

# リザーバー計算を用いた小規模言語モデル

## Small Language Model Using Reservoir Computing

藤巻侑暉

知能システムコース 1022194  
公立はこだて未来大学 香取研究室

January 28, 2026

# 目次

---

**1. 研究背景**

**2. モデル**

**3. 結果**

**4. まとめ・議論**

LLM の高性能化に伴い、学習効率の低さが課題として指摘されている  
モデル規模が年々増加している

- 学習に必要なデータ量
- モデルのパラメータ数
- 計算資源

# 背景

---

LLM の高性能化に伴い、学習効率の低さが課題として指摘されている  
モデル規模が年々増加している

- 学習に必要なデータ量
- モデルのパラメータ数
- 計算資源

## スケーリング則

学習に用いる計算量  $C$ , データサイズ  $D$ , パラメータ数  $N$  と, 検証データに対するクロスエントロピー損失との間にべき乗則が成立する現象

## データのクオリティを上げる

学習データの選別および合成データの活用によるデータ量が 100 倍以上のモデルに対して, コード生成タスクにおいて匹敵する性能を達成

# 目的

---

## データのクオリティを上げる

学習データの選別および合成データの活用によるデータ量が 100 倍以上のモデルに対して、コード生成タスクにおいて匹敵する性能を達成

## 関連研究：リザーバー計算を用いた言語モデル

---

### データのクオリティを上げる

学習データの選別および合成データの活用によるデータ量が 100 倍以上のモデルに対して、コード生成タスクにおいて匹敵する性能を達成

## 更新式

---

$$\begin{aligned} \mathbf{x}(t) &= (1 - \alpha)\mathbf{x}(t-1) \\ &+ \alpha \left( W^{\text{in}} \mathbf{d}(t) + W^{\text{rec}} \mathbf{r}(t-1) \right) \end{aligned} \quad (1)$$

$$\mathbf{r}(t) = \tanh(\mathbf{x}(t)) \quad (2)$$

$$\mathbf{y}(t) = W^{\text{out}} \mathbf{r}(t) \quad (3)$$

$$\mathbf{e}_k(t) = \mathbf{y}_{\text{true},k}(t) - \mathbf{y}_k(t) \quad (4)$$

$$W^{\text{out}}(t+1) = W^{\text{out}}(t) + \frac{\mu}{B} \sum_{k=1}^B \mathbf{e}_k(t) \mathbf{r}_k(t)^\top \quad (5)$$

$\mathbf{d}(t) \in \mathbb{R}^{N_{\text{in}}}$ : 時刻  $t$  の入力

$\alpha$  はリーク率

$W^{\text{in}} \in \mathbb{R}^{N_{\text{rec}} \times N_{\text{in}}}$ : 入力層からリザバー層への結合重み

$W^{\text{rec}} \in \mathbb{R}^{N_{\text{rec}} \times N_{\text{rec}}}$ : リザバー層内の再帰結合重みである

$N_{\text{in}}$ : 入力ベクトルの次元数

$N_{\text{rec}}$ : リザバー層のニューロン数

$\mathbf{y}(t) \in \mathbb{R}^{N_{\text{out}}}$ : 時刻  $t$  におけるリザバー出力



## 次単語予測精度 (**Perplexity**) の評価・考察

---

Table:  $\rho(W^{\text{rec}})$  および  $D_{\text{emb}}$  の変化に伴う Perplexity

$\rho(W^{\text{rec}})$	0.80	0.85	0.90	0.95	1.00	1.05
$D_{\text{emb}}$	平均 Perplexity (PPL)					
32	7849.77	<u>7849.08</u>	7850.66	7854.91	7861.96	7871.70
64	9226.47	<u>9223.86</u>	9223.99	9227.24	9233.84	9243.76
128	<u>10464.76</u>	10467.36	10471.31	10476.32	10482.18	10488.89

# 文法判断能力 (BLiMP) の評価・考察

Table:  $\rho(W^{\text{rec}})$  と  $D_{\text{emb}}$  の変化に伴う BLiMP 精度

スペクトル半径 $\rho(W^{\text{rec}})$	0.80	0.85	0.90	0.95	1.00	1.05
次元数 $D_{\text{emb}}$	精度 (Macro Average)					
32	55.68%	55.57%	55.48%	55.49%	<u>55.50%</u>	<u>55.53%</u>
64	<u>56.05%</u>	<u>55.97%</u>	<u>55.77%</u>	<u>55.51%</u>	55.26%	54.93%
128	55.07%	55.01%	54.92	54.82	54.70%	54.66%

## まとめ

---

一般の時系列処理を行う RC と別の挙動を示す

Sample text

入力するベクトルの次元数は挙げ続けても性能は上がらない

Sample text in red box

Example

次元数 32 でスペクトル半径が 1 以上のモデルは計算量を大幅に落としながらも性能を落としづらい (頑健性を持っている)

# Theorem

---

Theorem (Mass–energy equivalence)

$$E = mc^2$$

# Figure

---

Uncomment the code on this slide to include your own image from the same directory as the template .TeX file.

# Citation

---

An example of the `\cite` command to cite within the presentation:

This statement requires citation [?].

# References

---