

## 1. はじめに

人間の脳は役割によって処理する場所が異なる。これは機能分化と呼ばれ、脳に似た構造を持つネットワークである Echo State Network[1] でも機能分化が起こると考えられる。機能分化が起こると、ネットワークのニューロンは、空間情報や時間情報の分類に特化する。これにより、分類の性能向上に寄与すると考えられる。しかし、機能分化した際に必ず分類精度が向上するとは限らず、同じパラメータで実験を行っても結果にバラツキがあり、何度も実験を行う必要がある。そこで、本研究では機能分化を誘発する相互情報量の分布を考慮した損失関数を提案し、安定した機能分化と精度の向上を試みる。

## 2. Echo State Network

Echo State Network(ESN) は、入力層、Reservoir 層、出力層の 3 層で構成される。ESN の特徴は、Reservoir 層の接続と重みを隣接行列でランダムに決め、出力層の重みのみを学習する点である。本研究では、図 1 に示す Reservoir 層のニューロンを入力する側と出力する側に分けた拘束条件付き ESN を用いる。

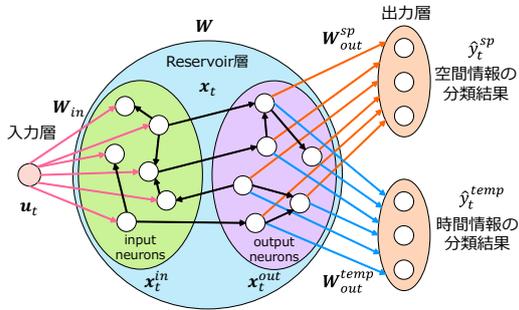


図 1: 拘束条件付き ESN

時刻  $t+1$  のニューロンの値  $x_{t+1}$  を式 (1) に示す。  $u_t$  は時刻  $t$  の入力データ、  $x_t$  は時刻  $t$  のニューロンの値、  $W_{in}$  は入力層の隣接行列、  $W$  は Reservoir 層の隣接行列、  $f(\cdot)$  は  $\tanh$  関数である。時刻  $t$  の空間情報のクラスの確率  $\hat{y}_t^{sp}$  は、式 (2) のように時刻  $t$  の output ニューロンの値  $x_t^{out}$  に出力層の重み  $W_{out}^{sp}$  を乗じ、  $g(\cdot)$  の softmax 関数に与えて求める。時間情報も同様に求める。

$$x_{t+1} = f(W_{in}u_t + Wx_t) \quad (1)$$

$$\hat{y}_t^{sp} = g(W_{out}^{sp}x_t^{out}) \quad (2)$$

## 3. 提案手法

安定して機能分化させるために、機能分化を定義し、定義した部分にニューロンがあると誤差が小さくなるような損失関数を設計する。

### 3.1. 機能分化の定義

機能分化した ESN の相互情報量の分布を図 2(a) に示す。このように時間情報と空間情報に寄与するニューロンが分離している。分離することで Reservoir 層内のニューロンが時間情報や空間情報の分類に特化すると考えられている。図 2(b) に示す赤枠にニューロンが多く集まると機能分化し、片方の情報に特化するニューロンや両方に寄与するニューロンが位置する。青枠のニューロンは寄与率が低いいため機能分化していないニューロンと言える。

### 3.2. 損失関数

図 2(c) の原点との距離で機能分化の傾向を計算できると考える。合計距離が大きい場合は、機能分化の傾向が大きくなり、小さい場合は機能分化の傾向が小さいとわかる。また、空間時間と時間情報に寄与するニューロンが分かれる場合やどちらにも寄与する場合も、分布に大きな偏りがあった場合は機能分化といえない。そのため、原点との距離を補助するために対角線からのバラツキも用いる。

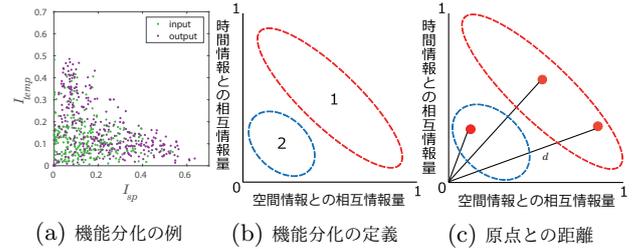


図 2: 機能分化の定義と指標

本実験では、式 (3) に示す対角線からのバラツキ  $L_{var}$  と原点との距離  $L_{dist}$  の項を損失関数に加え、相互情報量の分布を誘導する。  $\alpha$  と  $\beta$  により、スケールを調整する。

$$L = C_{sp} + C_{temp} + \alpha L_{var} + \beta L_{dist} \quad (3)$$

## 4. 評価実験

損失関数に対角線からのバラツキと原点との距離を追加し、相互情報量や精度にどのような変化があるか調査する。

### 4.1. 実験概要

入力データは、時間情報と空間情報を組み合わせたパターンであり、時間情報は周波数の異なる 3 種類の  $\cos$  波、空間情報は入力次元を 2, 4, 8 分割した 3 種類で表現する。学習に 12,000 step、評価に 10,000 step 用いる。最適化アルゴリズムには遺伝的アルゴリズムを用い、損失の少ないネットワークを探索する。

### 4.2. 実験結果

対角線からのバラツキと原点との距離を損失関数に追加した際の相互情報量と精度を図 3 に示す。図 3(a) に示す  $\alpha = -1, \beta = -0.1$  はバラツキと距離を大きくするように係数を決めており、ニューロンが空間情報と時間情報に分かれ、機能分化したことが確認できる。図 3(b) に示す  $\alpha = 1, \beta = -0.1$  はバラツキを小さく、距離を大きくするように係数を決めており、ニューロンが対角線付近に集まり、どちらの情報にも寄与している。しかし、どちらの係数も精度が低い。これは精度よりバラツキや距離の項が重視されているためであり、係数を変更して加える誤差を小さくする必要がある。図 3(c) に示す  $\alpha = -0.1, \beta = -0.01$  はパラメータ探索の結果、機能分化しつつ、精度が良かった際の分布である。

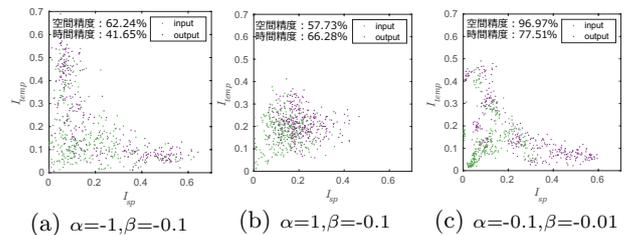


図 3: 相互情報量の分布と精度

## 5. おわりに

本稿では、損失関数に対角線からのバラツキや原点との距離を追加し、相互情報量の分布がどのように変化するか観察し、その時の精度を調査した。調査の結果、バラツキや距離を追加することで相互情報量の分布を誘導することができた。しかし、バラツキと距離の誤差が大きいと機能分化しても精度は低くなる。係数  $\alpha = -0.1, \beta = -0.01$  は機能分化しつつ、精度が他のパラメータより高い結果になった。しかし、この係数も高精度では無く、最適で無いため、今後は  $\alpha, \beta$  の最適な値を探索する。

## 参考文献

- [1] H.Jaeger, *et al.*, "Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication", Science, 2004.