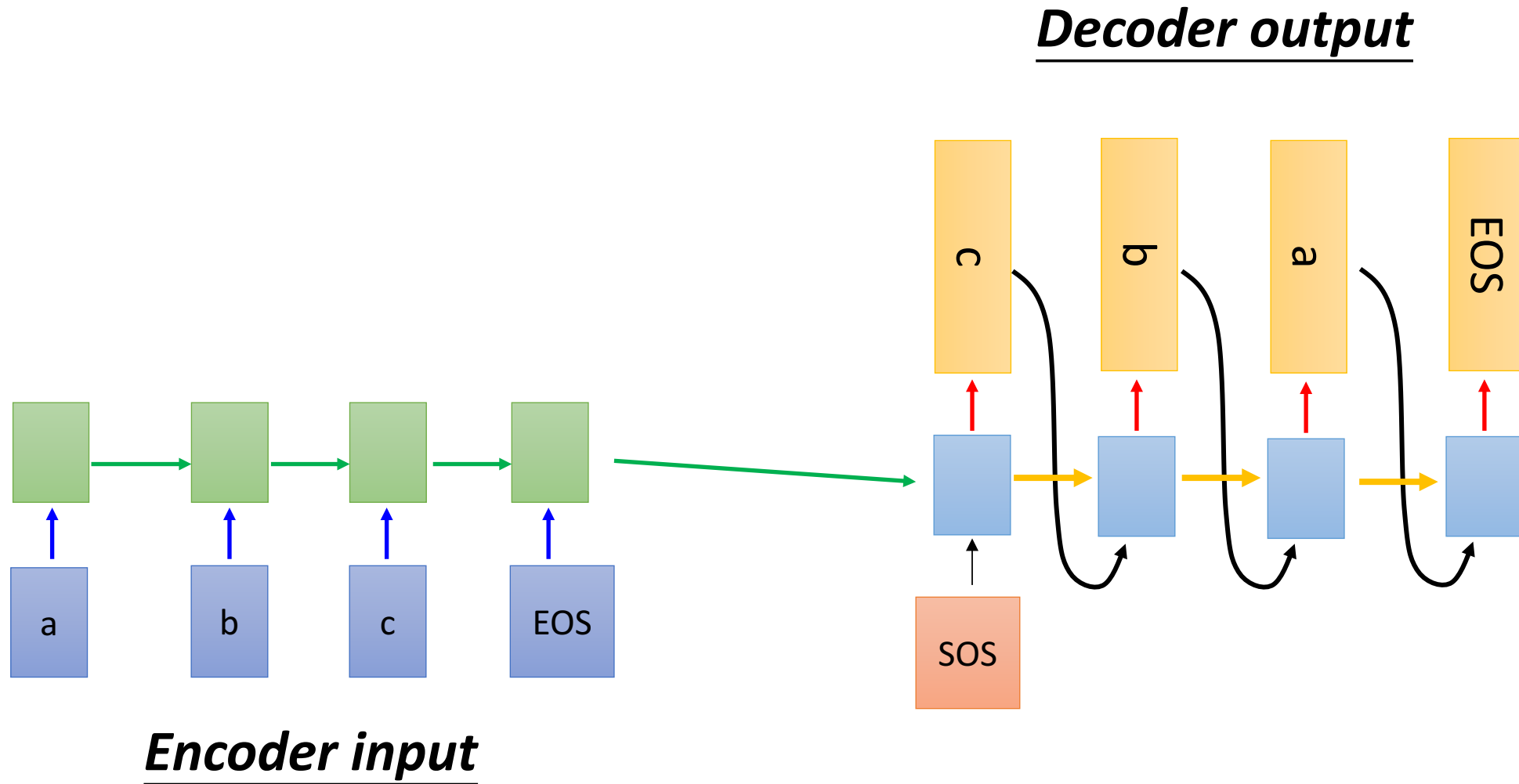


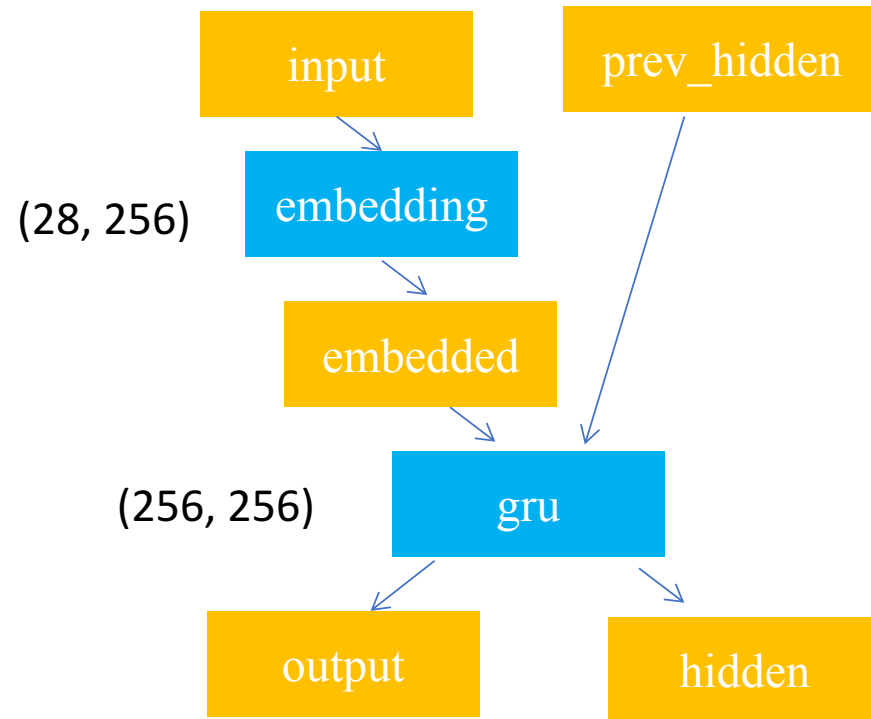
1.10 Report

Taoyuan Wang

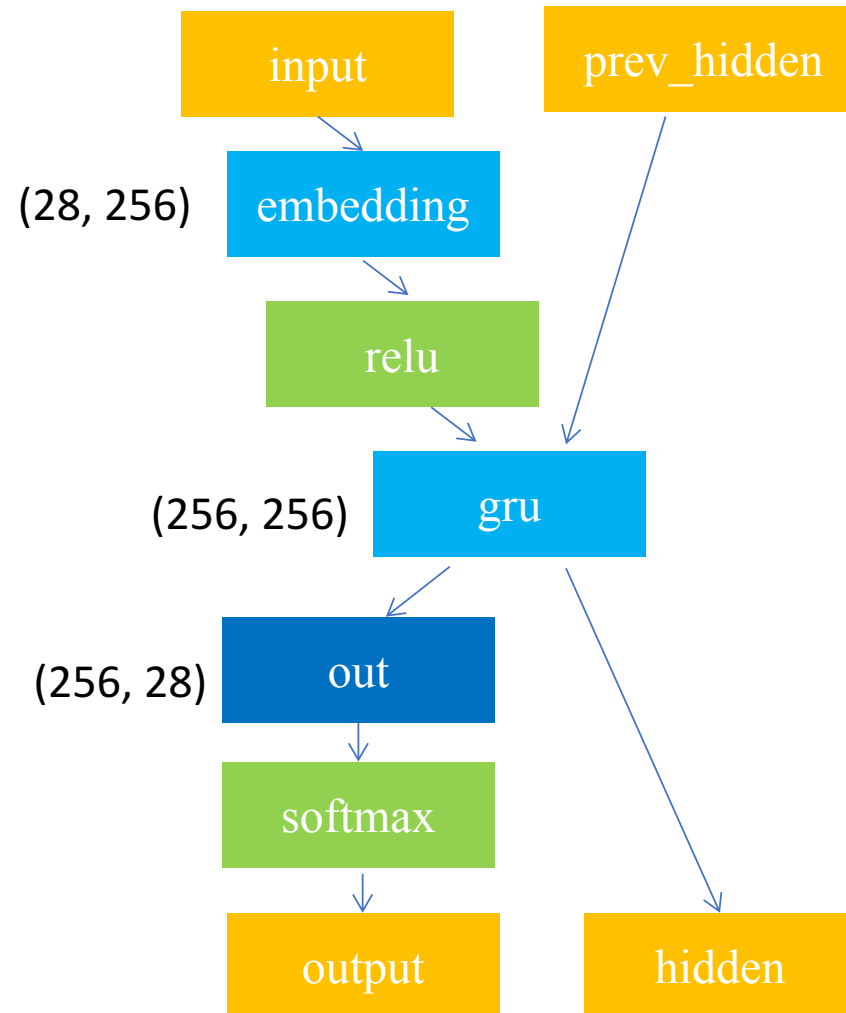
Structure



Encoder



Decoder



28: 26 alphabets + SOS + EOS

256: hidden layer

Experiment

- Random inverse pairs: 20000
- Epoch: 10000
- Learning Rate: 0.01
- Optimizer of Encoder and Decoder: SGD
- Loss Function: NLLLoss (Negative Log Likelihood)
- Trained data accuracy: 49.5%
- Untrained data accuracy: 33.3%
- Training time: 68 min @ deep

```
6m 33s (- 59m 3s) (10000 10%) 1.0354
13m 26s (- 53m 44s) (20000 20%) 0.7540
20m 19s (- 47m 26s) (30000 30%) 0.6540
27m 10s (- 40m 45s) (40000 40%) 0.6374
34m 3s (- 34m 3s) (50000 50%) 0.5988
40m 56s (- 27m 17s) (60000 60%) 0.5701
47m 46s (- 20m 28s) (70000 70%) 0.5494
54m 35s (- 13m 38s) (80000 80%) 0.5287
61m 29s (- 6m 49s) (90000 90%) 0.5054
68m 21s (- 0m 0s) (100000 100%) 0.4828
```

```
in-ac: 0.495
out-ac: 0.333424209378
```

```
input = a b c d
output = d c b a <EOS>
```

```
input = b d d
output = d b b <EOS>
```

```
input = c a b
output = b a c <EOS>
```

```
input = d c b a
output = a b c d <EOS>
```

Embedding

```
>>> # an Embedding module containing 10 tensors of size 3
>>> embedding = nn.Embedding(10, 3)
>>> # a batch of 2 samples of 4 indices each
>>> input = Variable(torch.LongTensor([[1,2,4,5],[4,3,2,9]]))
>>> embedding(input)
```

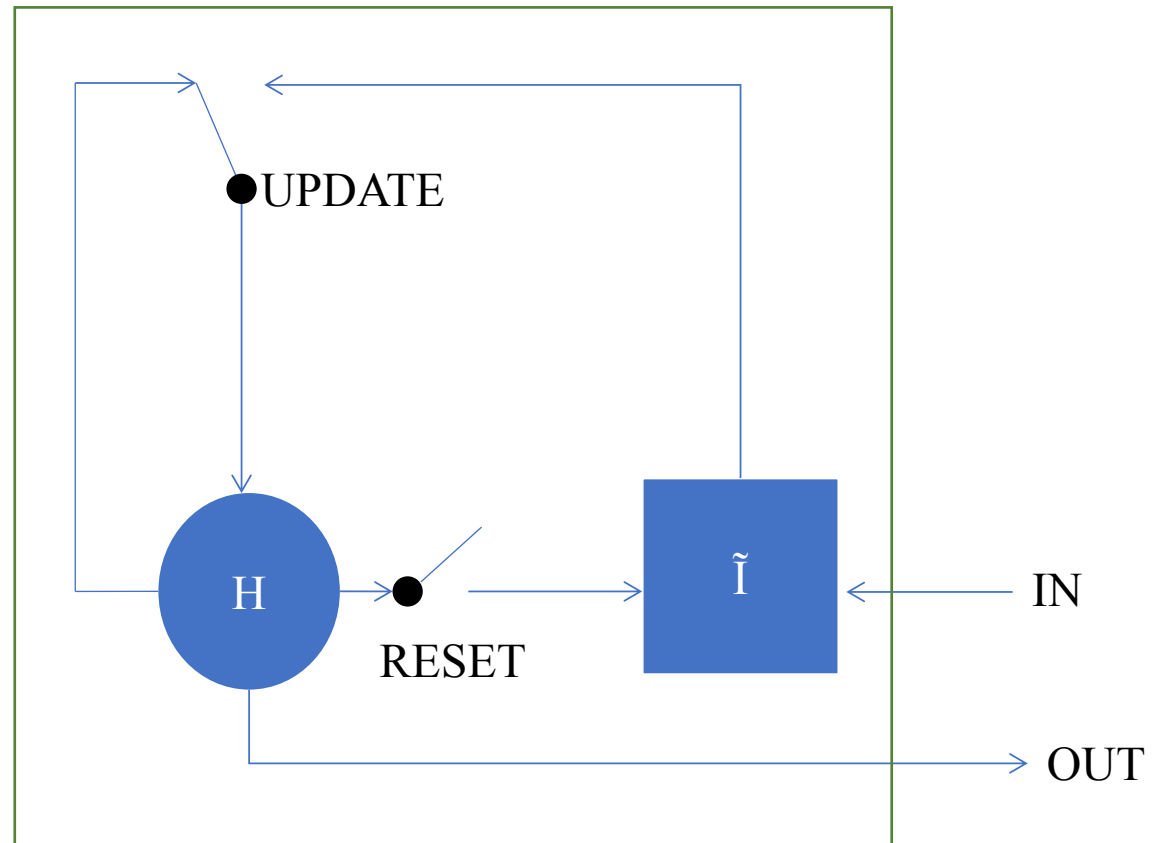
Variable containing:

```
(0 ,.,.) =
-1.0822  1.2522  0.2434
 0.8393 -0.6062 -0.3348
 0.6597  0.0350  0.0837
 0.5521  0.9447  0.0498
```

```
(1 ,.,.) =
 0.6597  0.0350  0.0837
-0.1527  0.0877  0.4260
 0.8393 -0.6062 -0.3348
-0.8738 -0.9054  0.4281
```

```
[torch.FloatTensor of size 2x4x3]
```

GRU

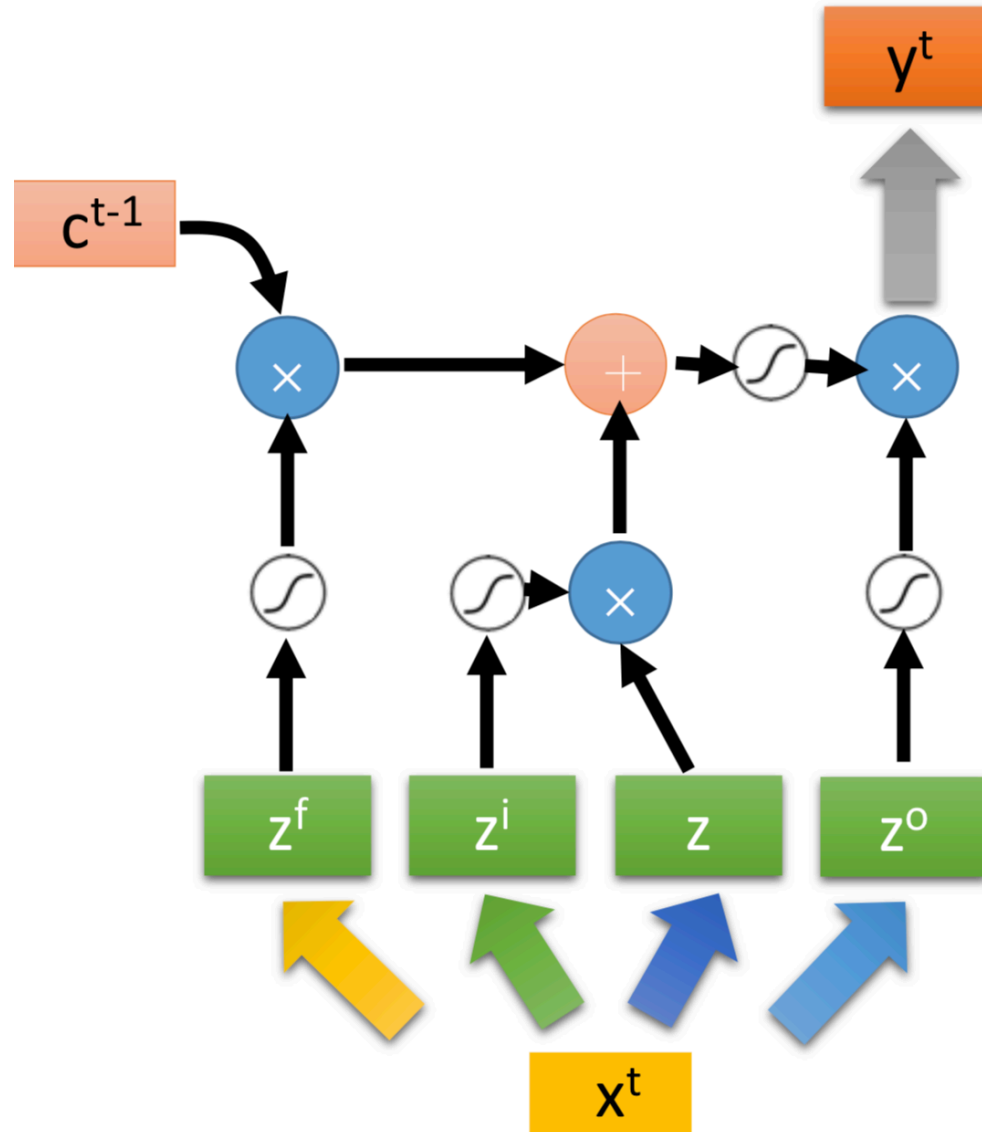


No Output Gate

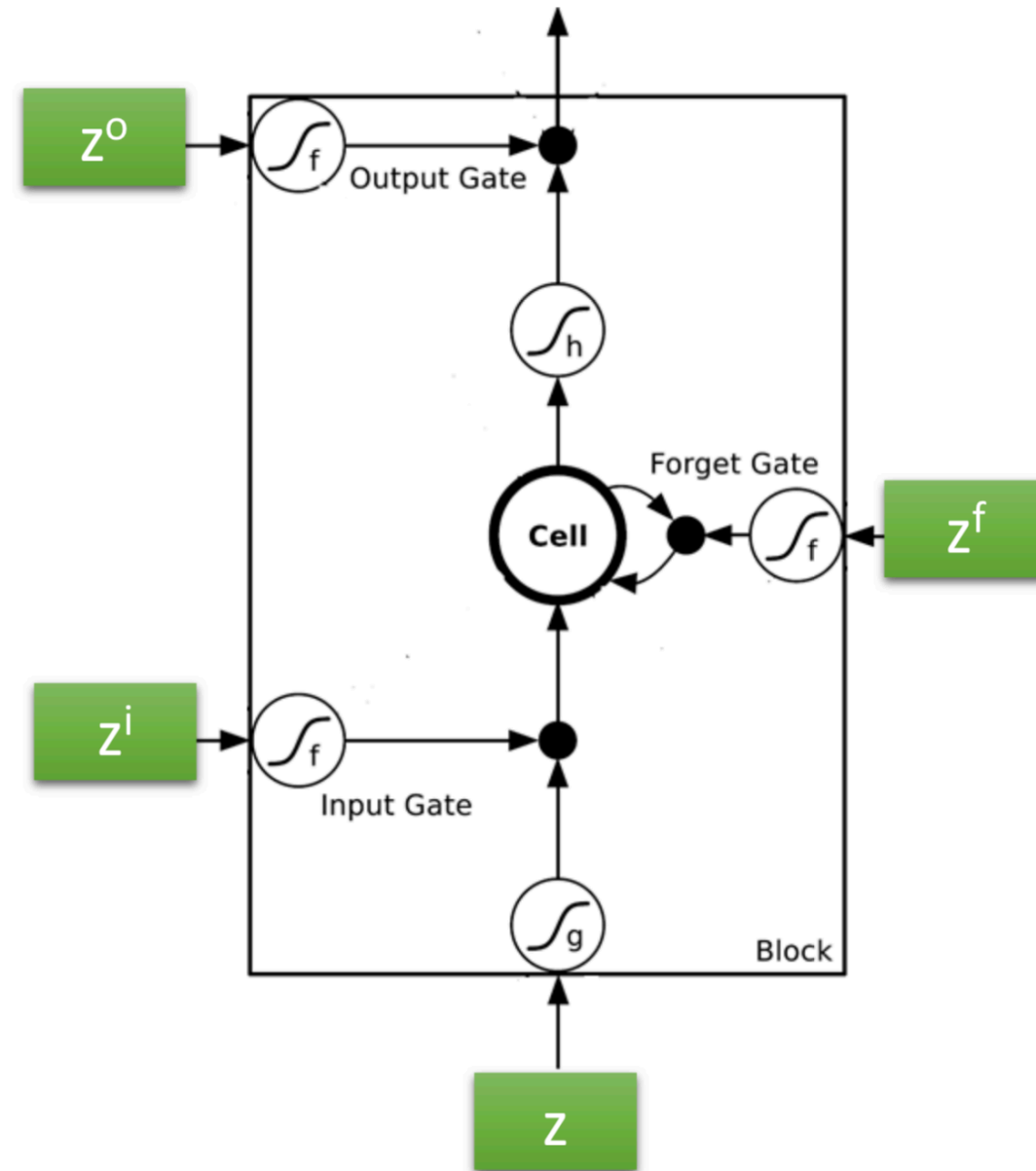
Update Gate = Input Gate + Forget Gate

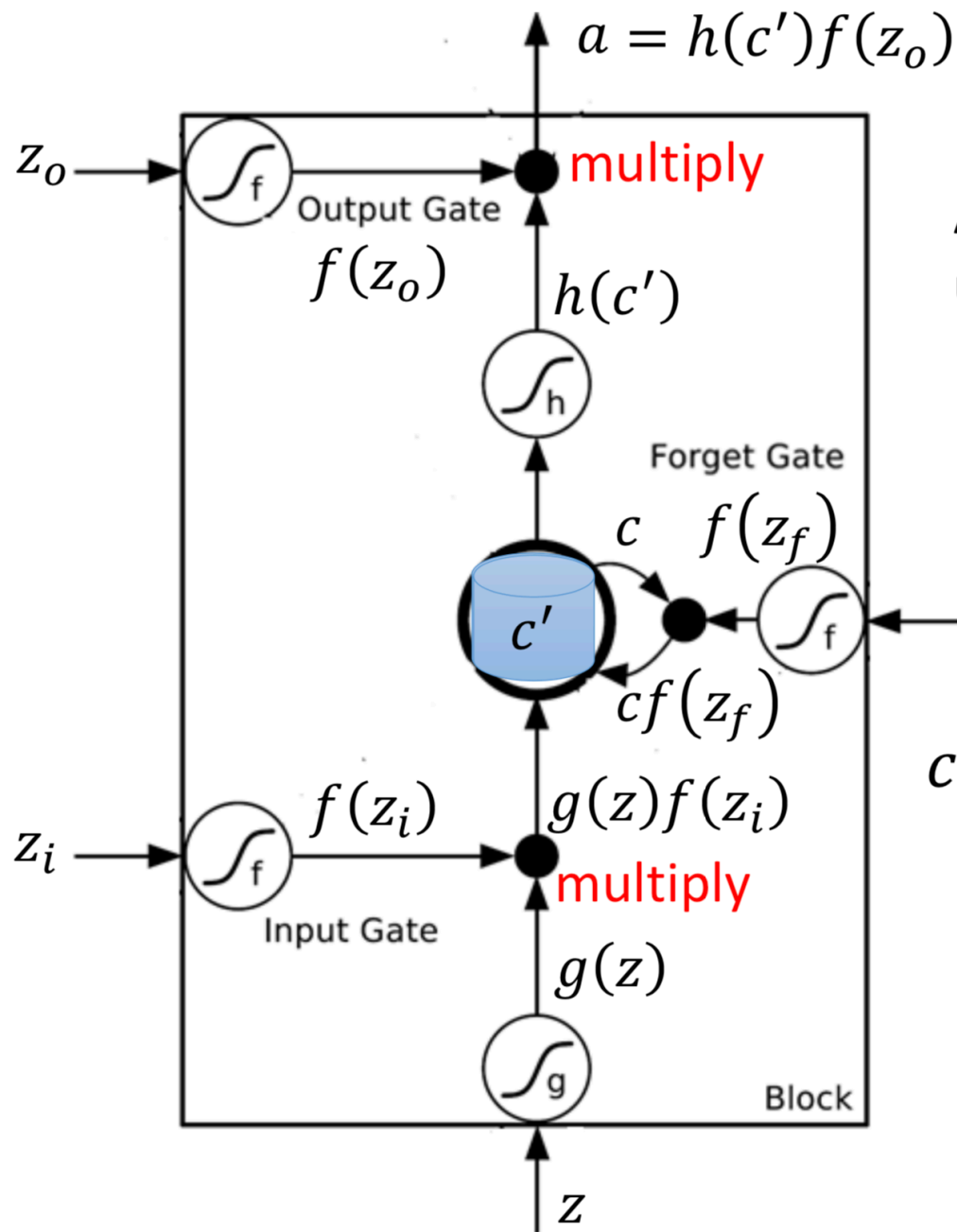
Reset Gate \rightarrow previous hidden state

LSTM



GRU: Input Gate open
then Forget Gate close





Activation function f is usually a sigmoid function

Between 0 and 1

Mimic open and close gate

$$c' = g(z)f(z_i) + cf(z_o)$$