

自己組織化を用いたリカレントニューラルネットワークの学習手法

池田成満, 高橋宏知

音声, 脳, 神経回路, リザーブ計算, 可塑性

1. はじめに

本研究では, 分散培養神経回路への特定パターンの繰り返し入力が, 分散培養神経回路のその特定パターンに対する識別能力を向上させる可能性を示した. また分散培養実験の結果を元に, リカレントニューラルネットワークによるモデルを提案した. 提案手法は話者識別タスクにおいて, リカレントニューラルネットワークである echo state network より高い性能を発揮し, 話者識別における state-of-the-art である i-vector/PLDA に匹敵する性能を示した. 特に, 話者の学習, 識別に使用できる音声データが少ない場合には提案手法が i-vector/PLDA を上回った. これは, 分散培養神経回路の刺激入力に対する可塑的な仕組みを模擬した人工ニューラルネットワークが実タスクにも活用できる可能性を示している.

2. 性別識別実験

本実験では, 神経細胞の分散培養系に対して, 特定の時空間パターンを繰り返し入力することで, 神経回路の自己組織的な変化を誘導する. この繰り返し刺激の前後において, 特定のパターン入力に対する誘発応答がどの程度識別可能かを定量化し, 繰り返し刺激の前後で識別率を比較する. 時空間パターンとして男女の日本語話者の音声を用いることで性別の識別タスクを行なった. 音声の変換には SREAK[1]を用いた. 繰り返し刺激には合計約 1 時間の男女の音声を用いている. また, 性別識別率は繰り返し刺激に用いた話者グループ (既知話者グループ) の繰り返し刺激に用いた発話 (既知発話) を用いた場合, 既知話者グループの繰り返し刺激に用いていない発話 (未知発話) を用いた場合, 繰り返し刺激に用いていない話者グループ (未知話者グループ) の未知発話を用いた場合で行なった. 男女の 2 クラス分類を leave-one-out で学習することで平均識別率を算出した. ただし, 話者の識別には線形回帰を用いた. 識別に用いる特徴量として, 刺激終了後からある時間遅れ後の, ビン幅 100 ms で算出した分散培養神経回路の発火率を用いた. 繰り返し前後における識別率が有意に変化したかを確認するためカイ二乗検定を行なった (有意水準 0.05).

図 1 に結果を抜粋して示す. 図にて, グレーの棒グラフは繰り返し刺激前の男女の平均識別率であり, 黒い棒グラフは繰り返し刺激後の男女の平均識別率を表している. 上段は既知話者グループにおける既知発話を用いた各時間遅れにおける平均識別率である. 刺激直後から近い時間領域にて識別率に有意な差が見られた. また, 中段は既知話者グループにおける未知発話を識別に用いた際

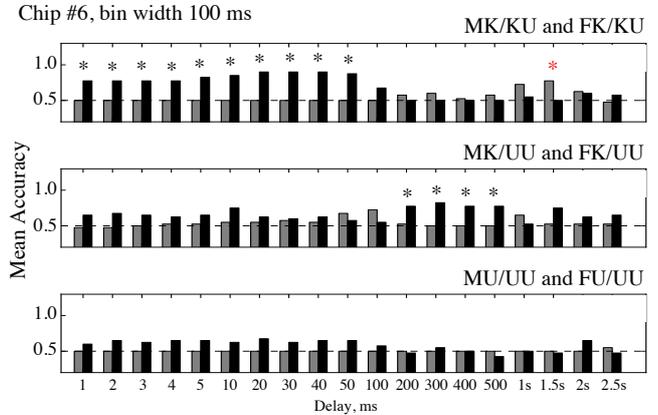


Fig. 1 Result of gender identification

の平均識別率である. 未知発話では, すこし時間がかかったあとで男女の識別ができるようになることを示唆している. 下段は未知話者グループの未知発話を用いた際の識別率であるが, こちらでは有意な差は得られなかった. この結果から, 特定パターンの繰り返し刺激で分散培養神経回路は特定パターンの識別能力を獲得することが示唆された. これは, 分散培養神経回路は特定パターン入力で自己組織的に変化し, その特定パターンの識別能力を向上させる可能性を示している.

時間遅れによる影響を見ると, 既知発話に対する平均識別率は 50 msec 以下のものでも有意に上昇している. 刺激の誘発応答は, 刺激後 20 msec までは刺激に直接反応した細胞による活動で, それ以降は活動が伝搬による活動だと考えられている [2]. 従って, 比較的小さい時間遅れにて識別率が上昇したことから, 刺激入力に対する自己組織化は刺激電極から近い領域に存在する神経細胞間で主に発生することが示唆される.

3. 分散培養神経回路のモデル化

前節の実験により, 分散培養神経回路は特定パターン入力で自己組織的に変化し, その特定パターンの識別能力を向上させる可能性を示している. そこで, 本節では分散培養神経回路の簡易的なモデル化を行い, 従来手法との識別率の比較を行なった.

まず, 前述の実験にて得られた結果は, 刺激に対する自己組織化は刺激チャンネル付近で主に発生することを示唆している. そこで, モデルを次のように構築する. 分散培養神経回路を, 1 層の中間層からなる RNN だと仮定する. すると, 刺激に対する自己組織化は刺激チャンネル付近, つまり入力層から中間層への結合荷重に影響を与えると言える. また, それ以外の結合には自己組織化は寄与しないと考えると, 中間層の結合荷重は固定だと考えられる. また, 中間層の活動を識別に用いるため, 中間層か

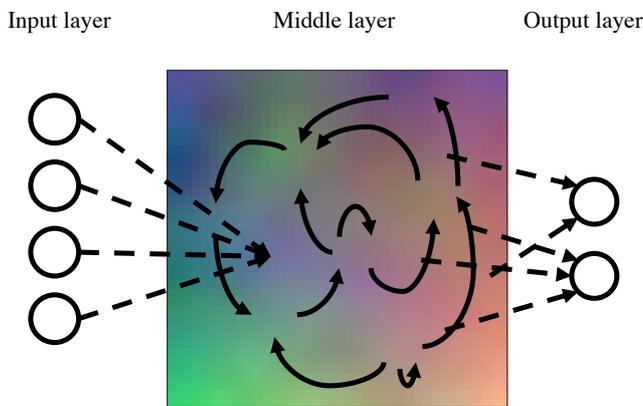


Fig. 2 Schematic illustration of proposed method

ら出力層への結合荷重をおき、それらを教師ありで学習することで入力に対する識別をおこなう。このモデルとして、self-organizing map[3]及び echo state network (ESN) [4]を用いた。図2に提案手法の図を示す。

識別タスクとして話者識別をおこなった。なお、100人の話者グループに対し識別タスクを行った。また比較手法には話者識別における state-of-the-art である i-vector/PLDA[5], 及び ESN を用いた。ESN は中間層のユニット数が 200 のものと 1600 のものを用いた(簡略化のため ESN200, ESN1600 とそれぞれ呼ぶ)。ただし、提案手法の中間層のユニット数は 1600 である。

図3に識別率の比較を示す。なお、白の射線が i-vector/PLDA, グレーの棒グラフが ESN200 であり、グレーの斜線が ESN1600, 黒の棒グラフが提案手法の識別率である。本実験で用いた話者学習用の発話は一人当たり 2.0 秒であり、識別に用いた音声は一人当たり 5.0, 2.0, 1.0, 0.5 秒である。図から明らかなように、提案手法は ESN200, ESN1600 よりも大幅に識別率が高い。なお、ESN1600 の識別率はチャンスレート(破線)よりも低くなっている。これは ESN1600 が過学習を起こしたためであると考えられる。一方で、提案手法は過学習した ESN と同数の中間層のユニット数を持つ。にも関わらず過学習を起こしていないのは、入力層から中間層への結合荷重が自己組織的に学習され、次元圧縮のような効果をもたらしたためである可能性がある。次に、i-vector/PLDA と比較する。テストに用いた発話が長い場合、i-vector/PLDA は提案手法よりも高い識別率を出せている。しかし、発話長が短くなる

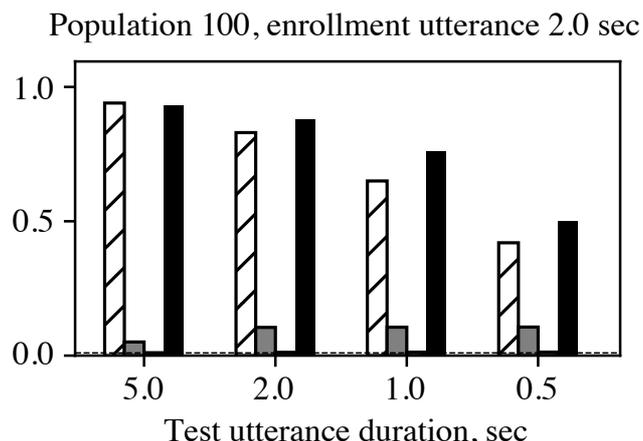


Fig. 3 Result of speaker identification

につれて提案手法が i-vector/PLDA を上回るようになることがわかる。これは、i-vector/PLDA は統計的な手法がベースにあるため、識別に利用できる発話長が短い場合、音響特徴量の十分な分布が得られずに、識別率が劣化することが原因と考えられる[6]。

この結果より、分散培養神経回路、及びその自己組織化様式のモデル化は工学的にも有用であることが示された。

4. 結論

本研究では、分散培養神経回路への特定パターンの繰り返し入力が、分散培養神経回路のその特定パターンに対する識別能力を向上させる可能性を示した。また分散培養実験の結果を元に、リカレントニューラルネットワークによるモデルを提案した。提案手法は話者識別タスクにおいて、リカレントニューラルネットワークである echo state network より高い性能を発揮し、話者識別における state-of-the-art である i-vector/PLDA に匹敵する性能を示した。特に、話者の学習、識別に使用できる音声データが少ない場合には提案手法が i-vector/PLDA を上回った。これは、分散培養神経回路の刺激入力に対する可塑性な仕組みを模擬した人工ニューラルネットワークが実タスクにも活用できる可能性を示している。

参考文献

- [1] Skinner, Margaret W., et al. "Identification of speech by cochlear implant recipients with the Multiplex (MPEAK) and Spectral Peak (SPEAK) speech coding strategies I. Vowels." *Ear and hearing* 17.3 (1996): 182-197.
- [2] Jimbo, Y., Kawana, A., Parodi, P., & Torre, V. (2000). The dynamics of a neuronal culture of dissociated cortical neurons of neonatal rats. *Biological cybernetics*, 83(1), 1-20.
- [3] Kohonen, Teuvo. "The self-organizing map." *Proceedings of the IEEE* 78.9 (1990): 1464-1480.
- [4] Jaeger, Herbert, and Harald Haas. "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication." *science* 304.5667 (2004): 78-80.
- [5] Dehak, Najim, et al. "Front-end factor analysis for speaker verification." *IEEE Transactions on Audio, Speech, and Language Processing* 19.4 (2011): 788-798.
- [6] Kanagasundaram, Ahilan, et al. "I-vector based speaker recognition on short utterances." *Proceedings of the 12th Annual Conference of the International Speech Communication Association*. International Speech Communication Association (ISCA), 2011.