

系列信号の長期的依存関係を学習するための自己・相互注意機構を用いたリザーバーコンピューティングの提案

小西 文昂[†] 廣瀬 明[†] 夏秋 嶺[†]

東京大学大学院工学系研究科電気系工学専攻[†]

1 まえがき

ニューラル機械翻訳(Neural Machine Translation: NMT)において著しい性能向上と学習時間の短縮に成功した Transformer [1] は、近年、時系列データの予測など様々な系列データに対して適用されている。本論文は、再帰型ニューラルネットワーク(Recurrent Neural Networks: RNNs) の一種であるリザーバーコンピューティング (Reservoir Computing: RC) [2] に、Transformer の重要な構成要素になっている自己注意機構と相互注意機構を組み込むことによって、系列信号フィルタリングのための長期的な依存関係を捉えることができるモデルを提案する。さらに、われわれは数値シミュレーションを行い、相互注意スコアが学習によって適切に変化することを示した。

2 リザーバーコンピューティング

RC は、入力層と”リザーバー”と呼ばれる再帰結合を持つ隠れ層、出力層の3層から構成される再帰型ニューラルネットワークの一種である[2]。一般的な RNN は、再帰結合を含む全ての結合荷重を学習するため、勾配消失問題や学習コストが大きいことが問題である。一方で、RC はリザーバー内に大量のニューロンを配置することで、入力信号を非線形活性化関数によって高次元空間に写像し、線形分離可能性を高める。これにより、入力層からリザーバーへの結合荷重やリザーバー再帰結合荷重をランダムな初期値に固定し、リザーバーから出力層への荷重のみを学習することで、大幅な学習コストの削減を実現できる。

3 相互注意機構・自己注意機構

相互注意機構(cross-attention)は、seq2seq[3]と呼ばれるNMTのためのRNNモデルに対して適用されたことをきっかけに注目を集めた仕組みである。予測したい直前のデコーダ側のLSTM内部状態とエンコーダ側の各単語入力時のLSTM内部状態とのスコアを計算することで、ソースとターゲットの単語の照応関係を捉えることができる。

並列処理のために RNN を用いず Attention プロ

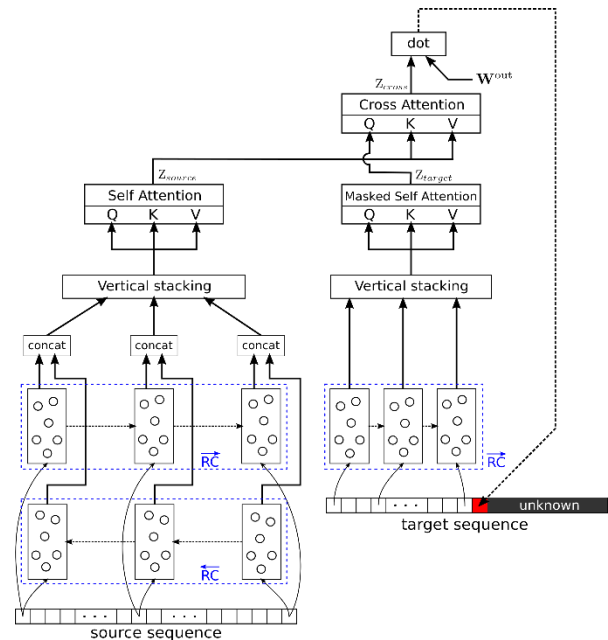


図1: 提案手法の構成。ソース系列(左部)とターゲット系列(右部)それぞれの自己注意計算と系列間の相互注意計算を行う。

ックのみで構成された Transformer[1]では、相互注意機構だけでなく、自己注意機構(self-attention)も採用された。自己注意機構は、エンコーダ・デコーダ内での潜在表現ベクトルのスコアをそれぞれ計算することで、各入力文章内の照応関係を捉えることができる。

4 従来研究

近年、複雑な系列データ予測などに対する注意機構の性能の高さから、RCに自己注意機構を組み込んだシステムがいくつか提案されている。Liuらによって提案された Attention-based ESN[4]は Attention 計算が RC 部よりも前で行われるため、Attention 内の線形変換行列を学習できず、遺伝的アルゴリズムを用いて最適化する必要がある。また、Lyuらは、発話や脳波など様々な分類タスクのために、リザーバーを複数並列に用意し、Multi-head Self-attention Memory Encoder (MSME) と Convolutional Memory Learner (CML)を組み合わせたモデルを提案した[5]。しかし、これらの従来手法は相互注意機構が組み込まれていないため、系列間の関係性を捉えることができない。

Proposal of reservoir computing by using self- and cross-attention mechanisms for learning long-term sequential dependencies, Bungo Konishi[†], Akira Hirose[†], and Ryo Natsuaki[†], Department of Electrical Engineering and Information Systems Graduate School of Engineering, The University of Tokyo[†].

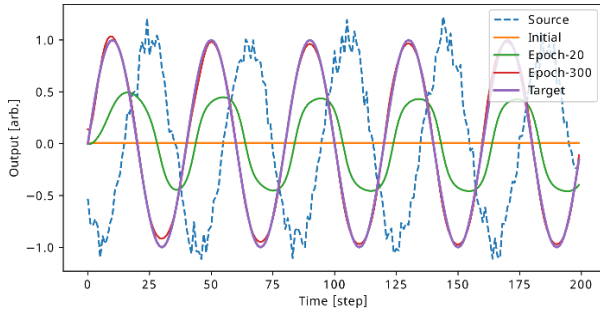


図 2: ソース・ターゲット信号と学習時における出力信号の変化。

5 提案手法

われわれは、ソース系列信号 $f(g(t))$ からターゲット系列信号 $g(t)$ を抽出する系列信号フィルタリングのためのモデルを提案する。例えば、ノイズと遅延を含むソース系列信号からノイズや遅れを持たないターゲット系列信号を予測することを想定している。

本提案手法の構成を図 1 に示す。ソース系列信号とターゲット系列信号は、それぞれ RC 部によって高次元の特徴量に変換される。入力系列が画像や文章である場合は、図 1 のように双方向の RC を用意し、結合することで 1 つの特徴量を得る。リザーバ部のダイナミクスを次に示す。

$$\mathbf{x}(t) = (\mathbf{I} - \Lambda)\mathbf{x}(t) + \Lambda \tanh(\mathbf{W}^{\text{in}}\mathbf{u}(t) + \mathbf{W}\mathbf{x}(t-1)) \quad (1)$$

$\mathbf{x}(t), \mathbf{u}(t)$ はそれぞれ時刻 t におけるリザーバ内部状態と入力信号値、 $\mathbf{W}^{\text{in}}, \mathbf{W}$ はそれぞれ入力層からリザーバへの荷重行列とリザーバ内の再帰結合荷重、 \mathbf{I} は単位行列、 Λ は漏れ率行列である。

リザーバ部によって変換された信号をそれぞれまとめて、1 つの行列を作成後、Attention ブロックへ入力する。自己注意計算も相互注意計算も内部のダイナミクスは同様である。各 Attention ブロックからの出力 \mathbf{z}_i は、次式で表される。

$$\mathbf{z}_i = \text{softmax}\left(\frac{\mathbf{W}_i^{\text{Q}}(\mathbf{W}_i^{\text{K}})^{\text{T}}}{\sqrt{d_{\text{model}}}}\right) \mathbf{V}_i \mathbf{W}_i^{\text{V}} \quad (2)$$

$\mathbf{W}^{\text{Q}}, \mathbf{W}^{\text{K}}, \mathbf{W}^{\text{V}}$ はそれぞれ、Attention ブロックへの入力 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ に対する線形変換行列である。ソフトマックス関数の勾配消失を防ぐため、入力データの次元数 d_{model} で正規化を施している。この $\text{softmax}(\cdot)$ は Query と Key に与えられた信号の線形変換後の類似度を示している。

そして、相互注意計算後、次式により、出力結合荷重 \mathbf{W}^{out} との内積を計算することで所望のターゲット信号を出力する。

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

学習は Attention ブロック内の線形変換行列と \mathbf{W}^{out} のみを勾配法を用いて行うため、ソース側のリザーバ計算はあらかじめ処理できる。

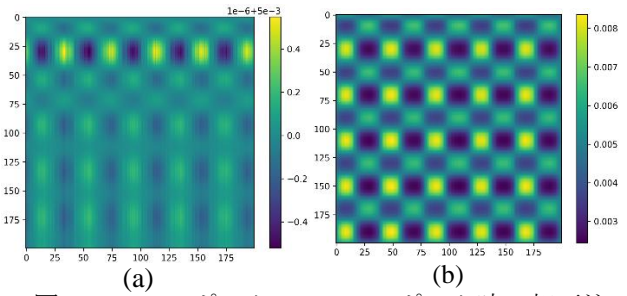


図 3: (a) 0 エポック, (b) 300 エポック時の相互注意機構のスコア行列 (softmax 値) の可視化結果 (縦軸: ターゲット信号, 横軸: ソース信号)。

6 数値シミュレーション実験

図 2 は、時間遅れ ($\tau = 25$) と加法性ガウスノイズ (SNR = 17dB) を持つソース側 sin 信号から、遅れ・ノイズを含まないターゲット側 sin 信号への変換を学習したときの出力信号の変化を示している。また、図 3 は、(a) 0 エポックと (b) 300 エポックでの相互注意機構ブロック内の softmax スコア行列の可視化結果である。(b) 300 エポックにおけるスコアは、例えばターゲット信号 (縦軸) が 150 [step] の場合、ソース信号 (横軸) が遅れと周期が対応する 5, 45, 85, 125, 165 [step] のときにスコアが高くなっており、学習が進むにつれて、系列間の関係性を学習できていることがわかる。

7 むすび

われわれは、2 つの系列信号の長期的な依存関係を学習するための自己注意機構と相互注意機構を持つ新しい RC 手法を提案した。本手法は、リザーバ計算や Attention ブロック内の線形写像によって変換された後の複雑な系列信号の自己相関や相互相関を捉えることができる。また、数値実験により、Attention ブロック内の類似度を表す行列を可視化することで、RC の解釈可能性を高められることを説明した。

参考文献

- [1] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
- [2] Jaeger, Herbert. "The "echo state" approach to analysing and training recurrent neural networks-with an erratum note." German National Research Center for Information Technology GMD Technical Report 148.34 (2001): 13.
- [3] Sutskever, Ilya, et al. "Sequence to sequence learning with neural networks." Advances in neural information processing systems 27 (2014).
- [4] Liu, Chongdang, et al. "Attention Based Echo State Network: A Novel Approach for Fault Prognosis" ICMLC (2019): 489-493.
- [5] Lyu, Huizi, et al. "Multiscale echo self-attention memory network for multivariate time series classification." Neurocomputing 520 (2023): 60-72.