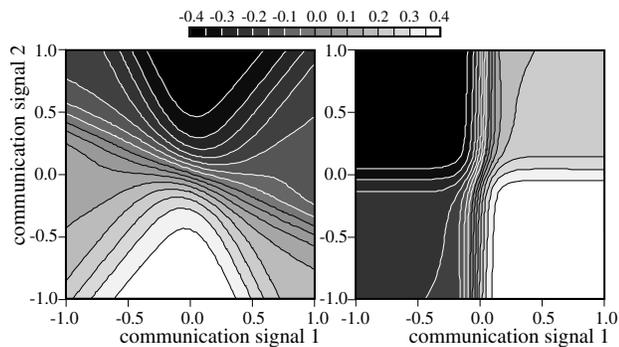


図 5 2つのコミュニケーション信号の値に対する受信者の動作信号
Fig. 5 The motion command as a function of the communication signals.

す．動作信号の絶対値が上限を越えると，受信者は送信者を通り越してしまい，下限に満たないと，受信者まで到達することができず，いずれの場合もその行動直後の報酬は得られない．

ノイズが付加されていない場合は，図 3(a)の信号 1 のようにコミュニケーション信号が中間的な値をとる傾向があり，信号の変化も滑らかであった．逆に，ノイズを付加することによって図 3(b)のように，信号が-1と1に二値化される傾向があった．動作信号を見ると，いずれの場合も，ほぼ最適値の範囲内に入っている．また，ノイズを付加しない場合(図 4(a))と付加した場合(図 4(b))を比較すると，付加した場合の方が動作信号を4つのレベルにきれいに離散化しており，コミュニケーション信号を二進数として解釈しようとしていることがわかる．また，付加していない場合でも，コミュニケーション信号，動作信号ともある程度離散化される傾向が見られる．これは，前述のように発信者の試行成分が受信者にとってノイズとしてとらえられることによる部分もあると考えられる．

さらに，受信者におけるノイズの効果を調べるために，図 5 に，二つのコミュニケーション信号に対する動作信号の値を色で表した図を，ノイズが付加されていない場合とされた場合について示す．ノイズがない場合(図 5(a))には，コミュニケーション信号に対して，動作信号が滑らかに変化しており，ノイズが付加されると(図 5(b))，コミュニケーション信号を4個の領域に区分し，その境界では，動作信号が急激に変わるようになっていることがわかる．これより，ノイズを付加することにより，単に発信者がコミュニケーション信号を二値化して表現するだけでなく，受信者も動作信号を離散化しており，受け取ったコミュニケーション信号を二値化された信号として解釈しようとしていると考えられることができる．

次に，ノイズの大きさを変化させたときの離散化の度合いの変化を調べた．まず，離散化の度合いに関して，二値化指数を

$$bin = \sum_i^{N_c} \sum_j^{N_d} |com_{i,j}| / (N_c \cdot N_d) \quad (6)$$

(N_c はコミュニケーション信号の数， N_d は受信者の位置のサンプル数， com は，試行成分およびノイズ成分を除いたコミュニケーション信号の値である．)と定義した．この値は，出力

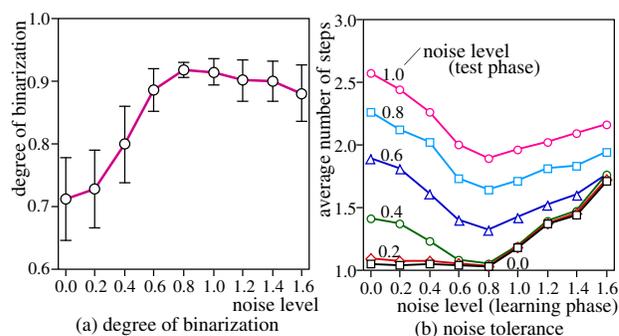


図 6 (a) ノイズの付加によるコミュニケーション信号の二値化指数
(b) 学習後のノイズの大きさによる到達ステップ数の変化
Fig. 6 (a) The degree of binarization according to the noise level.
(b) The noise tolerance according to the noise level in the learning phase.

が，受信者の位置によらず常に 1.0 または -1.0 であれば 1.0 で最大値となり，出力が常に 0.0 の場合に最小値 0.0 となる．この指標は，二値化以外の離散化の程度を調べる指標にはなっていないが，コミュニケーション信号に対しては，二値化以外の離散化はあまり観察されなかったので，二値化指数を用いて離散化を評価した．学習中のノイズレベルに対する二値化指数の値を図 6(a) に示す．この図は，同一ノイズレベルで乱数系列を変化させて 50 回の学習を行い，各ノイズレベルでの平均と標準偏差を表している．ちなみに，図 3 の場合の二値化指数は (a)(b) の場合それぞれ 0.69, 0.92 である．また，学習後の状態がどれだけノイズに強いかを調べるために，学習後のテスト時のノイズレベルを 0.0 から 1.0 まで変化させて接触までのステップ数を調べた．テスト時のノイズレベルごとに，学習中のノイズレベルに対する接触までのステップ数をプロットしたものを図 6(b) に示す．

まず，二値化指数は，学習時のノイズが大きくなるほど大きくなる傾向があり，同時に標準偏差も小さくなっていることがわかる．しかし，ノイズレベルが 0.8 付近からは逆に少しずつ指数が減少し，標準偏差も増大する傾向にある．次に，ノイズ耐性を見ると，学習後のテスト時でのノイズレベルが小さいときは，学習中のノイズレベルによってあまり性能は変化しないが，テスト時のノイズレベルを上げていくと，学習時のノイズレベルが高くなるにしたがって性能が向上していることがわかる．このことから，強化学習によって，ノイズの影響を減少させるために信号の二値化が行われ，その結果，ノイズ耐性も向上したと考えられる．しかしながら，この場合も，学習時のノイズレベルが 0.8 を越えると逆に性能が悪くなっていることがわかる．この場合，テスト時のノイズがない場合でも，性能低下が見られることから，これは，学習自体がうまくいっていないと想定される．また，ノイズレベルが 1.0 以上になると，どんなに信号を二値化してもノイズによって信号が反転してしまうということから考えても，学習がうまくいっていないのかと推測できる．

次に，図 3 で示した場合の実際のコミュニケーション信号において，学習とともに，二値化がどのように進んでいるかを図

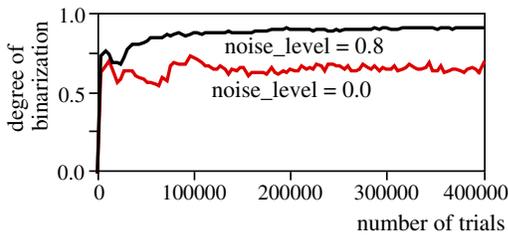


図 7 学習による二値化指数の変化

Fig. 7 Change of the degree of binarization.

7 に示す．ノイズが付加されている場合とされていない場合で，学習の初期の段階で差がつき，ノイズが付加されている場合は，その後も徐々に二値化指数が大きくなっていることがわかる．実際のコミュニケーション信号の変化を見たところ，最初の 40000 試行程度でほぼ信号の変化の傾向が決まり，その後，各信号が 1.0 や -1.0 に貼り付いていく様子が観察できた．

5. リカレントネットの導入

次に，図 2 の階層型ニューラルネットを Elamn ネットに置き換えた場合について示す．この場合，まず，発信者においては gaussian の出力が，受信者においてはコミュニケーション信号がニューラルネットに入力される．この際，予め中間層の値は 0 にリセットしておき，フィードバック部分の入力はすべて 0 とした．そして，中間層の値を計算した後，今度は外部入力部分を 0 にして，中間層からのフィードバックループ部分を T 回計算する．その後，その中間層の値から出力を求めた．この T を待ち時間と呼ぶ．中間層ニューロン数は 10 個で，前節と同じとした．

学習は，BPTT(Back Propagation Through Time) を用いて，待ち時間 T だけ時間をさかのぼって学習させた．また，フィードバック部分は，セルフフィードバックの重み値を 4.0 とし，その他を 0.0 とした．各ニューロンの出力関数の微係数の最大値は 0.25 であるため，セルフフィードバックの重み値 4.0 とかけるとちょうど 1.0 になる．したがって，誤差の情報が発散せずに過去に効率的に伝達できると考えられるからである．また，セルフフィードバックが 4.0 よりも少しでも大きくなると，双安定となるため，連想記憶の機能を形成することが容易に学習できると考えられる．

さらに学習を効率的に行うため，全試行数 3000000 試行のうち，最初の 1000000 試行では $T = 0$ と固定し，次の 1000000 試行では， T を乱数で決める確率を徐々に増やし，最終的にすべての時に乱数で決定するようにし，最後の 1000000 試行では，乱数で決定する割合を徐々に減らし，その分， $T = 5$ とする確率を徐々に増やし，最後には，常に $T = 5$ になるようにスケジューリングした．また，比較のために，最初から最後まで常に $T = 0$ として，リカレントを用いないで学習した場合のシミュレーションも行った．

レベル 0.8 のノイズを付加した場合に，学習時に待ち時間 $T = 0$ としてリカレントを利用しなかった場合とリカレントで学習した場合のそれぞれについて，テスト時に $T = 0$ の場合と

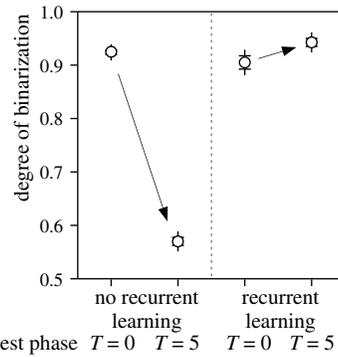


図 8 リカレントネットを導入した場合の二値化指数

Fig. 8 The degree of binarization when the recurrent neural networks are used.

$T = 5$ の場合のコミュニケーション信号の二値化指数 (式 (6) 参照) を調べた．その結果を図 8 に示す．また，それぞれの場合のコミュニケーション信号のサンプルを図 9 に示す．この図からわかるように，リカレントの学習をしていない場合は，前節の結果と同様に，ある程度の二値化がされているが，待ち時間 T を 5 とすると，図 9(b) のように，中間的な値をとるようになる．この時，フィードバック部分の結合の重み値は，初期値のまま，セルフフィードバック部分のみ 4.0 であるため，フィードバック部分が双安定とならず，徐々に絶対値が小さくなっていると考えられる．一方，リカレントの学習をした場合は， $T = 0$ での二値化指数は，リカレントの学習をしない場合よりむしろ小さいが， $T = 5$ の時点では，図 9(d) を見てもわかるように，二値化が進んでいることがわかる．これより，リカレントネットの導入により，さらに二値化が促進されていることがわかる．

次に，前節同様，受信側の離散化についても調べた結果をリカレントの学習をした場合のみ， $T = 0$ の場合と $T = 5$ の場合について図 10 に示す．この図を図 5 と比較すると， $T = 5$ では，4 つの動作信号レベルの境界が図 5 よりさらにはっきりしており，受信側でも，リカレントネットを導入することによって，動作信号の離散化が促進されていることがわかる．

6. 考 察

最後に，なぜ単に強化学習で学習するだけで，二値化が促進されるのかについて考えてみる．2. でも少し述べたように，強化学習はより報酬を得るような学習，つまり「必要」を実現するための学習である．したがって，この場合，信号を二値化することにより，ノイズの影響を減少させ，結果的に報酬を得る確率が大きくなったと考えることができる．しかし，この場合，発信者も受信者もともに学習しており，たとえば発信者が二値化をしても，受信者が二値化されたことを何らかの形で理解しなければ学習は進まないと考えられる．

そこで，もう少し細かく理由を考えてみる．受信側は，より良い行動を得るために，コミュニケーション信号の差を拡大し，出力に反映させようという学習が進む．そうすると，重み値の

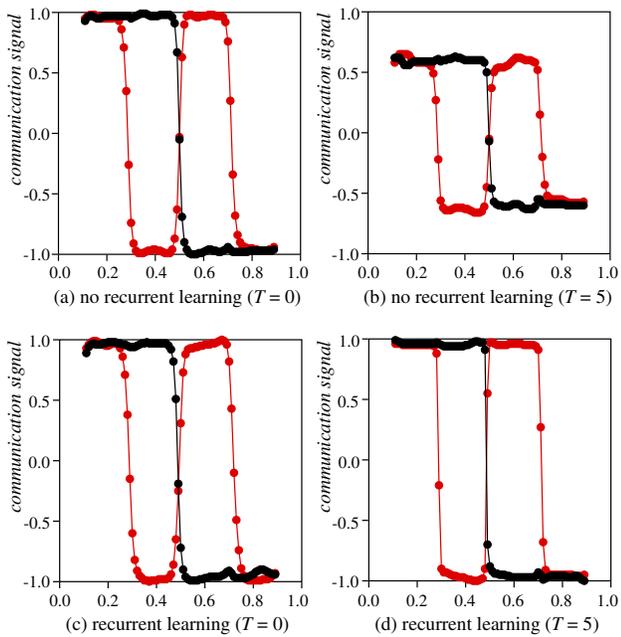


図9 リカレントネットを導入した場合のコミュニケーション信号。それぞれの二値化指数は，(a)0.92, (b)0.57, (c)0.91, (d)0.94。

Fig. 9 The output when recurrent neural network was used. The degree of binarization is 0.92, 0.57, 0.91, 0.94 for each of (a), (b), (c), (d) respectively.

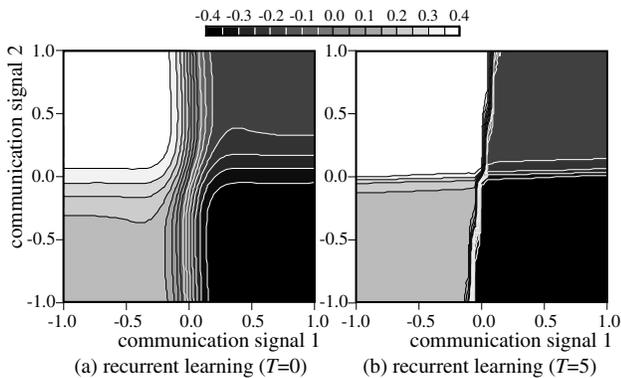


図10 リカレントネットを用いた場合のコミュニケーション信号に対する動作信号

Fig. 10 The motion command as a function of the two communication signals when the RNN was used.

絶対値が全体的に大きくなり，中間層ニューロンがシグモイド関数の飽和領域を用いて情報を表現するようになる．すると，飽和領域では，ノイズの影響が小さくなるため，発信側も中間的な値を取らないように学習が進む．そうすると，今度は受信側もノイズによる影響を小さくするために中間的な値を取らないように学習が進む，両者が一体となってコミュニケーション信号と動作信号の離散化が進むのではないかと考えられる．しかし，このメカニズムを完全に解明するためには，今後さらなる検証が必要である．

7. ま と め

ニューラルネットで生成したアナログコミュニケーション信号を単に強化学習で学習させるだけで，コミュニケーション信

号に付加するノイズの大きさとともに，信号が離散化することがわかった．また，受信側も，受け取った信号をさらに離散化し，行動に反映させることもわかった．さらに，リカレントネットを導入することで，信号の離散化がさらに進むことがわかった．

筆者は，現在まで，センサ信号を直接ニューラルネットに入力し，その出力をアクチュエータに送り，そのニューラルネットを強化学習に基づいて学習させることによって，認識，記憶を含むさまざまな機能を自律的，目的論的，調和的に獲得することを提唱してきた．そして，主に行動のための学習と言われてきた強化学習が，あらゆる機能の学習に有効であることを示してきた [11]．本研究もその一環であり，最も困難と考えてきたシンボル処理に対する研究の第一歩であるが，ニューラルネットによるシンボル創発への可能性を示せたことは，この研究の流れを確固たるものとする意義深いことであると考えている．本研究が，高次知能実現の新たな道を切り拓くことを期待する．今後は「論理的思考」の必要性からシンボルが発現するという問題についても考えていきたい．

謝辞 本研究の一部は，日本学術振興会科学技術研究費 基盤研究 (B) #13780295 および #15300064 の補助を受けた．ここに謝意を表する．

文 献

- [1] R. W. Sperry: "Hemisphere disconnection and unity in conscious awareness", *American Psychologist*, **23**, pp. 723-733 (1968).
- [2] S. Hanard: "Symbol grounding problem", *Physica D*, **42**, p. 335 (1990).
- [3] J. L. Elman: "Finding structure in time", *Cognitive Science*, **14**, pp. 179-211 (1990).
- [4] T. Omori, A. Mochizuki, K. Mizutani and M. Nishizaki: "Emergence of symbolic behavior from brain like memory with dynamic pattern", *Neural Networks*, **12**, 7-8, pp. 1157-1172 (1999).
- [5] D. Zipser: "Recurrent network model of the neural mechanism of short-term memory", *Neural Computation*, **3**, pp. 179-193 (1991).
- [6] K. Shibata and M. Sugisaka: "Dynamics of a recurrent neural network acquired through the learning of a context-based attention task", *Proceedings of the 7th International Symposium on Artificial Life and Robotics (AROB)*, Vol. 1, pp. 152-155 (2002).
- [7] K. Shibata and M. Nakanishi: "Discretization of analog communication signals by noise addition in communication learning", *Proceedings of the 9th International Symposium on Artificial Life and Robotics (AROB)*, Vol. 2, pp. 351-354 (2004).
- [8] G. M. Werner and M. G. Dyer: "Evolution of communication in artificial organisms", In *Artificial Life II*, pp. 659-687 (1991).
- [9] N. Ono, T. Ohira and A. T. Rahmani: "Emergent organization of interspecies communication in q-learning artificial organs", *Advances in Artificial Life*, pp. 396-405 (1995).
- [10] A. G. Barto, R. S. Sutton and C. W. Anderson: "Neuronlike adaptive elements that can solve difficult learning", *IEEE Transaction on Systems, Man and Cybernetics*, **13**, pp. 835-846 (1983).
- [11] 柴田：「強化学習とロボットの知能 -あめとむちで知能は作れるか? -」, 第16回人工知能学会全国大会論文集, パネルディスカッション「強化学習とその諸相」原稿, 2A1-05 (2002).