ニューラルネットワークを使った強化学習による行動学習を通した音声認識機能の創発

大分大学　○江越正大　柴田克成

Emergence of speech recognition function through action learning by reinforcement learning using a neural network

Seidai Egoshi and Katsunari Shibata, Oita University

Abstract: In recent years, the ability of speech recognition has been greatly improved by introducing Deep Learning. However, we think the present speech recognition has two essential problems. First, since recognition is learned separately from the other functions such as action generation, the learner cannot understand the meaning of the recognized words. Secondly, it is hard to provide the teacher signals for a lot of words in advance. In order to solve these problems, we propose to introduce the "End-to-End Reinforcement Learning" approach in which voice data are the input of a neural network and its action output is learned by reinforcement learning. In a simple simulation, a robot was given a voice to instruct appropriate one of 4 possible actions and learned to generate an appropriate continuous-valued action for the voice only from the reward at the goal. After learning, the robot could approach the goal according to the voice guidance. That means that the robot could recognize the voices correctly.

1　序論

近年、深層学習 (Deep Learning) を用いたパターン認識が従来手法を凌ぐ成果を出して [1]、注目を集めている。今後 AI 技術がさらに進むことによって、ロボットはこれまでのように工場で単純作業を行うだけでなく、例えば介護のような、人間と関わる場での活躍が期待される。このような状況では、ロボットのコミュニケーション能力の重要性が増大すると考えられる。

コミュニケーションの中で重要な役割を果たす音声認識でも、深層学習とリカレントネットを用いることで認識精度は大きく向上している [2]。しかし、高いコミュニケーション能力の実現という観点から、現在の音声認識には 2 つの根本的な問題点があると筆者らは考えている。1 つ目は、認識がロボットの行動生成などと分離して学習されている点である。ロボットは認識の学習時に行動を考慮していないため、意味の理解を伴った音声認識を行うことができない。2 つ目は、どのように教師信号を生成するかという点である。一般的なパターン認識では、われわれが使う全ての単語に対して出力を予め用意して教師信号を与える。しかし、我々は日常のやり取りの中で音声認識の学習をすることができるし、そもそも、あらゆる単語に対応した認識ニューロンを予め設けることも柔軟性という観点から得策ではないと考えられる。

これらの問題を解決するために、本論文では、汎用人工知能実現に向けた End-to-End 強化学習のアプローチ [3] を音声認識にも導入することを提案する。人間の脳のように、並列で柔軟な処理を行うことができるニューラルネットによって、センサからモータまでの全ての処理を構成し、強化学習を行う。ニューラルネットと強化学習を組み合わせることで、ニューラルネット内部に適切な行動を実現するための様々な機能を、経験から自律的に獲得することができる。このことは、我々の研究室や、DeepMind によって、すでにいくつかの例が示されている [3][4]。さらに、我々の研究室ではリカレントネットを導入することで、「記憶」や「予測」といった時間的な処理を必要とする機能の創発にも成功しており [3]、時系列データの学習が必要である音声認識についても同様な枠組みで学習できる可能性がある。これによって、ロボットも日常生活のやり取りの中で言葉の意味の理解を伴った音声認識を学習によって獲得できるのではないかと期待される。そこで、本研究では、ニューラルネットを使った強化学習を用いて、音声を入力として簡単な行動学習を行うことにより音声認識機能が創発する可能性を検討する。

2　研究方法

2.1　強化学習と音声認識

強化学習は報酬と罰から学習する手法で、ロボット (エージェント) は試行錯誤を繰り返して、罰を避け、報酬を受け取るような行動の学習を行う。この行動を通して、エージェントは行動生成に必要となる内部表現を獲得し、ゴールへ到達できるようになる。例えば TV ゲームの学習のように、画像入力を使って強化学習を行った場合、ネットワーク内部で必要となる画像認識の機能を獲得する。同様に、リカレントネットを用いて時系列へと拡張し、音声入力を使って強化学習を行った場合、音声認識機能を獲得すると考えられる。このとき行う音声認識は、行動生成のために必要に応じて自律的に形成される中間的な内部表現として獲得され、事前に認識出力

を規定する必要はなく、かつ音声認識と行動出力が繋がることで意味の理解を伴った音声認識を行うことが期待される。

本論文では、ネットワークは階層構造のネットワークにフィードバック結合を有する層を設けたリカレントネットを用いた。これによって、時系列データを入力とした学習を行う。ネットワーク内のニューロンの内部状態 $u_{j,t}^{(l)}$ は

$$u_{j,t}^{(l)} = \sum_{i=1}^{N^{(l-1)}} w_{j,i}^{(l)} o_{i,t}^{(l-1)} \left(+ \sum_{i=1}^{N^{(l)}} w_{FB}{}_{j,i}^{(l)} o_{i,t-1}^{(l)} \right) \quad (1)$$

で求められる。ここで、$N^{(l)}$ は l 層のニューロン数、$w_{j,i}^{(l)}$ は l 層 j 番目ニューロンの i 番目の重み値、$o_{i,t}^{(l)}$ は l 層 i 番目ニューロンの出力である。右辺第 2 項のカッコ内はフィードバック結合のある層で使用する。また、各ニューロンの出力関数はシグモイド関数から 0.5 を引いたもの、

$$o_{j,t}^{(l)} = \frac{1}{(1 + e^{-u_{j,t}^{(l)}})} - 0.5 \quad (2)$$

を使用した。

強化学習はロボットが実際に動くことを想定し、連続値を扱うことのできる Actor-Critic を使用する。Actor-Critic では、エージェントは時刻 t におけるニューラルネットへセンサ信号 $\mathbf{S}_t$ を入力とし、2 つの要素を持つ動作出力ベクトル $\mathbf{A}_t$ で移動し、乱数ベクトル $\mathbf{R}_t$ を加えて探索を行う。乱数 $\mathbf{R}_t$ の各要素は $[-0.5, 0.5]$ の値域を持つ。エージェントはゴールへ到達すると報酬 r_t を受け取る。学習は次式の TD 誤差 $\hat{r}_t$ を用いる。

$$\hat{r}_t = r_{t+1} + \gamma V_{t+1} - V_t \quad (3)$$

ここで、r_{t+1} は時刻 $t+1$ で与えられる報酬、V_t は時刻 t のネットワーク出力である状態評価値、γ は割引率で 0.9 に設定した。教師信号は、動作出力 (actor) 部 $\mathbf{A}_t$、評価出力 (critic) 部 V_t でそれぞれ

$$\mathbf{T}_{A_t} = \mathbf{A}_t + \alpha \hat{r}_t \mathbf{R}_t \quad (4)$$

$$T_{V_t} = V_t + \hat{r}_t = r_{t+1} + \gamma V_{t+1} \quad (5)$$

で求め、各教師信号が-0.4 から 0.4 に入らない場合は-0.4 または 0.4 とした。ここで、α は学習率で 0.5 に設定した。重み値は BPTT(Back Propagation Through Time) 法 [5] により、次式の最急勾配法に基づいて音声データの最初の時間まで遡って更新する。

$$\Delta w_{j,i}^{(l)} = -\eta \frac{\partial E}{\partial w_{j,i}^{(l)}} \quad (6)$$

ここで、η は学習係数、E は $E = \sum_{j=1}^{N^{(L)}} \frac{1}{2}(T_j - o_j^{(L)})^2$ で求められる 2 乗誤差である。

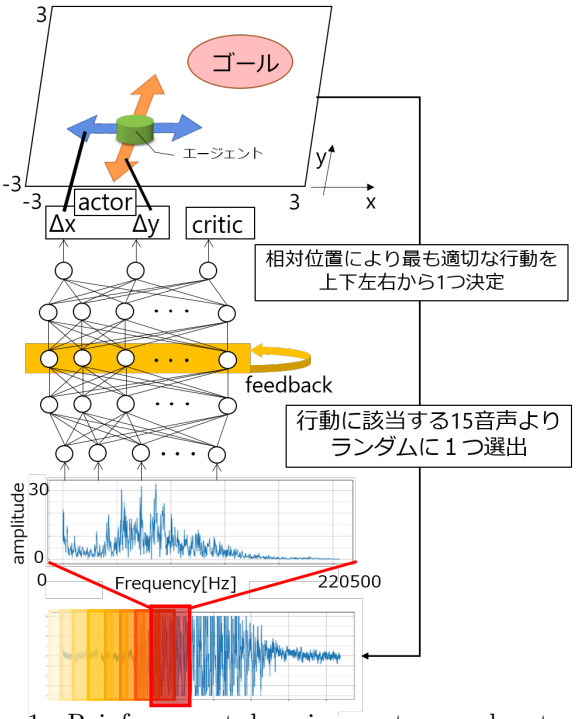
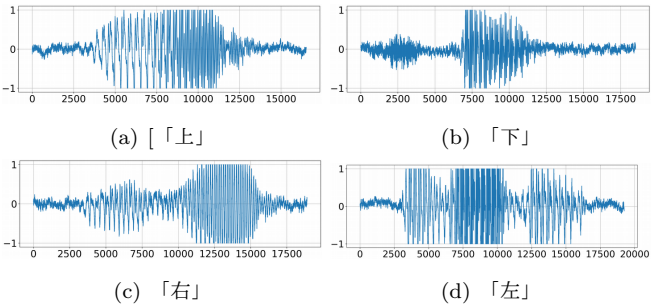


Fig. 1 Reinforcement learning system and network configuration used in this paper.



(a) [「上」 (b) 「下」

(c) 「右」 (d) 「左」

Fig. 2 An example of voice waveform for each instructed direction.

2.2 シミュレーション

ここでは、センサを持たないエージェントに対して、取るべき行動を音声で伝え、エージェントがゴールしたときに報酬を与えるというタスクを強化学習によって学習させた。そして、与えた音声に相当する行動をエージェントが行うことによって音声認識ができるようになったと判断する。シミュレーションは Fig.1 のように 6×6 の領域を用意し、ゴールとスタートの位置を毎回ランダムに決定した。エージェントは大きさを持たず、ゴールの半径は 0.5 とした。エージェントはネットワーク出力に乱数を加えた値によって移動し、これを 1 ステップとする。エージェントは、その中心がゴールに到達することで、0.4 の報酬を受け取る。各試行の終了条件はエージェントがゴールに到達し、報酬を受け取るか、移動回数の上限である 300 ステップに到達することである。ネットワークの重み値は毎ステップ更新を行い、試行回数は 50,000 回とした。

Table 1　The parameters used in the simulation.

ネットワーク層数		5
ニューロン数		1024-100-100-100-3
出力関数の値域		[-0.5,0.5]
初期重み値	出力層	[0.0] の一様乱数
	中間層	[-1,1] の一様乱数
	中間層 self FB 部	4.0
学習係数 η	中間層 FW 部及び出力層	1.0
	中間層 FB 部	0.1

外から人間がエージェントに対して指示を出していると仮定し、音声データを入力した。音声データは、成人男性 3 人が「上」「下」「右」「左」をそれぞれ 5 回発声して録音した、1 人 20 個の音声、計 60 個のデータを使用した。録音時のサンプリング周波数は 44.1kHz で、これらの録音した音声から、無音部を取り除き、人が聞いて意味を理解できる部分のみを残した。これによって、取り出された音声は Fig.2 のようになり、音声の長さは短いもので 0.26 秒、長いもので 0.59 秒程度になった。

エージェントへ入力する音声は、エージェントから見たゴールの相対位置を使って決定した。相対位置は Fig.4 のように、上下左右を斜めに 90° ずつ分けた 4 つの領域 ① ～ ④ で分類し、① の場合は「下」、② の場合は「上」、③ の場合は「左」、④ の場合は「右」に相当する音声を 15 個から 1 つランダムに選択する。選ばれた音声は 2048 フレーム毎に hamming 窓を適用し、フーリエ変換を行った。これによって得られた 1024 個の、0 ～ 22,050Hz までである周波数データを入力とした。フーリエ変換は 512 フレームずつ先の時間にシフトしており、19 ～ 47 ステップ分の入力となる。エージェントの移動距離は入力の最終ステップでの行動出力 $\mathbf{A}_t$ に乱数 $\mathbf{R}_t$ を加えたものとした。

シミュレーションに用いたパラメータを Table 1 に示す。ネットワークは開発に Preferred Networks, Inc. の Chainer を用いた。入力が時系列データであるため、フィードバック結合を有するリカレントネットを使用し、中間層を 3 層とし、まん中の 2 層目にフィードバック結合を設けることで、入力からフィードバック、フィードバックから出力の両方で非線形演算ができるようにした。学習係数は、10,000 試行毎に半分に減らした。出力層の初期重み値は初期値による影響を抑えるため、0 に設定した。また、中間層 2 層目のフィードバック結合では、自分自身への結合であるセルフフィードバック部の重み値を、自身の出力の値を保持し、かつ適切に誤差伝搬するように 4.0 に設定した。

3　結果

50,000 試行の学習をした際の学習曲線をゴール到達までのステップ数の変化として Fig.3 に示す。Fig.3 を見ると、10,000 試行程度までステップ数が減少し、その後あまり大きな変化は見られず、学習が完了しているように見える。10,000 試行後と 50,000 試行後のエージェントの軌跡を Fig.4 に示す。ゴールは中心に固定され、中心から半径 2 の円上の 12 個の点からエージェントはスタートし、乱数を加えずに行動させた。10,000 試行後を見ると、斜めに動く軌道も見られるが、いずれの場合もゴールに到達している。50,000 試行後は各領域での相当する音声指令の方向にほぼ動いており、領域の境界部分では、音声の切り替わりに対応してエージェントが境界に沿って進んでいることがわかる。

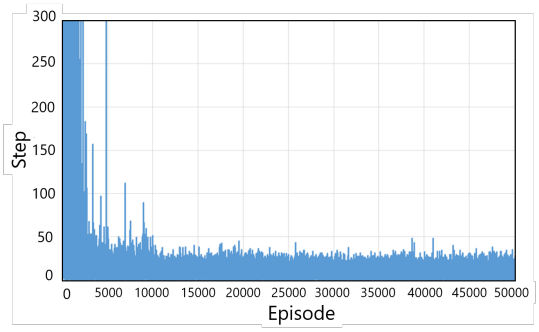
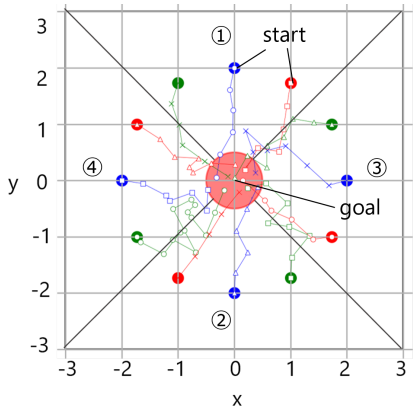
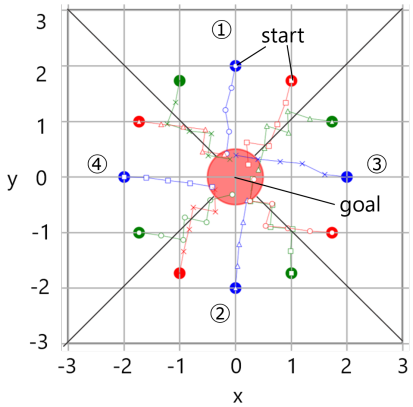


Fig. 3　Learning curve (Number of steps to reach the goal).

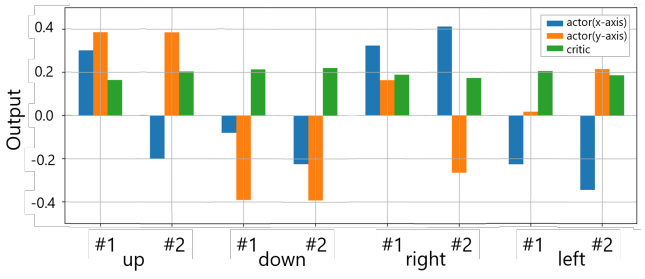


(a)after 10,000 episodes

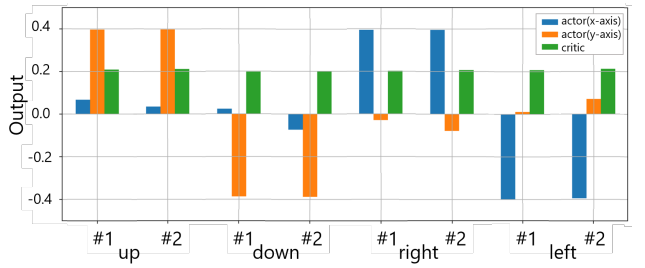


(b)after 50,000 episodes

Fig. 4　Agent trajectories: 12 locations are on a circle with a radius of 2.



(a)after 10000 episodes



(b)after 50000 episodes

Fig. 5　The network outputs(1 critic, 2 actors) for two of the voice data for each instructed moving direction.

10,000 試行後、50,000 試行後について、例として上下左右の音声をそれぞれ 2 つ選んで入力した際の出力を Fig.5 に示す。教師信号が-0.4 から 0.4 の範囲で与えられるため、各音声に与えられる理想出力は Table 2 となる。しかし、Fig.5(a) の 10,000 試行目を見ると、理想的な出力との差がある程度大きい。これに対して Fig.5(b) の 50,000 試行後は、理想的な出力との差が小さくなっている。Fig.4(a) で、エージェントが軸方向に沿って直進せずに、斜めに進む場合が多いのは、この理想的な出力との差によるものであることがわかる。

Table 2 の x と y の 2 出力に対する理想出力を使用して、各動作指示に対する全 15 音声データについての 1 出力あたりの平均 2 乗誤差 (Root Mean Square Error) の学習による変化を求めた結果を Fig.6 に示す。誤差は 10,000 試行以降も残っていて、徐々に小さくなっていることがわかる。

本論文では結果を示していないが、強化学習を使用せず、教師あり学習を行った場合、1,000 試行程度で学習を行うことができた。この結果と比較すると、Fig.6 ではかなり学習に時間がかかっているが、これは教師信号を直接与えずに試行錯誤に基づいて学習していることが大きな理由と考えられる。

4　結論

本研究では、ニューラルネットを使った End-to-End 強化学習によって行動学習を通した音声認識を行った。非常に簡単なタスクではあるが、音声からロボットの行動を直接出力し報酬のみから学習することで、予め認識出力を規定することなく、適切な行動ができるようになった。また、適切な行動生成の必要性から与えられた音声の認識を獲得することができたことにより、意味の理解を伴った学習の可能性を示したと考えている。

Table 2　Ideal outputs used for error computation.

	上	下	右	左
A_x	0.0	0.0	0.4	-0.4
A_y	0.4	-0.4	0.0	0.0

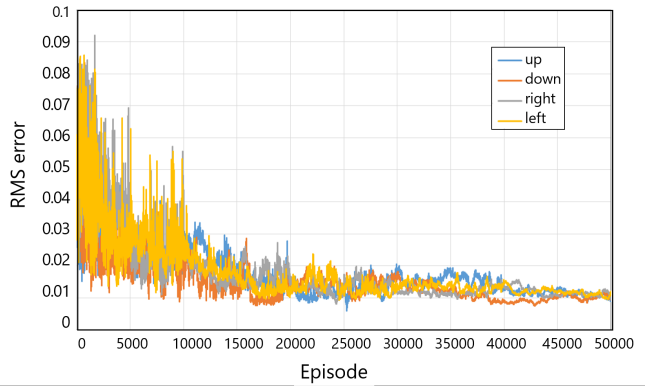


Fig. 6　Loss curve for each direction (RMS erroe from the ideal outputs).

本研究の問題点は 2 つある。1 つ目は入力が継続的に与えられていないという点である。実験に使用した音声は人為的に切り取られた音声であり、入力や出力のタイミングが決められている。人間の介入なしに自律的な学習を行うためには入力は継続的に与える必要があるだろう。2 つ目は状態評価 (critic) の学習が行えていない点である。音声入力は行動の指示のみで critic の出力に必要な情報が与えられおらず、critic が有効に働いていない。そのため、ゴールに到達した時の報酬のみからしか actor 部の学習を行っていないと考えられる点である。これらを改善していく必要がある。

謝辞

本論文は JSPS 科研費 (15K00360) の補助を受けた。

参考文献

1) A. Krizhevsky et al.: ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems 25(NIPS), 2012

2) A. Hannun et al.: Deep Speech 2: End-to-End Speech Recognition in English and Mandarin, Computation and Language (cs.CL),arXiv:1512.02595, 2015

3) 柴田克成, 後藤祐樹: 深層学習が示唆する end-to-end 強化学習に基づく機能創発アプローチの重要性と 思考の創発に向けたカオスニューラルネットを用いた新しい強化学習, 認知科学, Vol. 24, No. 1, pp. 96-117, 2017

4) V. Mnih et al.: Playing Atari with Deep Reinforcement Learning, NIPS Deep Learning Workshop 2013, 2013

5) D. E. Rumelhurt et al.: Learning internal representation by error propagation, Parallel Distributed Processing, Vol. 1, Chapter 8, pp. 318-362, MIT Press, 1986