ディープラーニングと物理学

4.2　再帰的ニューラルネットワークと誤差逆伝播法

須賀勇貴

茨城大学大学院　理工学研究科　量子線科学専攻　2 年

April 15, 2023

時系列データについて

系列データ

個々の要素が順序付きの集まりとして与えられるデータのこと
(ex)

- 動画データ $\to$ 順序付きの自然画像データ
- 文章データ $\to$ 順序付きの文字画像データ
- 会話データ $\to$ 順序付きの音声データ

長さが T の系列データは以下のように表現できる

$$\begin{pmatrix} x(1) \\ x(2) \\ x(3) \\ \vdots \\ x(T) \end{pmatrix} = |x(t)\rangle \quad (t = 1, 2, 3, \cdots, T)$$

時系列データについて

(ex) "This is an apple ." という文章データを系列データとして扱う場合

$$
\begin{aligned}
|x(1)\rangle &= |\mathsf{This}\rangle \\
|x(2)\rangle &= |\mathsf{is}\rangle \\
|x(3)\rangle &= |\mathsf{an}\rangle \\
|x(4)\rangle &= |\mathsf{apple}\rangle \\
|x(5)\rangle &= |.\rangle
\end{aligned}
$$

文字データなどは 1-of-K ベクトルなどにより数値ベクトルとして表現される

$$
\boldsymbol{x}^t = \left(\overset{\text{a}}{\overset{\searrow}{0}} \;\; \overset{\text{aah}}{\overset{\downarrow}{0}} \;\; \cdots \;\; \overset{\text{applause}}{\overset{\searrow}{0}} \;\; \overset{\text{apple}}{\overset{\downarrow}{1}} \;\; \overset{\text{applet}}{\overset{\swarrow}{0}} \;\; \cdots \;\; \overset{\text{zzz}}{\overset{\swarrow}{0}} \right)^\top
$$

図: "apple" の 1-of-k ベクトル表示

再帰的ニューラルネットワーク (RNN) の考え方

これまで、扱ってきたデータはデータ間につながりがないものだった

$$\Downarrow$$

系列データをニューラルネットワークで扱えるようにしたい

$$\Downarrow$$

データ間のつながりを表現できるようなニューラルネットワークを構築すればよい！

再帰的ニューラルネットワーク (RNN) の考え方

素朴な考え方
⇒ 前の時刻の出力を次の時刻の入力に加えるようなニューラルネットワーク

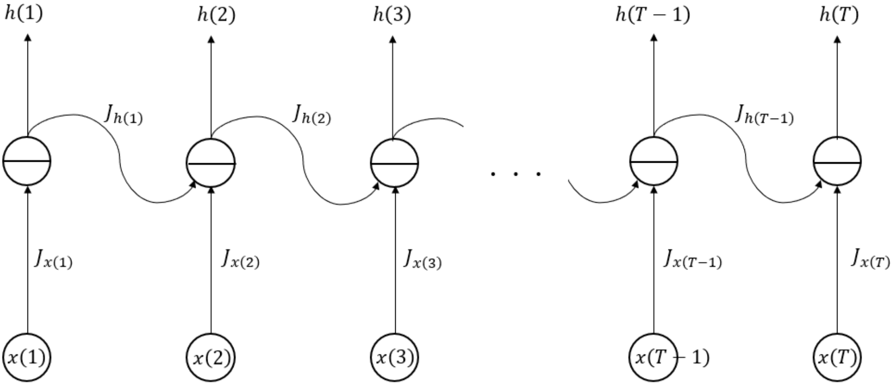
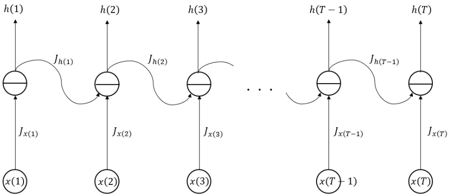


図: 最もシンプルな形の RNN

再帰的ニューラルネットワーク (RNN) の考え方

$$h(1) = \sigma_\bullet(J(1)x(1))$$
$$h(2) = \sigma_\bullet(J(2)x(2) + J_{h(1)}h(1))$$
$$\vdots$$
$$h(t) = \sigma_\bullet(J(t)x(t) + J_{h(t-1)}h(t-1))$$



右のブラケット表記を用いると

$$\begin{aligned} |h(t)\rangle &= \sigma_\bullet(J(t)\,|x(t)\rangle + J_{h(t-1)}\,|h(t-1)\rangle) \\ &= \sigma_\bullet \sum_m |m\rangle\,(\langle m|\,J(t)\,|x(t)\rangle \\ &\quad + \langle m|\,J_{h(t-1)}\,|h(t-1)\rangle) \end{aligned}$$

$$\begin{aligned} |x(t)\rangle &= (x(1), x(2), \cdots, x(T))^\top \\ |h(t)\rangle &= (h(1), h(2), \cdots, h(T))^\top \\ \mathbb{J}_x &= \mathrm{diag}(J_{x(1)}, J_{x(2)}, \cdots, J_{x(T)}) \\ \mathbb{J}_h &= \mathrm{diag}(J_{h(1)}, J_{h(2)}, \cdots, J_{h(T)}) \end{aligned}$$

再帰的ニューラルネットワーク(RNN)の考え方

各時刻での出力は以下のように簡単に求められる

$$
\begin{aligned}
h(1) &= \sigma_\bullet(J_{x(1)}x(1)) \\
h(2) &= \sigma_\bullet(J_{x(2)}x(2) + J_{h(1)}h(1)) \\
&\vdots \\
h(t) &= \sigma_\bullet(J_{x(t)}(t)x(t) + J_{h(t-1)}h(t-1))
\end{aligned}
$$

ブラケット表記より，まとめて以下のように書ける

$$
\begin{aligned}
|h(t)\rangle &= \sigma_\bullet(\mathbb{J}_x\,|x(t)\rangle + \mathbb{J}_h\,|h(t-1)\rangle) \\
&= \sum_m |m\rangle\,\sigma_\bullet\Big(\,\langle m|\,\mathbb{J}_x\,|x(t)\rangle + \langle m|\,\mathbb{J}_h\,|h(t-1)\rangle\,\Big)
\end{aligned}
$$

RNN における誤差逆伝播

誤差関数 L は典型的に以下のようになっているとする

$$L = \langle d(1)|h(1)\rangle + \langle d(2)|h(2)\rangle + \cdots + \langle d(T)|h(T)\rangle = \sum_{t=1}^{T} \langle d(t)|h(t)\rangle$$

t 番目の変化量に着目 (計算↓)

$$\begin{aligned}
\delta \langle d(t)|h(t)\rangle &= \langle d(t)|\, \delta \,|h(t)\rangle \\
&= \cdots \\
&= \langle \delta_t(t)|\, \delta\mathbb{J}_x \,|x(t)\rangle + \langle \delta_t(t)|\, \delta\mathbb{J}_h \,|h(t-1)\rangle \\
&+ \langle \delta_t(t-1)|\, \delta\mathbb{J}_x \,|x(t-1)\rangle + \langle \delta_t(t-1)|\, \delta\mathbb{J}_h \,|h(t-2)\rangle \\
&+ \cdots \\
&+ \langle \delta_t(1)|\, \delta\mathbb{J}_x \,|x(1)\rangle + \langle \delta_t(1)|\, \delta\mathbb{J}_h \,|h(0)\rangle \\
&= \sum_{\tau \le t} \Big(\langle \delta_t(\tau)|\, \delta\mathbb{J}_x \,|x(\tau)\rangle + \langle \delta_t(\tau)|\, \delta\mathbb{J}_h \,|h(\tau-1)\rangle \Big)
\end{aligned}$$

RNN における誤差逆伝播

ここで，

$$\delta \mathbb{J} = \sum_{m,n} |m\rangle \langle n| \, \delta J \ mn$$

とすると

$$\begin{aligned}
&\delta \langle d(t)|h(t)\rangle \\
&= \sum_{\tau \le t} \Big(\langle \delta_t(\tau)| \, \delta \mathbb{J}_x \, |x(\tau)\rangle + \langle \delta_t(\tau)| \, \delta \mathbb{J}_h \, |h(\tau - 1)\rangle \Big) \\
&= \sum_{\tau \le t} \sum_{m,n} \Big(\langle \delta_t(\tau)|m\rangle \langle n|x(\tau)\rangle \, \delta J_x^{mn} + \langle \delta_t(\tau)|m\rangle \langle n|h(\tau - 1)\rangle \, \delta J_h^{mn} \Big) \\
&= \sum_{m,n} \Big(\sum_{\tau \le t} \langle \delta_t(\tau)|m\rangle \langle n|x(\tau)\rangle \, \delta J_x^{mn} + \sum_{\tau \le t} \langle \delta_t(\tau)|m\rangle \langle n|h(\tau - 1)\rangle \, \delta J_h^{mn} \Big)
\end{aligned}$$

RNNにおける誤差逆伝播

以上の結果から誤差関数 L の変化量は

$$\begin{aligned}\delta L &= \sum_{t=1}^{T} \delta \langle d(t)|h(t)\rangle \\ &= \sum_{m,n} \Bigg(\sum_{t=1}^{T} \sum_{\tau \le t} \langle \delta_t(\tau)|m\rangle \langle n|x(\tau)\rangle \, \delta J_x^{mn} \\ &\qquad + \sum_{t=1}^{T} \sum_{\tau \le t} \langle \delta_t(\tau)|m\rangle \langle n|h(\tau-1)\rangle \, \delta J_h^{mn} \Bigg)\end{aligned}$$

と表すことができる．この結果より，パラメータ J_x^{mn} と J_h^{mn} はそれぞれ

$$\delta J_x^{mn} = -\epsilon \sum_{t=1}^{T} \sum_{\tau \le t} \langle \delta_t(\tau)|m\rangle \langle n|x(\tau)\rangle$$

$$\delta J_h^{mn} = -\epsilon \sum_{t=1}^{T} \sum_{\tau \le t} \langle \delta_t(\tau)|m\rangle \langle n|h(\tau-1)\rangle$$

という更新ルールにすれば，誤差関数の値が小さくなるようになる．

$\delta \left| h(t) \right\rangle$ の計算

パラメータに依存するのは $\mathbb{J}_{x,h}$ と $\left| h(t-1) \right\rangle$ なので

$$\begin{aligned}
\delta \left| h(t) \right\rangle &= \underbrace{\sum_m \left| m \right\rangle \sigma'_{\bullet}\Big(\left\langle m \right| \mathbb{J}_x \left| x(t) \right\rangle + \left\langle m \right| \mathbb{J}_h \left| h(t-1) \right\rangle \Big) \left\langle m \right|}_{=:\boldsymbol{G}(t)} \\
&\quad \times \delta\Big(\mathbb{J}_x \left| x(t) \right\rangle + \mathbb{J}_h \left| h(t-1) \right\rangle \Big) \\
&= \boldsymbol{G}(t)\Big(\delta\mathbb{J}_x \left| x(t) \right\rangle + \delta\mathbb{J}_h \left| h(t-1) \right\rangle + \mathbb{J}_h \delta \left| h(t-1) \right\rangle \Big) \\
&= \boldsymbol{G}(t)\delta\mathbb{J}_x \left| x(t) \right\rangle + \boldsymbol{G}(t)\delta\mathbb{J}_h \left| h(t-1) \right\rangle + \boldsymbol{G}(t)\mathbb{J}_h \delta \left| h(t-1) \right\rangle
\end{aligned}$$