

ディープラーニングと物理学

4.3 LSTM

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 17, 2023

LSTM の構造

LSTM(Long short-term memory)

RNN の中間層のユニットをメモリー・ユニットに置き換えもの

メモリー・ユニットには情報の保持や忘却を可能にする仕組みが組み込まれている

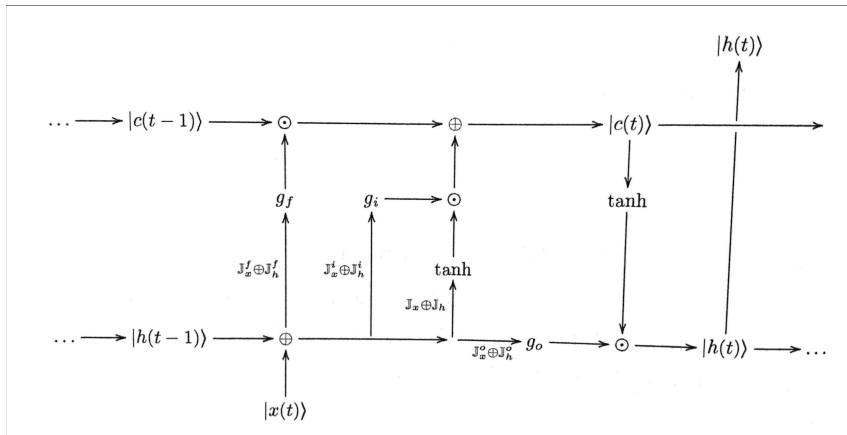


図: LSTM の模式図. \oplus はベクトルの和, \odot はアダマール積を表す

LSTM の構造

LSTM の核心となるのはメモリベクトル $|c(t)\rangle$ で、「一時的な記憶」を司る

g はゲート (gate) であり, それぞれ

g_f : 忘却ゲート

g_i : 入力ゲート

g_o : 出力ゲート

と呼ばれ, すべて成分ごとのシグモイド関数を取る

$$g_f = g_i = g_o = \sigma$$

模式図から以下の関係式がわかる

$$\begin{aligned}
 |h(t)\rangle &= \sum_m |m\rangle \langle m|g_o\rangle \langle m|\tanh(c(t))\rangle \\
 |c(t)\rangle &= \sum_m |m\rangle \langle m|c(t-1)\rangle \langle m|g_f\rangle \\
 &\quad + \sum_m |m\rangle \tanh\left(\langle m|\mathbb{J}_x|x(t)\rangle + \langle m|\mathbb{J}_h|h(t-1)\rangle\right) \langle m|g_i\rangle \\
 |g_f\rangle &= \sum_m |m\rangle g_f \left(\langle m|\mathbb{J}_x^f|x(t)\rangle + \langle m|\mathbb{J}_h^f|h(t-1)\rangle\right) \\
 |g_i\rangle &= \sum_m |m\rangle g_i \left(\langle m|\mathbb{J}_x^i|x(t)\rangle + \langle m|\mathbb{J}_h^i|h(t-1)\rangle\right) \\
 |g_o\rangle &= \sum_m |m\rangle g_o \left(\langle m|\mathbb{J}_x^o|x(t)\rangle + \langle m|\mathbb{J}_h^o|h(t-1)\rangle\right)
 \end{aligned}$$

誤差逆伝播法

前節の RNN の時と同様の手順で誤差の逆伝播を見ていく

$$\delta h(t) = \sum_m \left[\underbrace{\langle m | \delta | g_o \rangle}_{(A)} \langle m | \tanh(c(t)) \rangle + \langle m | g_o \rangle \underbrace{\langle m | \delta | \tanh(c(t)) \rangle}_{(B)} \right]$$

まず, (A) について

$$\begin{aligned} (A) &= \delta \sum_m |m\rangle g_o \left(\langle m | \mathbb{J}_x^o | x(t) \rangle + \langle m | \mathbb{J}_h^o | h(t-1) \rangle \right) \\ &= \sum_m |m\rangle g'_o(\bullet) \left[\delta \left(\langle m | \mathbb{J}_x^o | x(t) \rangle + \langle m | \mathbb{J}_h^o | h(t-1) \rangle \right) \right] \\ &= \sum_m |m\rangle g'_o(\bullet) \left[\langle m | \delta \mathbb{J}_x^o | x(t) \rangle + \langle m | \delta \mathbb{J}_h^o | h(t-1) \rangle + \underbrace{\langle m | \mathbb{J}_h^o \delta | h(t-1) \rangle}_{\star} \right] \end{aligned}$$

★ は時間をさかのぼって計算することができる．その際に \mathbb{J}_h^o が毎回1個ずつ出てくるので結局 \mathbb{J}_h^o をたくさんかけることになり, 勾配爆発/勾配消失が起こる

誤差逆伝播法

前節の RNN の時と同様の手順で誤差の逆伝播を見ていく

$$\delta h(t) = \sum_m \left[\langle m | \underbrace{\delta |g_o\rangle}_{(A)} \langle m | \tanh(c(t)) \rangle + \langle m | g_o \rangle \langle m | \underbrace{\delta | \tanh(c(t)) \rangle}_{(B)} \right]$$

次に, (B) について

$$(B) = \delta \sum_m |m\rangle \tanh(\langle m | c(t) \rangle) = \sum_m |m\rangle \tanh'(\bullet) \left[\langle m | \underbrace{\delta | c(t) \rangle}_{(C)} \right]$$

勾配の大きさは (C) で決まる. 計算すると

$$\begin{aligned} (C) = & \sum_m |m\rangle \left[\langle m | \delta | c(t-1) \rangle \langle m | g_f \rangle + \langle m | c(t-1) \rangle \underbrace{\langle m | \delta | g_f \rangle}_{(D)} \right] \\ & + \sum_m |m\rangle \left[\tanh'(\bullet) \left(\langle m | \delta \mathbb{J}_x | x(t) \rangle + \langle m | \delta \mathbb{J}_h | h(t-1) \rangle + \langle m | \mathbb{J}_h \delta | h(t-1) \rangle \right) \langle m | g_i \rangle \right. \\ & \left. + \tanh(\bullet) \underbrace{\langle m | \delta | g_i \rangle}_{(F)} \right] \end{aligned}$$

(C) 計算

$$\begin{aligned}
 (C) &= \delta \sum_m |m\rangle \langle m|c(t-1)\rangle \langle m|g_f\rangle \\
 &+ \delta \sum_m |m\rangle \tanh\left(\langle m|\mathbb{J}_x|x(t)\rangle + \langle m|\mathbb{J}_h|h(t-1)\rangle\right) \langle m|g_i\rangle \\
 &= \sum_m |m\rangle \left[\langle m|\delta|c(t-1)\rangle \langle m|g_f\rangle + \langle m|c(t-1)\rangle \underbrace{\langle m|\delta|g_f\rangle}_{(D)} \right] \\
 &+ \sum_m |m\rangle \left[\tanh'(\bullet) \left(\langle m|\delta\mathbb{J}_x|x(t)\rangle + \langle m|\delta\mathbb{J}_h|h(t-1)\rangle + \langle m|\mathbb{J}_h\delta|h(t-1)\rangle \right) \langle m|g_i\rangle \right. \\
 &\quad \left. + \tanh(\bullet) \underbrace{\langle m|\delta|g_i\rangle}_{(F)} \right]
 \end{aligned}$$