RNN & LSTM

Sung Eun Lee

September 06, 2024

Seoul National University

Contents

1. Introduction
2. Recurrent Neural Networks(RNNs)
3. Long Short-Term Memory(LSTM)

Introduction

Introduction

Sequence Data : Data in which elements are arranged sequentially in a specific order.

- Text : Sequence of words.
- Image : Sequence of Image.

Suitable for text data analysis :

1. Facilitates the extraction of **relational information** between words.
2. Facilitates the extraction of **contextual meaning** of words.

it is well-suited for extracting the information contained in text.

Recurrent Neural Networks(RNNs)

Notation

- x_t : Input Vector at time step t.
- h_t : Hidden State Vector at time step t.

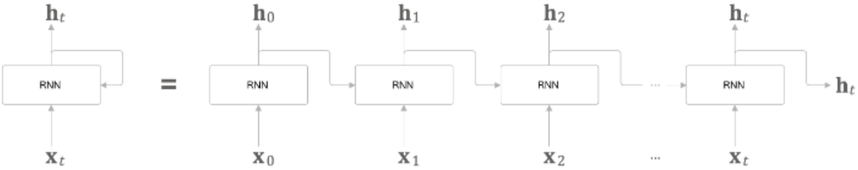

Figure1 : Recurrent Neural Networks Structure.

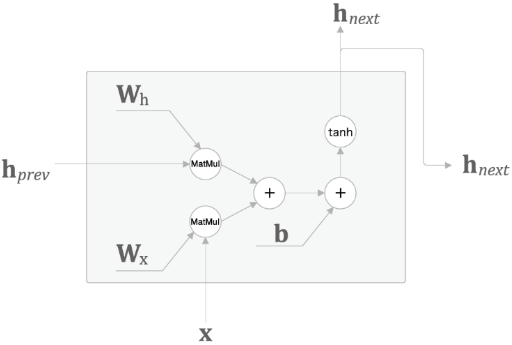

Figure2 : RNN Layer (Detailed).

- The weight matrices such as W_x and W_h are all shared, and this is referred to as **Weight Sharing**.

Formulas

- $h_{\text{next}} = \tanh\left(W_x \cdot x + W_h \cdot h_{\text{prev}} + b\right)$

RNNs

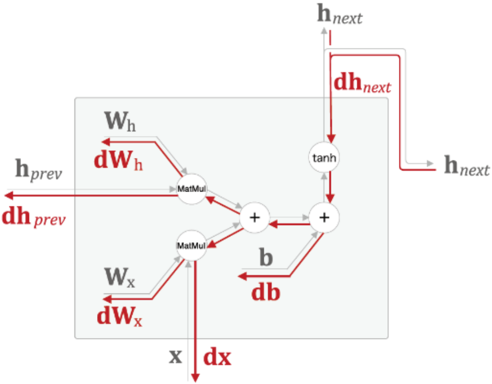

Figure3 : Back Propagation Through Time(BPTT).

- The **Vanishing Gradient** problem occurs during the BPTT process.
- The vanishing gradient problem leads to issues with **long-term dependencies**.

Long Short-Term Memory(LSTM)

Notation

- C_t : Memory cell at time step t
- $\sigma(\cdot)$: sigmoid function
- $\odot$: Hadamard Product
- W_x^g, W_h^g : Weight Matrix, $\quad g \in \{f, I_1, I_2, o, n\}$

  f : forget gate ,

  I_k : k-th Weight Matrix of the input gate , $k =$ 1,2

  o : output gate,

  n : new input.

LSTM Machanism

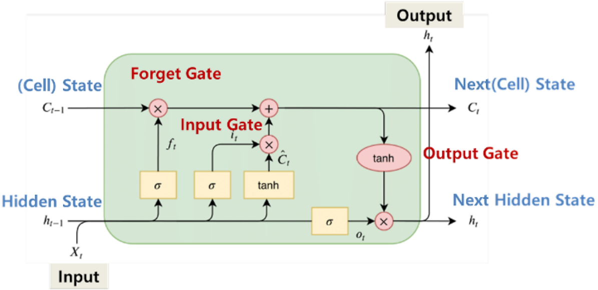

Figure4 : LSTM Structure.

- The key difference from RNNs is the introduction of **Memory Cells** and the concept of **gates**.

Forget gate

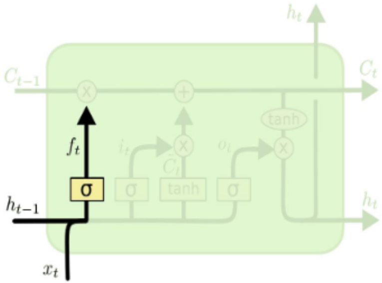

Figure5 : Forget gate structure

Forget Gate determines how much of the information from the previous time step should be forgotten.

Formulas

- $f_t = \sigma(x_t \cdot W_x^f + h_{t-1} \cdot W_h^f + b^f)$.

Input Gate

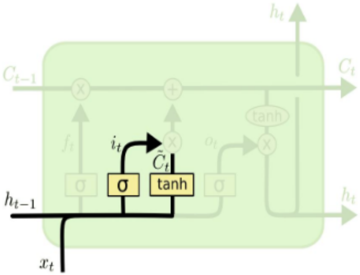

Figure6 : Input gate structure

Input Gate is a gate responsible for adding the information from the current word x_t.

Formulas

- $i_t = \sigma(x_t \cdot W_x^i + h_{t-1} \cdot W_h^i + b^i)$
- $\tilde{C}_t = \tanh\left(x_t \cdot W_x^n + h_{t-1} \cdot W_h^n + b^n\right)$

Update

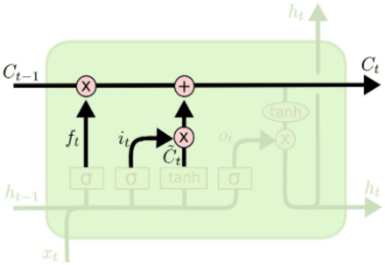

Figure7 : Update Process

This process updates the previous Cell State C_{t-1} to the new State C_t.

Formulas

- $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}$

Output Gate

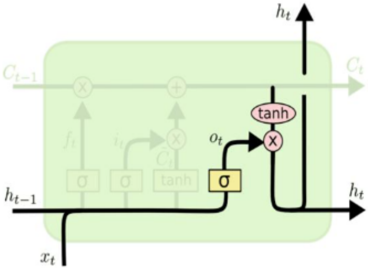

Figure8 : Output gate structure

In the Ouput Gate, the newly updated C_t is used along with the Input Vector x_t and the hidden state vector from the previous time step h_{t-1} to caculate the hidden state h_t for the current time step t.

Formulas

- $o_t = \sigma(x_t \cdot W_x^o + h_{t-1} \cdot W_h^o + b^o)$
- $h_t = o_t \odot \tanh(C_t)$

Q & A