Lecture 1(p2): Recurrent Nets

CSE599G1: Spring 2017
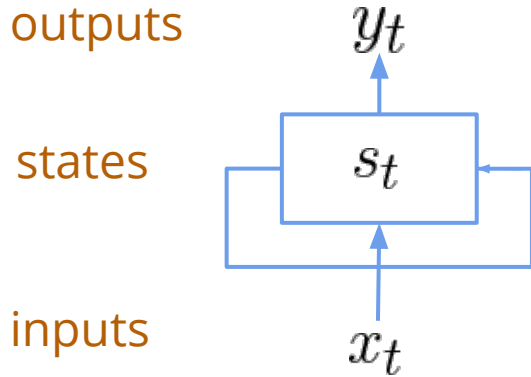
Recurrent Nets for Sequence Data

Why Recurrent Structure

- Solve problems of sequence data: speeches, languages
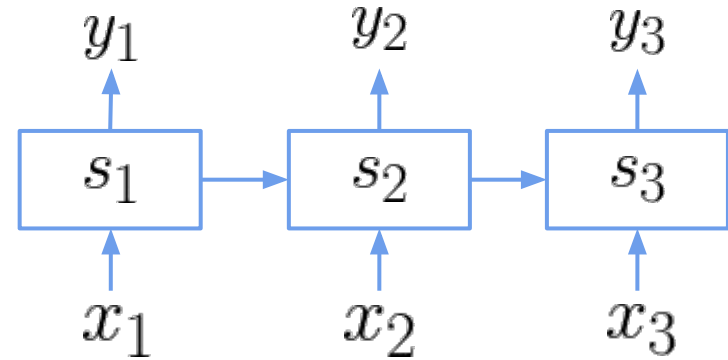- Captures sequence dependencies

output

state

input

hello   world   !

<go>   hello   world
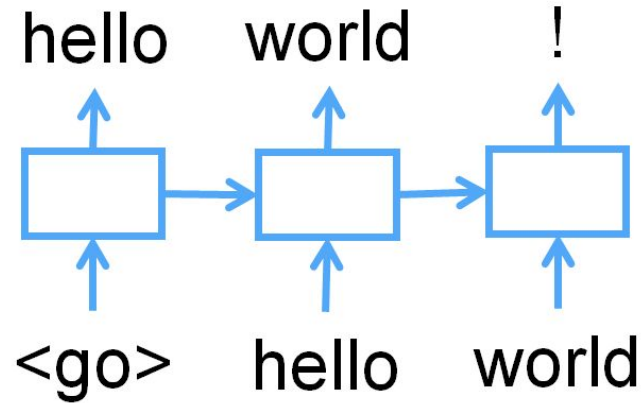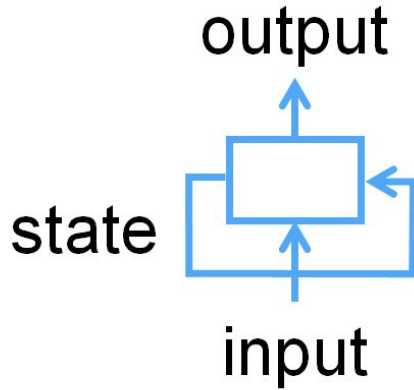
Recurrent Structure: Folded and Unrolled View

Folded View

Unrolled View

outputs y_t

states s_t

inputs x_t

y_1 y_2 y_3
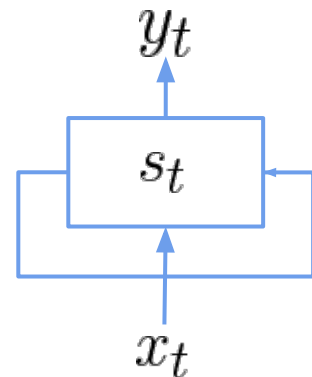
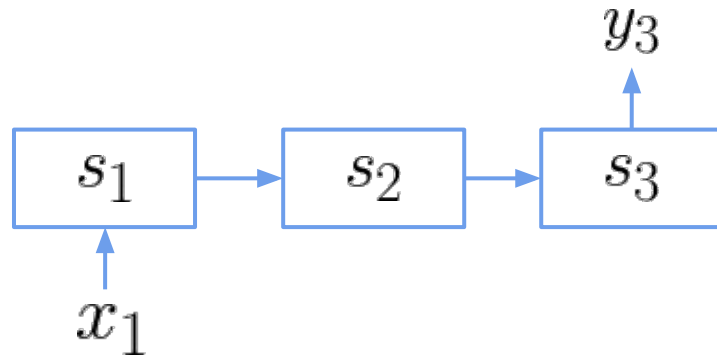s_1 → s_2 → s_3

x_1 x_2 x_3

Simple Recurrent Net

transition $\quad s_t = \sigma(W_1 s_{t-1} + W_2 x_t)$

emission $\quad y_t = \sigma(W_3 s_t)$

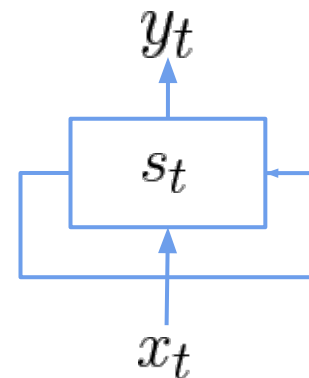Hard to capture long term dependency, exponential multiplication effect

LSTM: Update with Moving Average

Forget Rate

Remembering Rate

$$s_t = \alpha_t s_{t-1} + \beta_t \tanh(W[x_t, s_{t-1}])$$

How to set the forget rate and remembering rate?

Use another neural net module
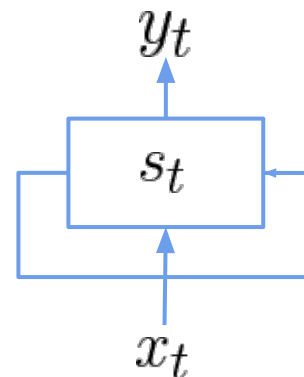
One variant of LSTM

Forget Rate   **Remembering Rate**

$$s_t = \alpha_t s_{t-1} + \beta_t \tanh(W[x_t, s_{t-1}])$$

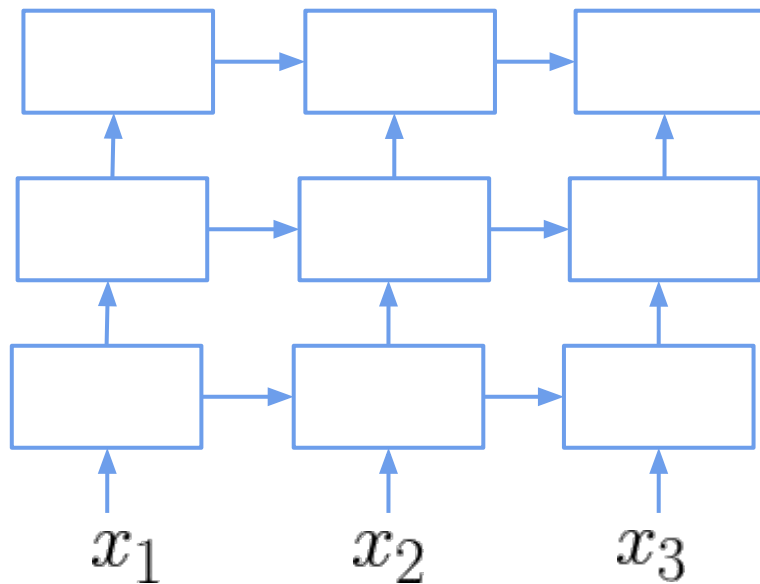$$\alpha_t = \sigma(W_f[x_t, s_{t-1}])$$

$$\beta_t = \sigma(W_r[x_t, s_{t-1}])$$

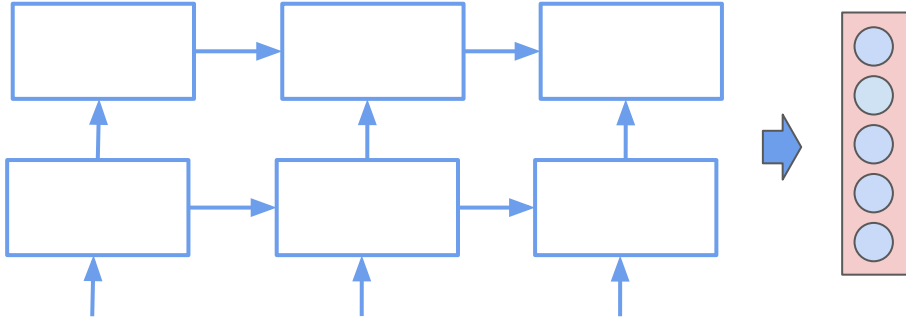Many other variant exists with similar spirit

Unrolled Stacked Recurrent Nets

- Provide Hierarchical representation about sequence
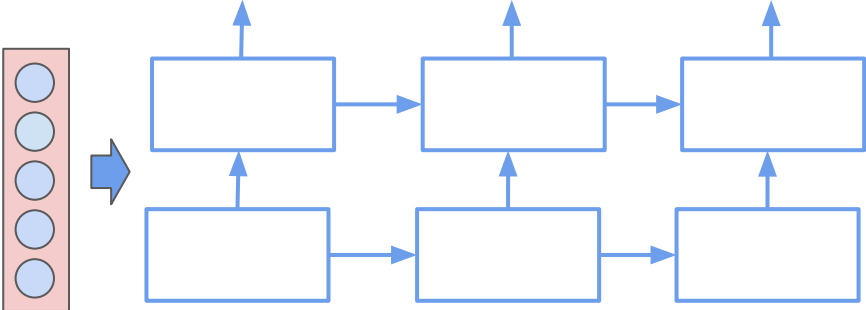- Feed output of one sequence to another RNN

LSTM as Compositional Building Block

Sequence Encoder
sequence to vector

Sequence Decoder
vector to sequence

Seq2Seq: Machine Translation Model
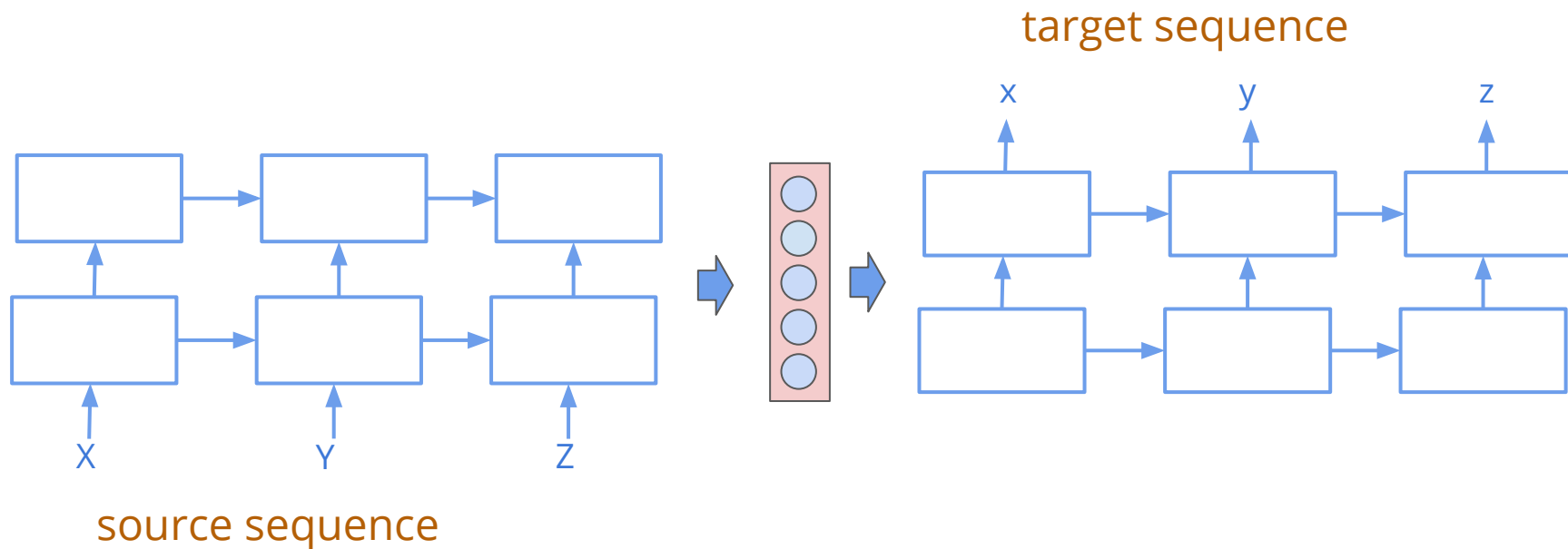

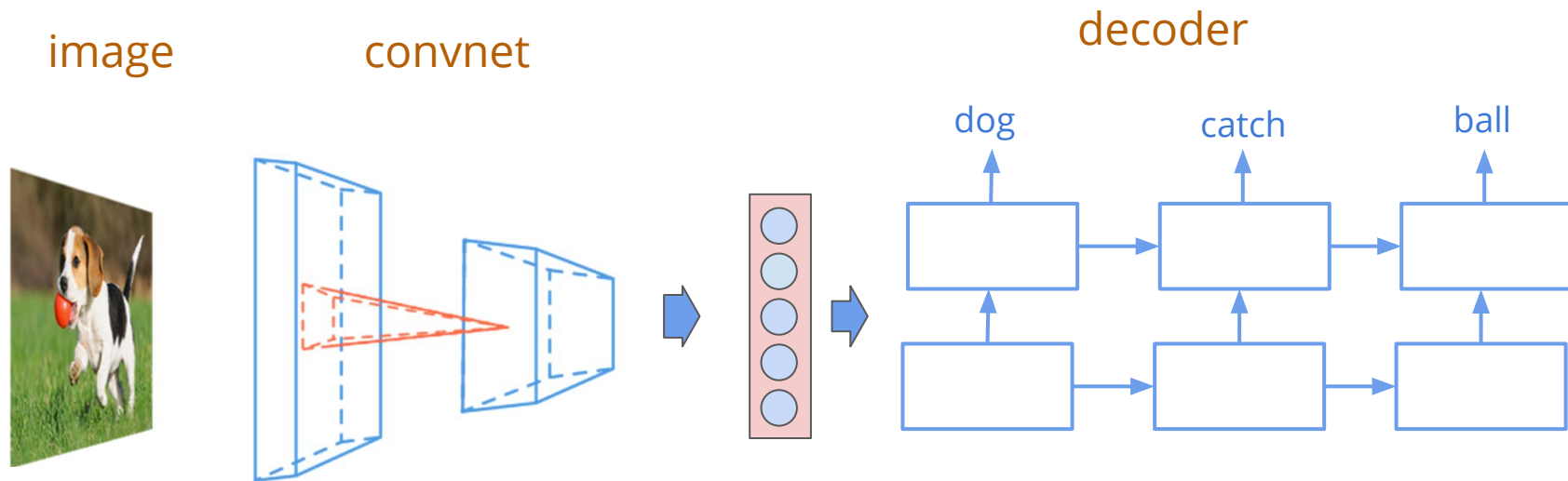target sequence

source sequence

Image2Seq: Image Caption Model

It is all about Composability!

Topics not Covered Today

- More complicated memory structures
  - Memory networks

- Objective derivation for generative models
  - Generative adversarial nets
  - Variational methods

- Gradient estimation with hard decision and interactions
  - Q learning, Policy gradient