weekly report: amino acid seq

Jitian Zhao

University of Wisconsin Madison

September 23, 2019

Outline

- low performance of Char-RNN
- code modification of Transformer

char-RNN:recap

- training: $\hat{X}_n = sample(P(X_n|X_1, ..., X_{n-1}))$
- generating: $\hat{X}_n = sample(P(X_n|\hat{X}_1, ..., \hat{X_{n-1}}))$

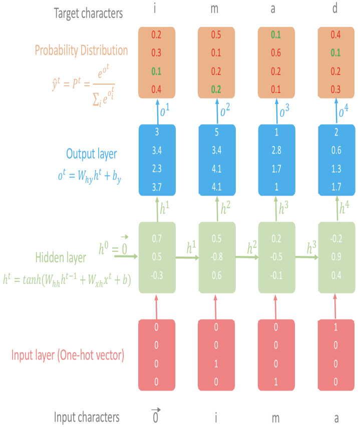

Figure: training process

Source: https://towardsdatascience.com/character-level-language-model-1439f5dd87fe

low performance of Char-RNN

- low reconstruction accuracy for test dataset: 33%
- low accuracy was not due to error propagation
- model insensitive to the given information

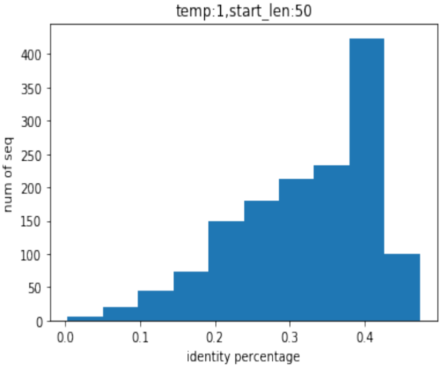

Figure: reconstruction accuracy

possible reasons of why it doesn't work

- simple structure: 1 gru layer + 1 fc layer
- does not learn from the position
- sequence diversity or data clustering
- ...

Transformer

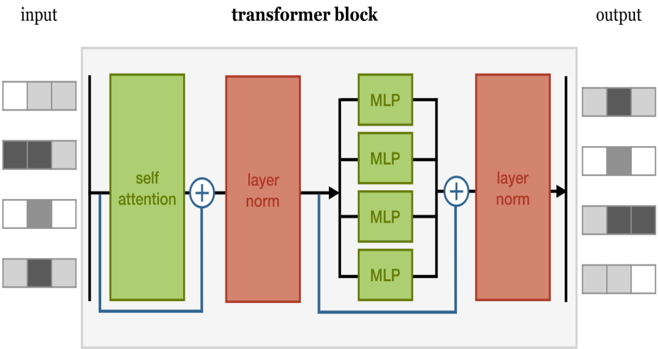

Figure: structure of transformer

Source: `http://peterbloem.nl/blog/transformers`

self-attention mechanism

- input: $X_1, ..., X_t$
- output: $Y_i = \sum_j w_{ij} X_j$
- weight: $w_{ij} = \text{softmax}(w'_{ij}), w'_{ij} = X_i^T X_j$
- one input vector X_i serves 3 needs: contribute to its own weight (query), other inputs weight (key), final weighted sum (value)

catch the position information : multi-head attention

- example: 'mary', 'gives', 'jane', 'a', 'flower'
- neighbor matters!
- different attention heads have different sets of matrices, which gives it power of discriminating neighbors.

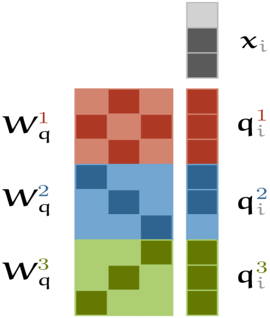

Figure: multi-head attention

Source: http://peterbloem.nl/blog/transformers

tailored modification for a.a. seq

- sequence padding
- random trunk $\longrightarrow$ randomly choose one sequence
- word encoding: different vocabulary