

Encode

Attention

Decode

1: masked\_fill(mask\_s: B, 1, Ss| -inf)  
2: softmax(dim=-1)

alphas: B, 1, Ss

scores: B, 1, Ss

unsqueeze(1)  
squeeze(2)

scores: B, Ss, 1

proj\_key: B, Ss, H

Linear(2H,H)

enc\_hidden: B, Ss, 2H

BiRNN(Es, H)

embs\_s: B, Ss, Es

Embedding(Vs, Es)

s: B, Ss

t: B, St

Embedding(Vt, Et)

embs\_t: B, St, Et

for i in max\_len:  
embs\_t[:, i] -> unsqueeze(1)

embs\_prev: B, 1, Et

context: B, 1, 2H

pre\_output\_vectors: B, max\_len, H

for i in max\_len:  
append(pre\_output)

pre\_output: B, 1, H

2: Linear(Et+H+2\*H, H)  
1: Dropout

pre\_output: B, 1, Et+2H+H

decoder\_states: B, max\_len, H

for i in max\_len:  
append(output)

output: B, 1, H

RNN(Et+2H, H)

decoder\_input: B, 1, Et+2H

cat

bmm

3. Linear(H,1)  
2: tanh  
1: broadcast +

query: B, 1, H

Linear(H,H)

query: B, 1, H

unsqueeze(1)  
last layer

dec\_hidden: L, B, H

2: tanh  
1: Linear(2H,H)

enc\_final: L, B, 2H

cat