后厂理工学院
houchangtech.com

HCT NLP Week 5

问答摘要与推理-
项目模型算法提升

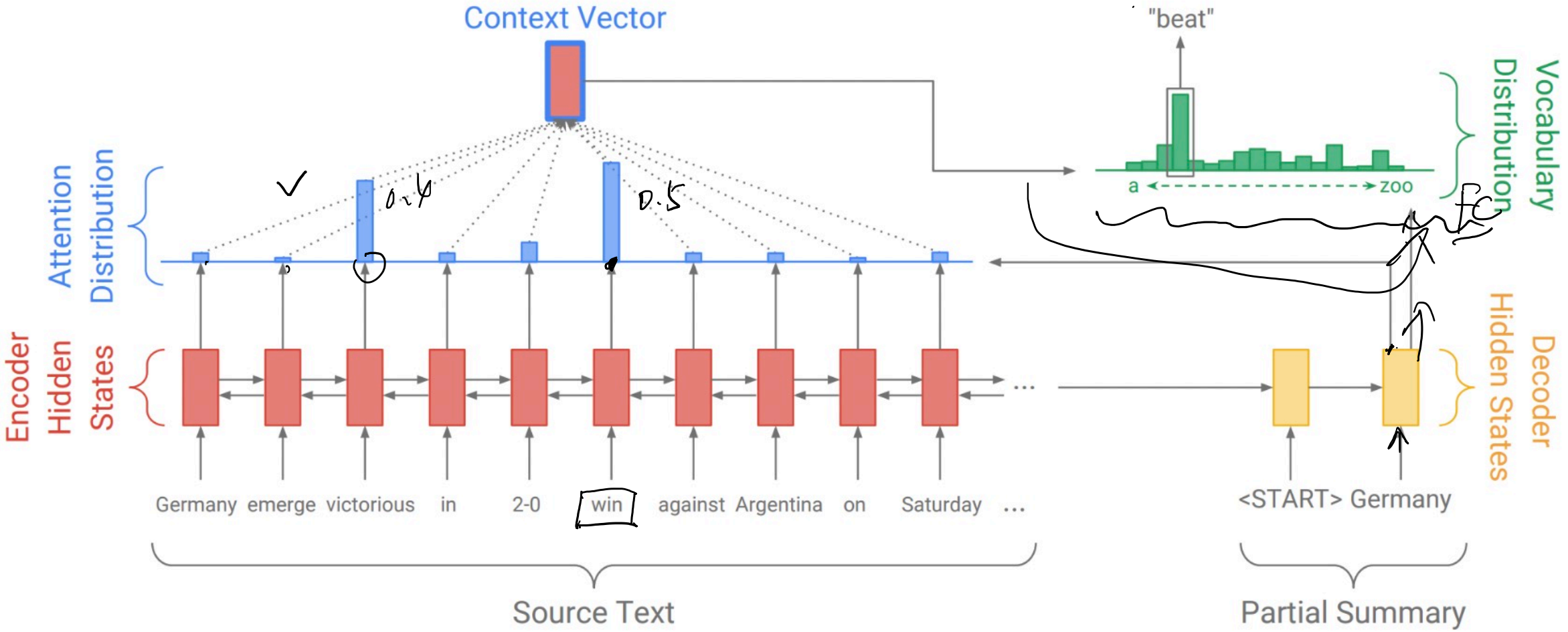
Outline

- OOV 和Word-repetition解决
- Training Strategies
- 抽提式文本摘要基本方法
- 相关代码实践

Outline

- OOV 和Word-repetition解决
- Training Strategies
- 抽提式文本摘要基本方法
- 相关代码实践

典型的seq2seq


Context Vector
"beat"
Vocabulary Distribution
Attention Distribution
Encoder Hidden States
Decoder Hidden States
a
zoo
0.4
0.5
Germany emerge victorious in 2-0 win against Argentina on Saturday ...
<START> Germany
Source Text
Partial Summary

问题

The encoder is not well trained via back propagation through time.

从模型的路径上看，encoder到实际输出有一定距离，从此限制了反向传播。

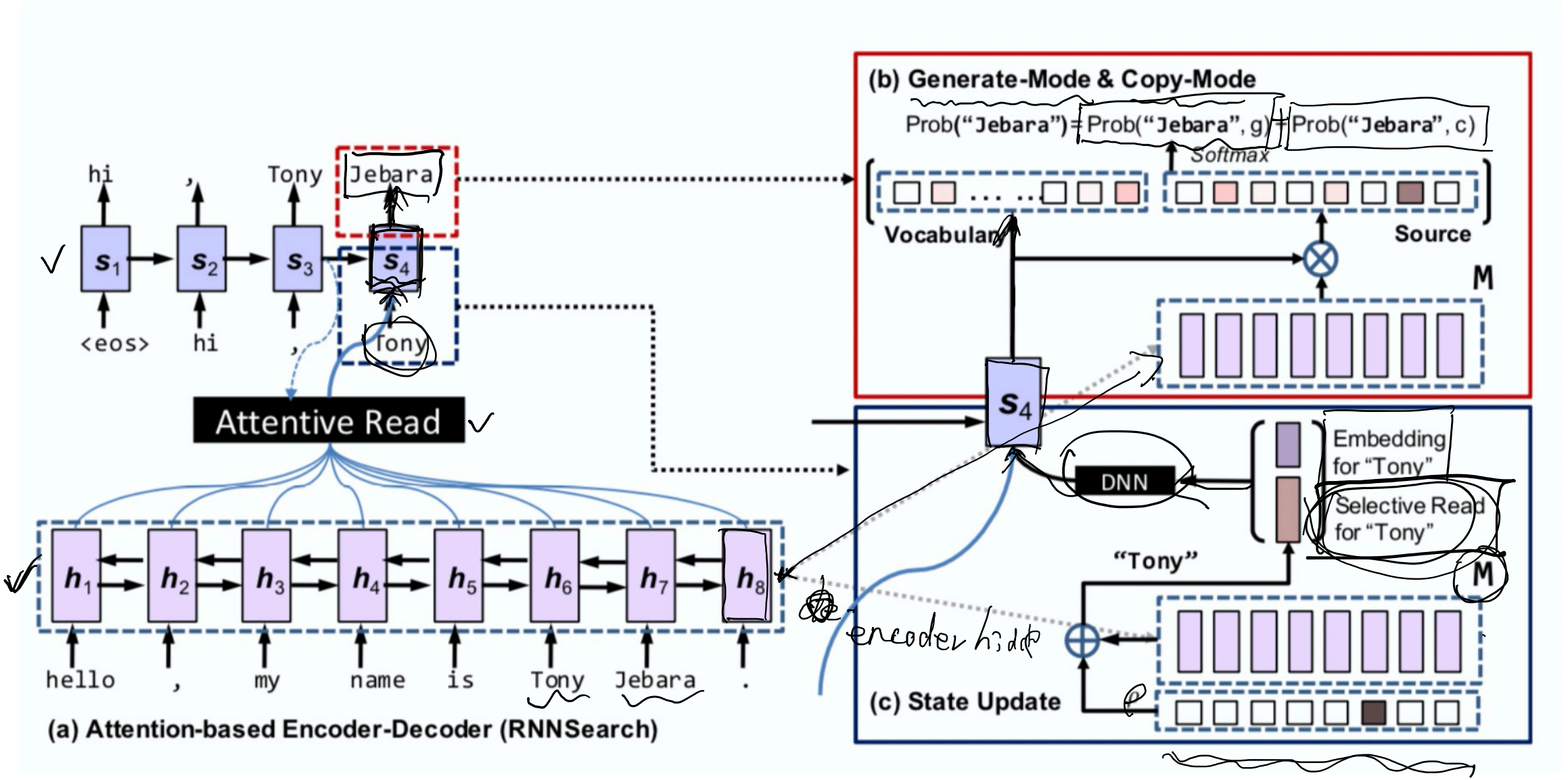
OOV（Out-of-vocabulary未登录词）

摘要总结的结果有的时候并不准确，比如摘要的结果可能输出德国队以2-1比分击败阿根廷，但是实际比分是2-0，出现这个的原因是out-of-vocabulary words（OOV）的出现

Word-repetition问题

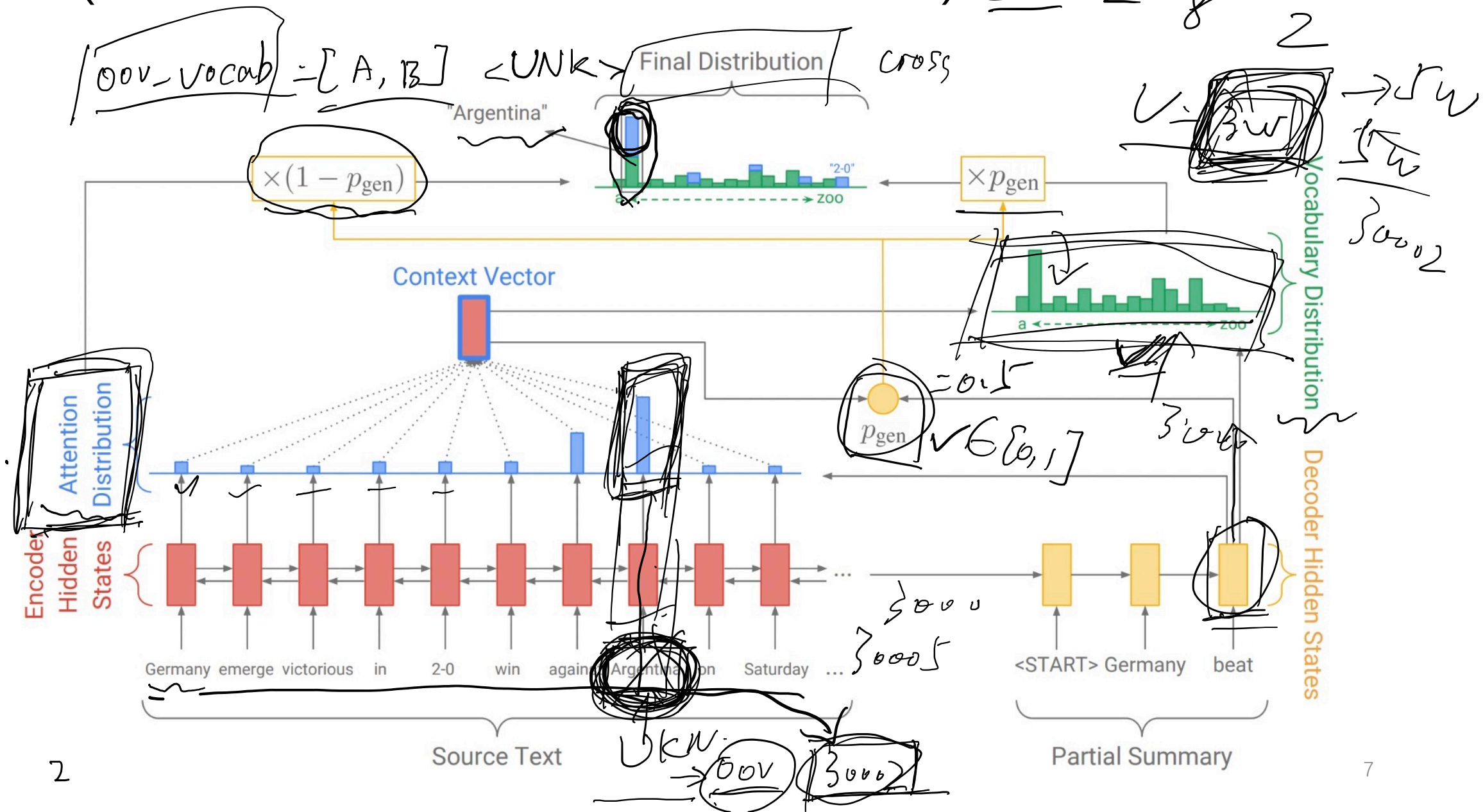
摘要结果会出现repeat重复的信息，比如重复出现德国队击败阿根廷队

CopyNet

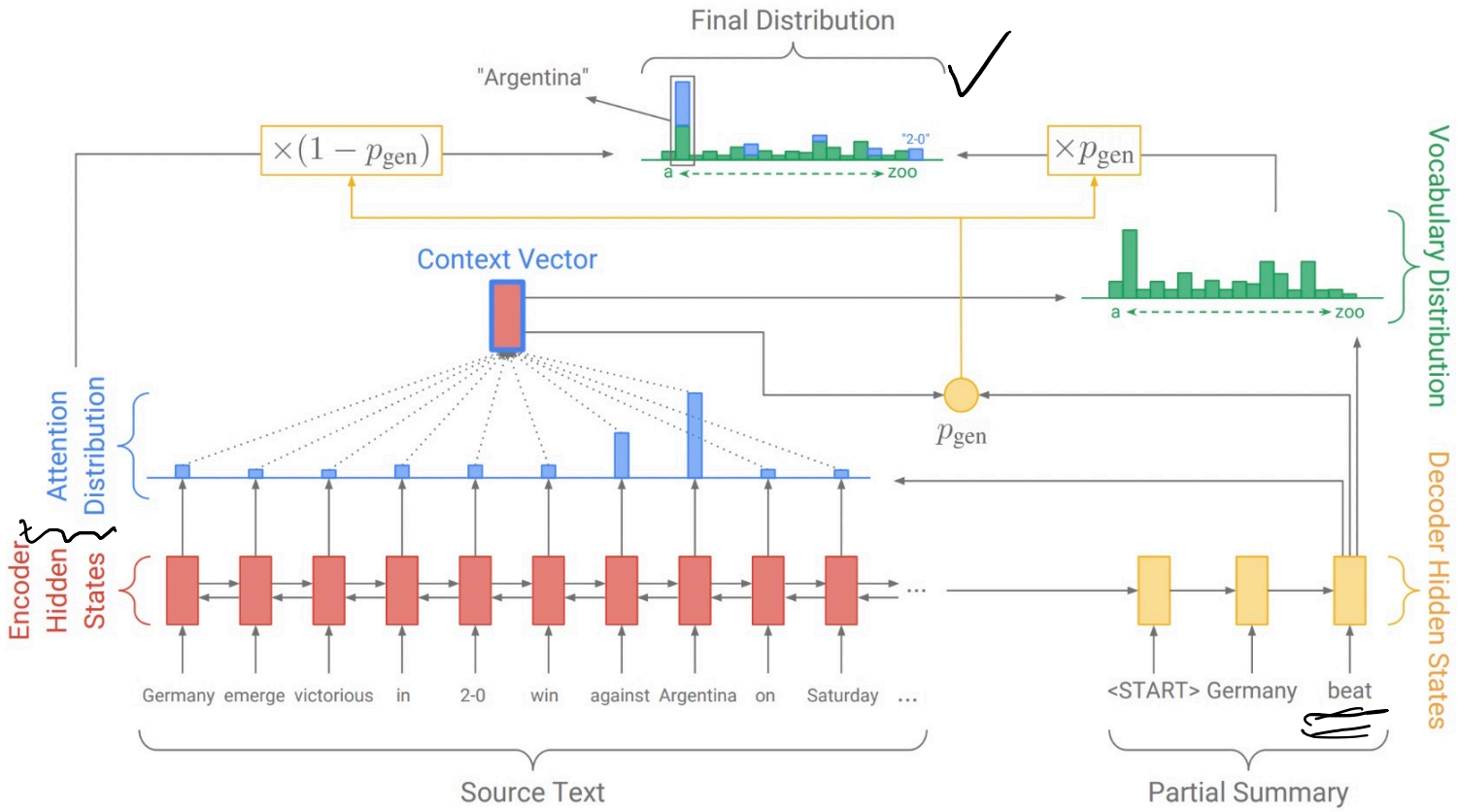


Incorporating Copying Mechanism in Sequence-to-Sequence Learning

PGN（Pointer-Generator Networks）



PGN（Pointer-Generator Networks）



$P_{gen} = \sigma(w_h^T h_t^* + w_s^T s_t + w_x^T x_t + b)$

w_1 w_2 w_3

context vector　decoder hidden　input

$P(w) = P_{gen} P_{vocab}(w) + (1 - P_{gen}) \sum_{i: w_i = w} \alpha_i^t$

1. pointer-generator network能够很容易的复制输入的文本内容，可以通过Pgen 来调节。

2. pointer-generator network能够从输入的文本内容中复制OOV词汇，这是最大的优点，这个也可以采用更小的词汇表vocabulary，较少计算量和存储空间。

3. pointer-generator network训练会更快，在seq2seq训练过程中用更少的迭代次数就能取得一样的效果。

Get To The Point: Summarization with Pointer-Generator Networks

Repetition Handling

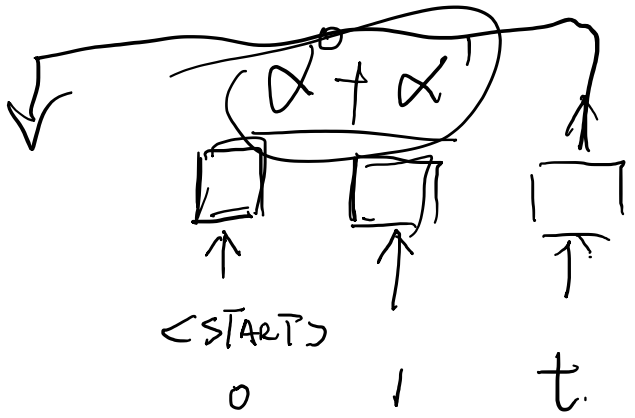
model generated summaries suffer from both word-level and sentence-level repetitions.

Temporal Attention ✓

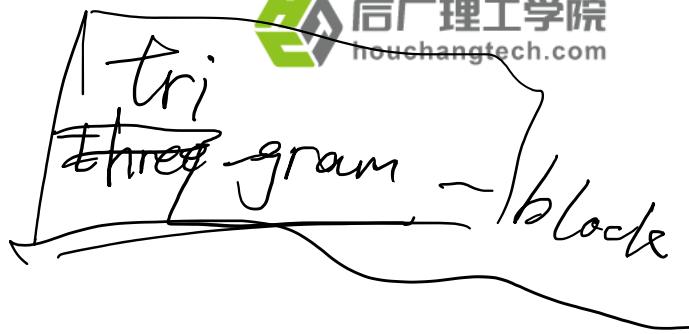
Intra-decoder Attention ✓

Coverage

Coverage



c^t

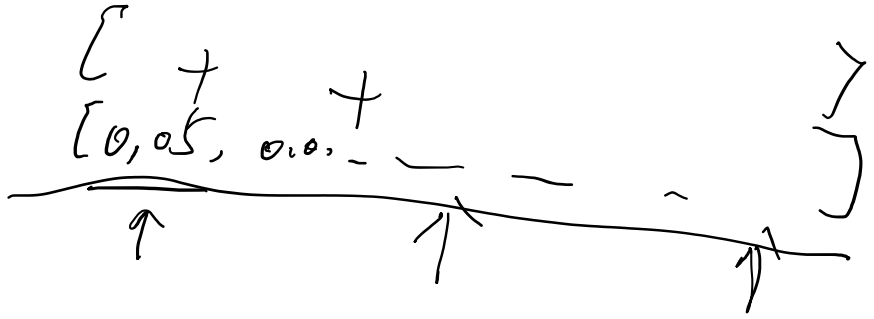


① $c^t = \sum_{t'=0}^{t-1} \alpha^{t'}$

$c^0 = 0 \quad t'=1$

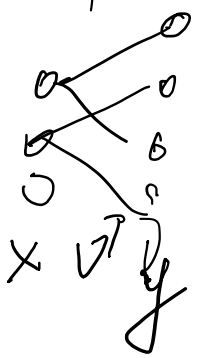
② $e_i^t = v^T \tanh(W_h h_i + W_s s_t + W_c c_i^t + b)$

tf.keras.layers.Dense(1)



③

batch_size, dim

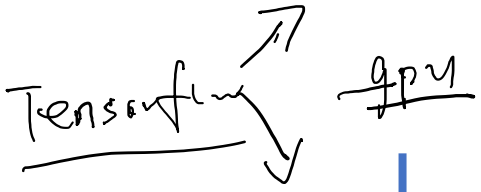


$covloss_t = \sum_i \min(\alpha_i^t, c_i^t)$

[batch_size, 2x0]

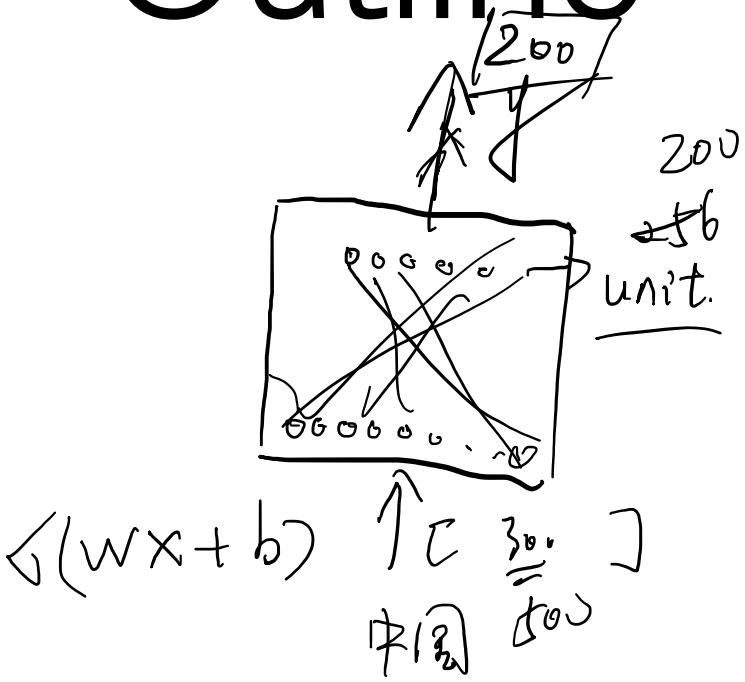
$= 0.5$

$loss_t = -\log P(w_t^*) + \lambda \sum_i \min(\alpha_i^t, c_i^t)$



20:57见

Outline



- OOV 和Word-repetition解决
- Training Strategies
- 抽提式文本摘要基本方法
- 相关代码实践

TRAINING STRATEGIES

A. Word-Level Training

two different methods for avoiding the problem of exposure bias.

1) Cross-Entropy Training (XENT)

↳ Teacher Forcing

2) Scheduled Sampling

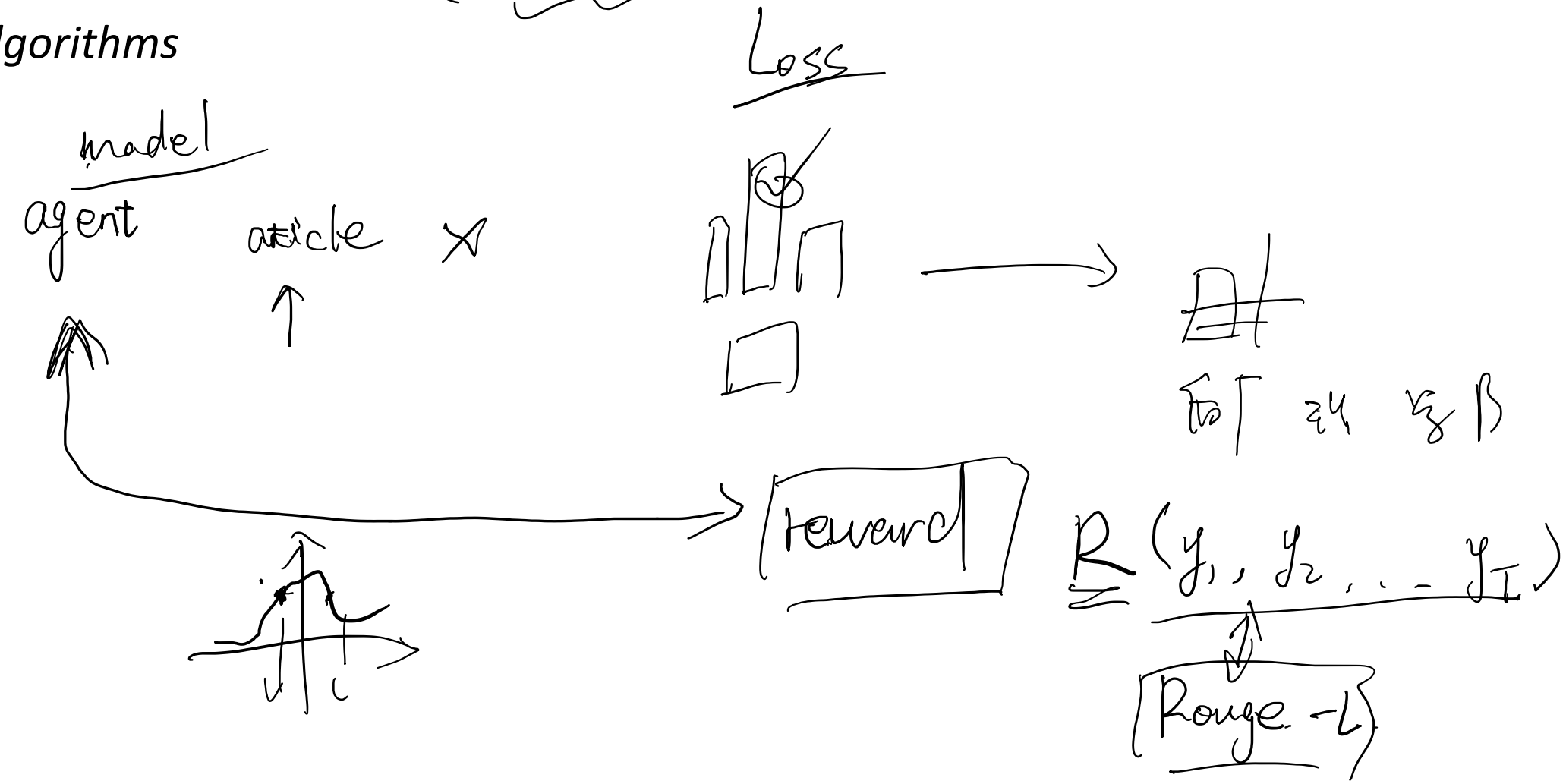
是一种解决训练和生成时输入数据分布不一致的方法。在训练早期该方法主要使用目标序列中的真实元素作为解码器输入，可以将模型从随机初始化的状态快速引导至一个合理的状态。随着训练的进行，该方法会逐渐更多地使用生成的元素作为解码器输入，以解决数据分布不一致的问题。该方法应用在模型的训练阶段，生成阶段不使用。

Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks

TRAINING STRATEGIES

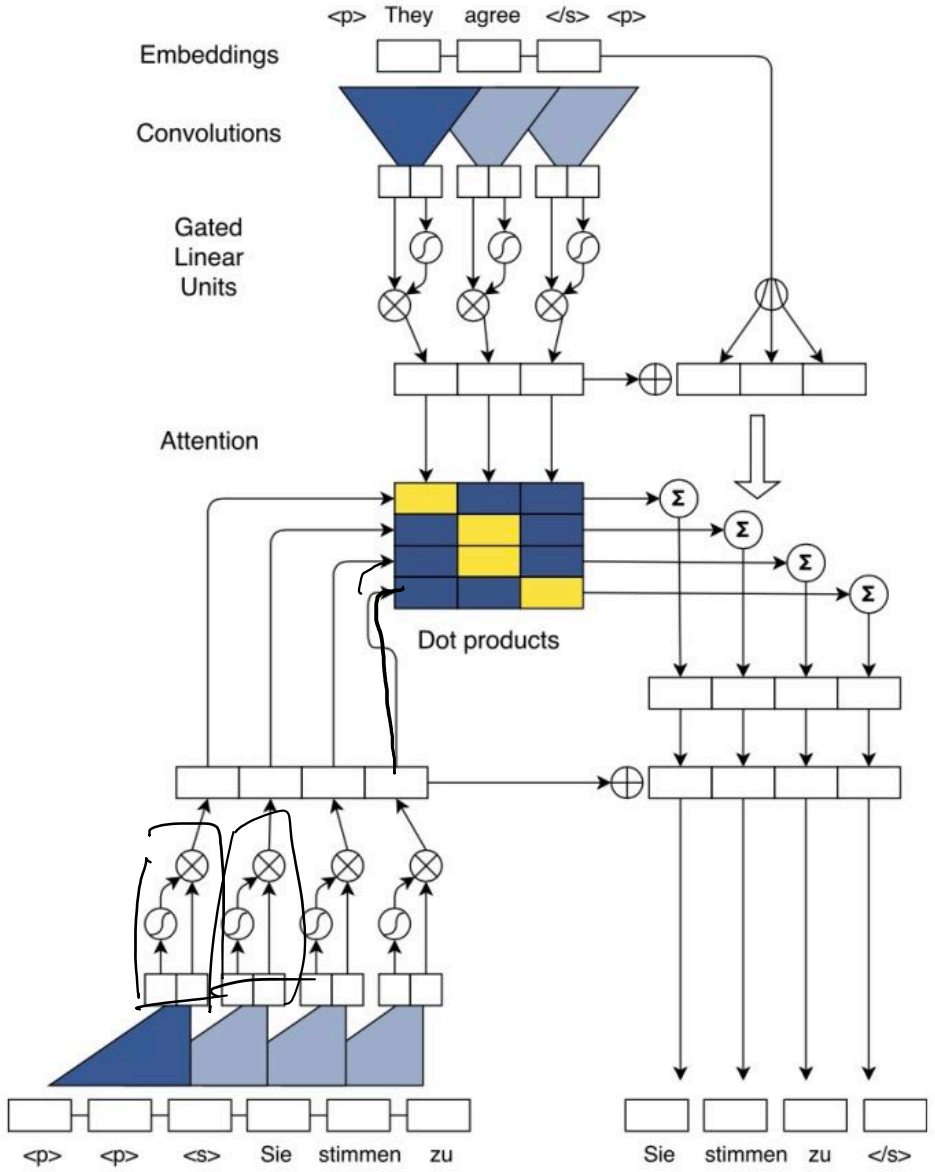
B. Sequence-Level Training

RL algorithms



Beyond RNN

CNN



Position Embedding

层叠CNN构成了hierarchical representation表示

融合了Residual connection、liner mapping的多层attention

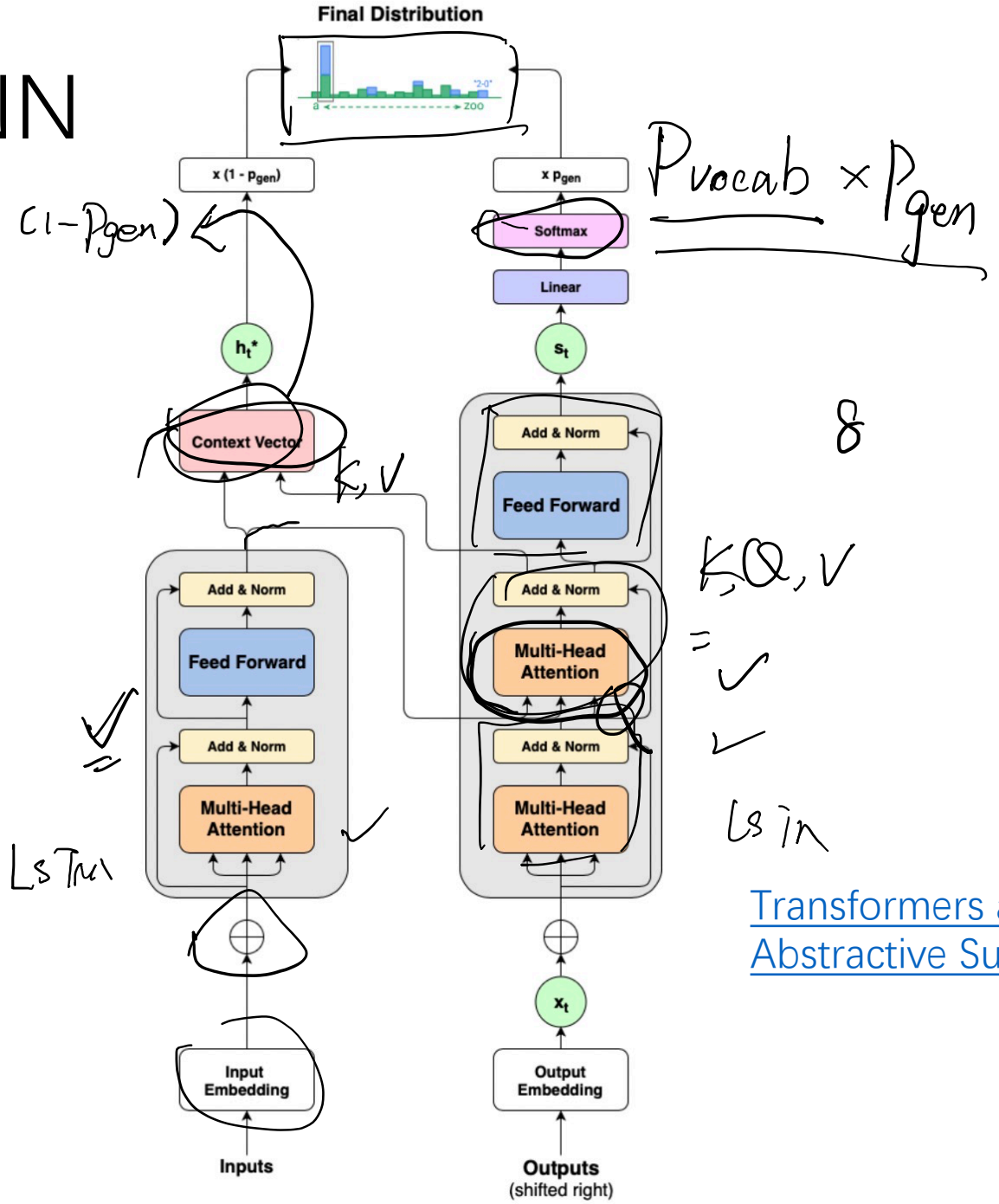
采用GLU做为gate mechanism

进行了梯度裁剪和精细的权重初始化，加速模型训练和收敛

Convolutional Sequence to Sequence Learning

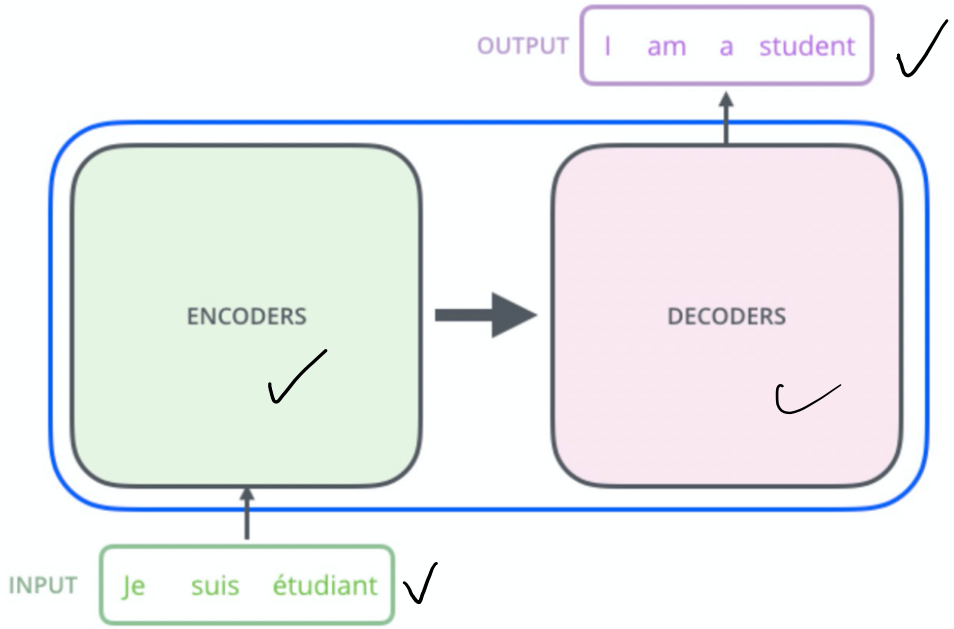
Beyond RNN

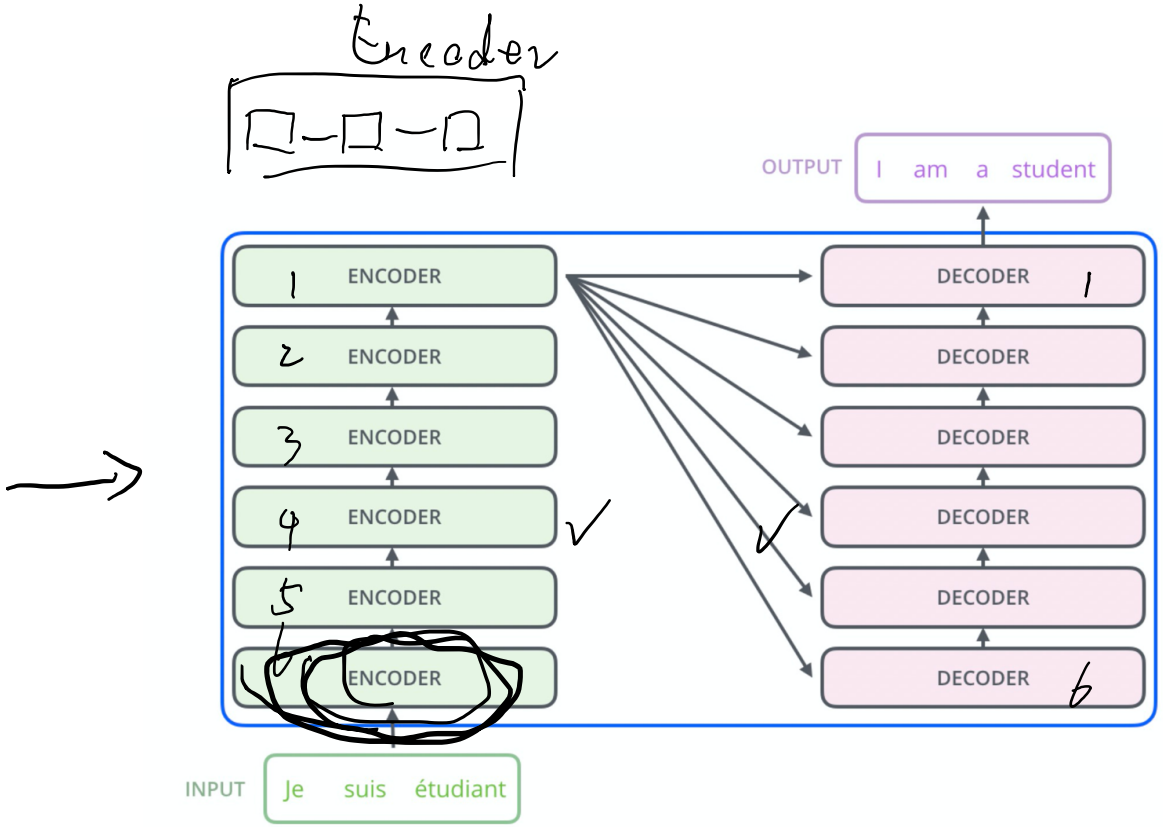
Transformer

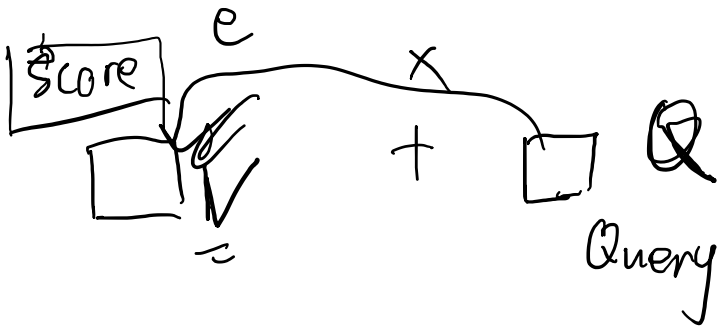


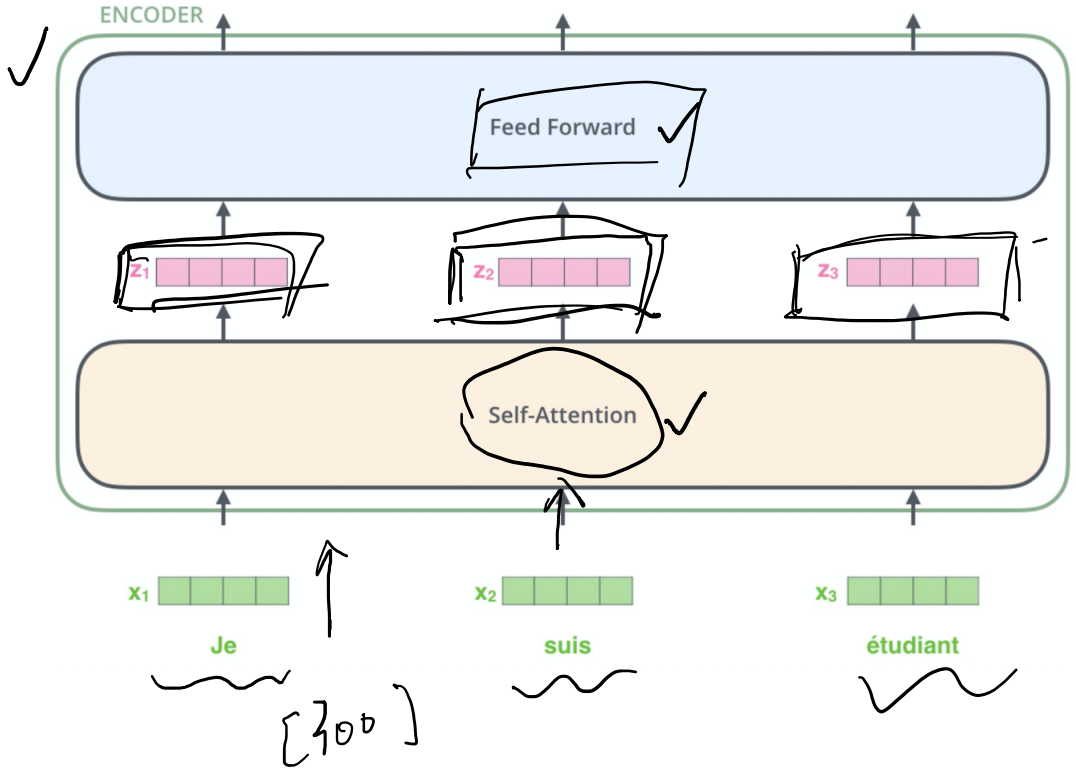
[Transformers and Pointer-Generator Networks for Abstractive Summarization]

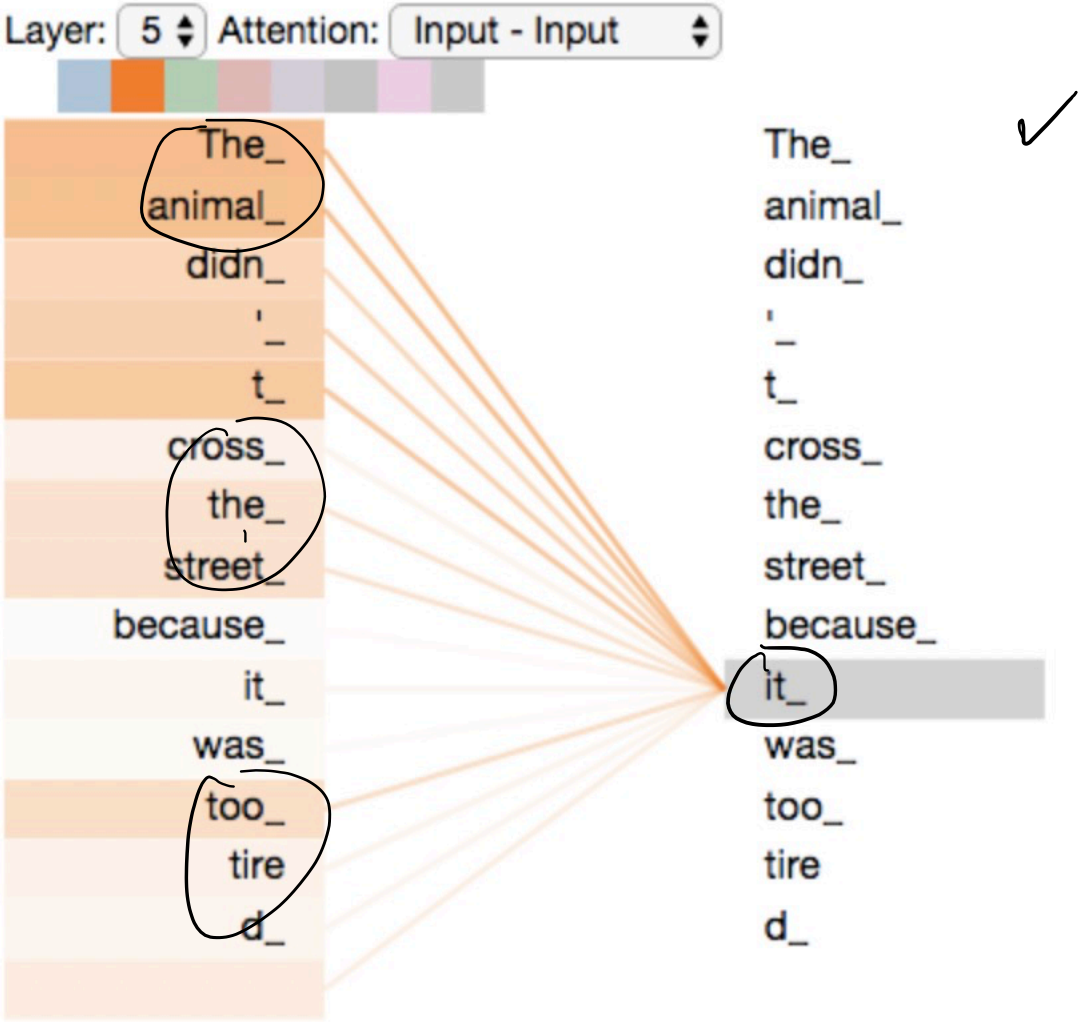

INPUT
Je suis étudiant
THE TRANSFORMER
OUTPUT
I am a student

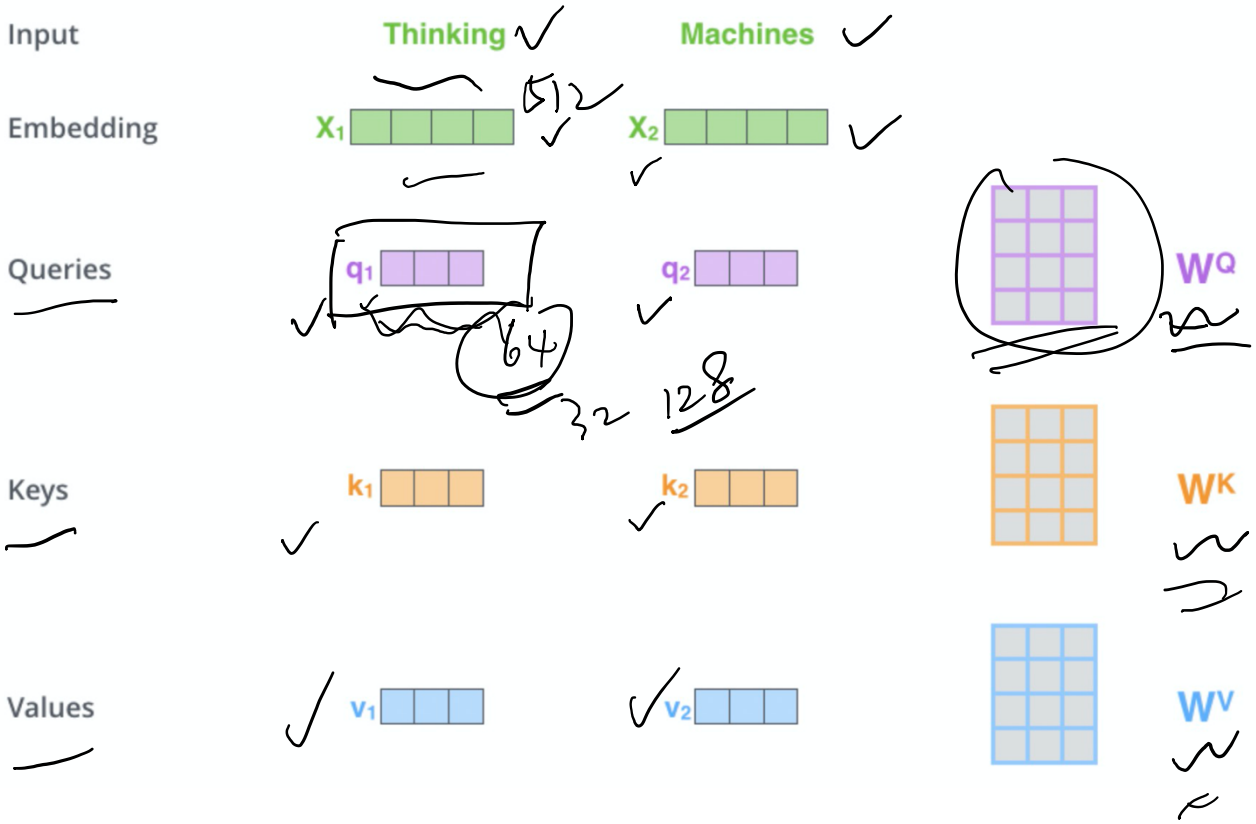

OUTPUT
I am a student
ENCODERS
DECODERS
INPUT
Je suis étudiant

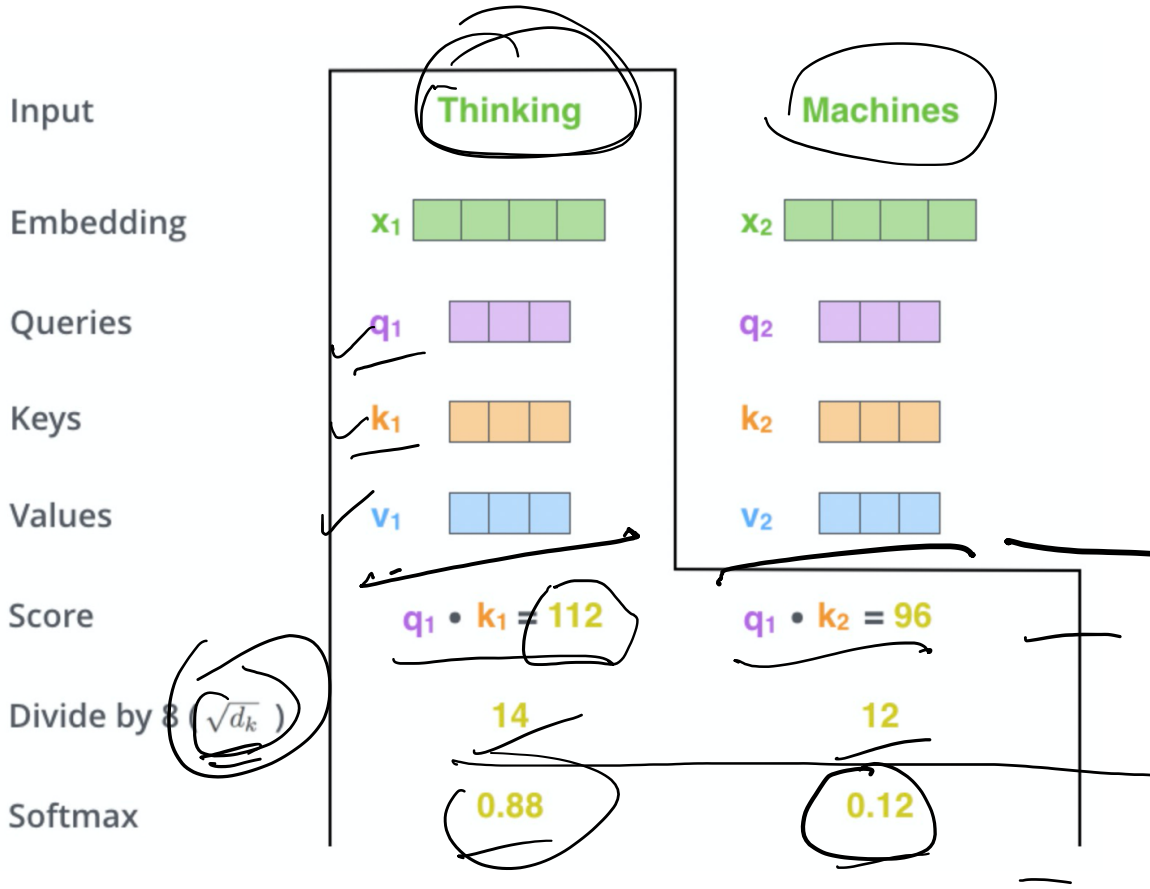

Encoder
OUTPUT
I am a student
ENCODER
ENCODER
ENCODER
ENCODER
ENCODER
ENCODER
DECODER
DECODER
DECODER
DECODER
DECODER
DECODER
INPUT
Je suis étudiant

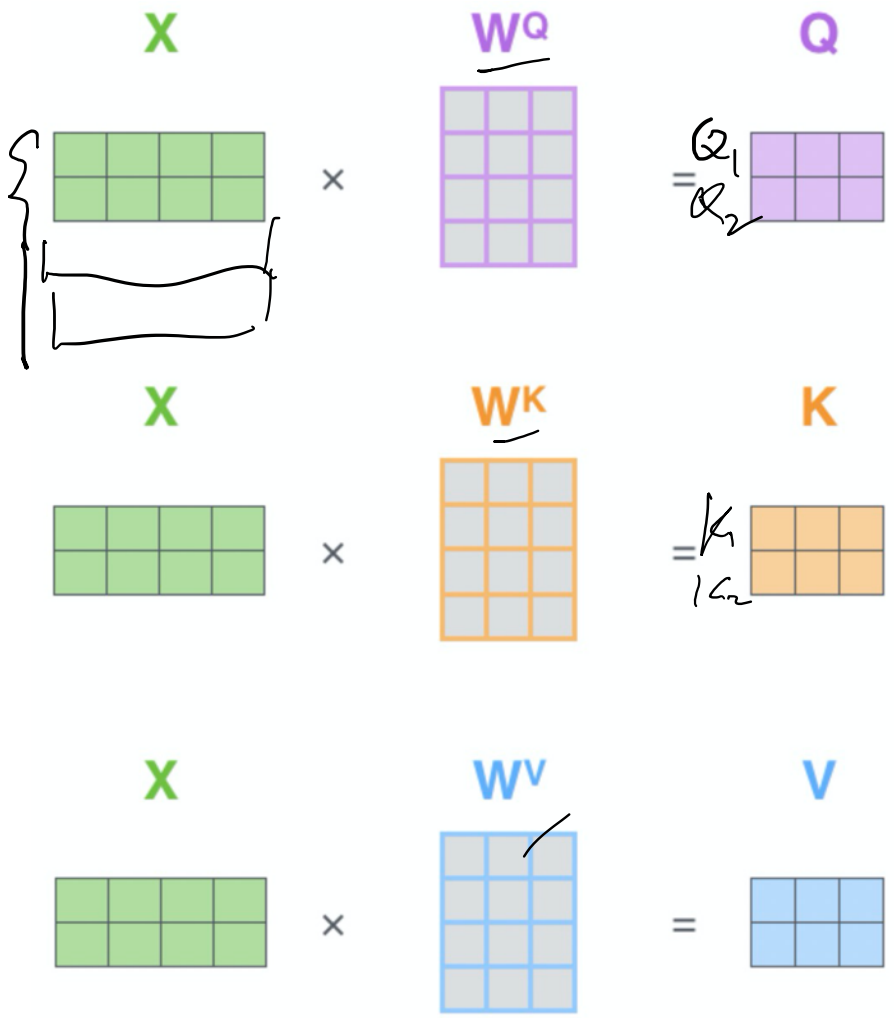

score
e
x
+
Q
Query

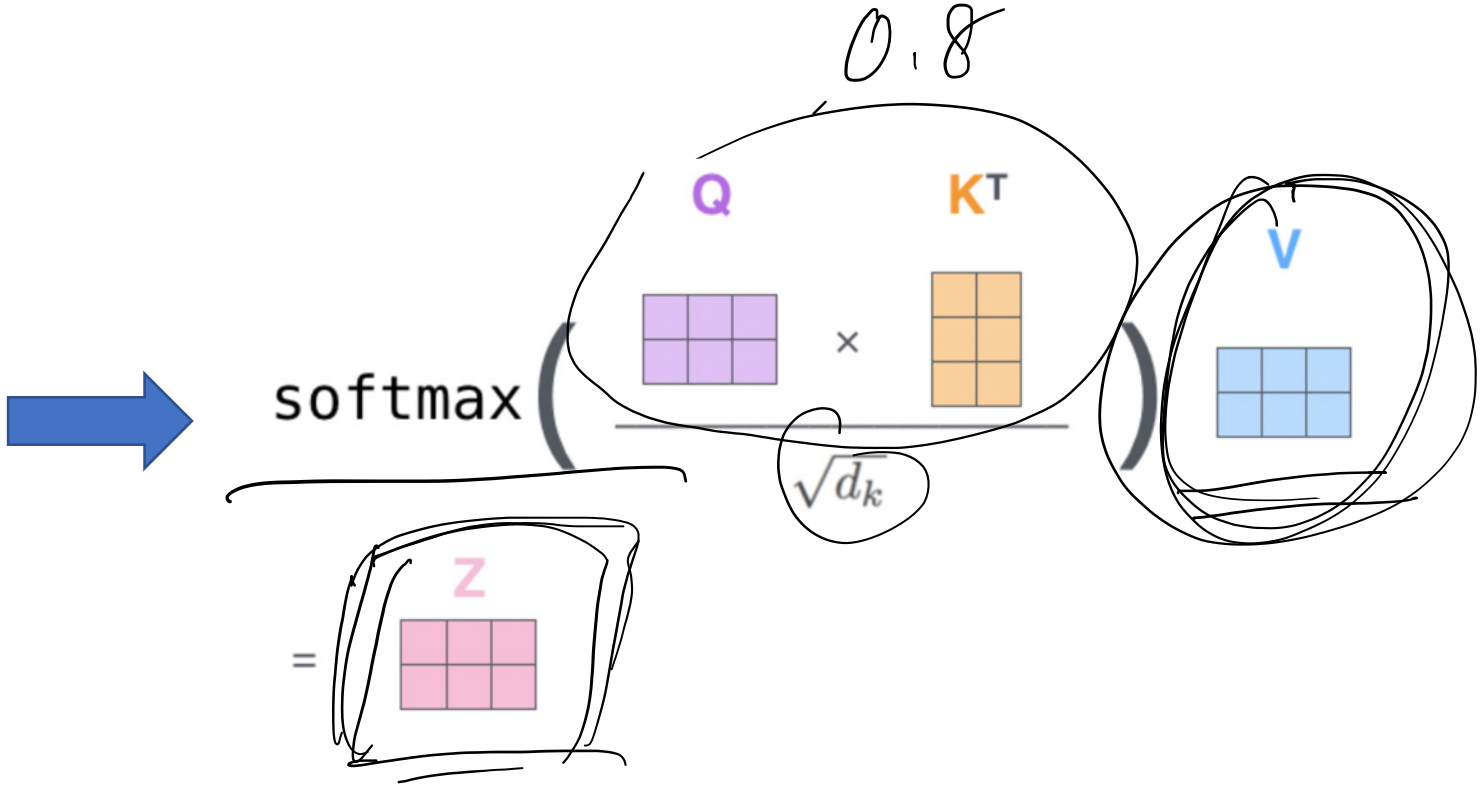

ENCODER
Feed Forward
Self-Attention
z1
z2
z3
x1
x2
x3
Je
suis
étudiant


Layer: 5
Attention: Input - Input
The_
animal_
didn_
'_
t_
cross_
the_
street_
because_
it_
was_
too_
tire
d_

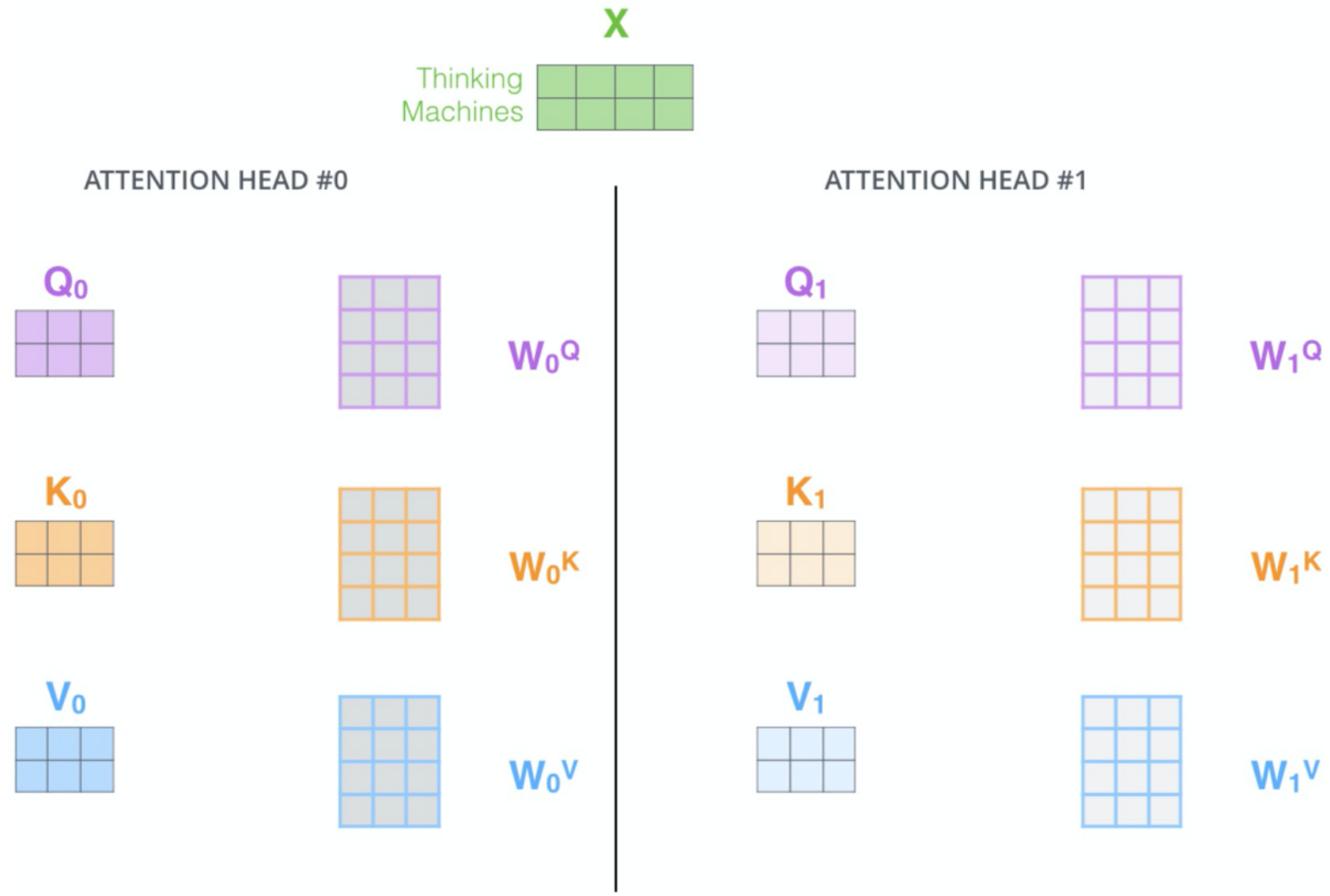

Input
Thinking
Machines
Embedding
X_1
X_2
Queries
q_1
q_2
Keys
k_1
k_2
Values
v_1
v_2
W^Q
W^K
W^V


Input
Thinking
Machines
Embedding
x_1
x_2
Queries
q_1
q_2
Keys
k_1
k_2
Values
v_1
v_2
Score
$q_1 \cdot k_1 = 112$
$q_1 \cdot k_2 = 96$
Divide by 8 ($\sqrt{d_k}$)
14
12
Softmax
0.88
0.12


X
W^Q
Q
X
W^K
K
X
W^V
V


softmax
Q
K^T
V
$\sqrt{d_k}$
Z

Multi-headed



Multi-headed

1) This is our input sentence*

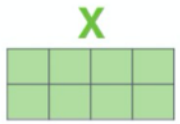
2) We embed each word*

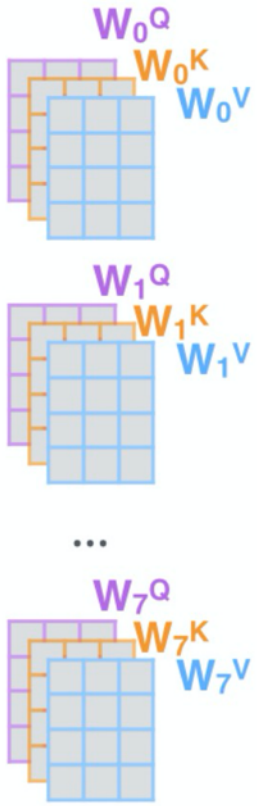
3) Split into 8 heads. We multiply X or R with weight matrices

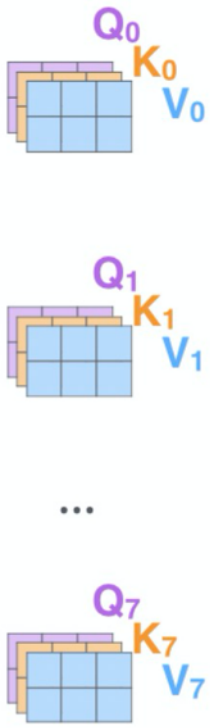
4) Calculate attention using the resulting Q/K/V matrices

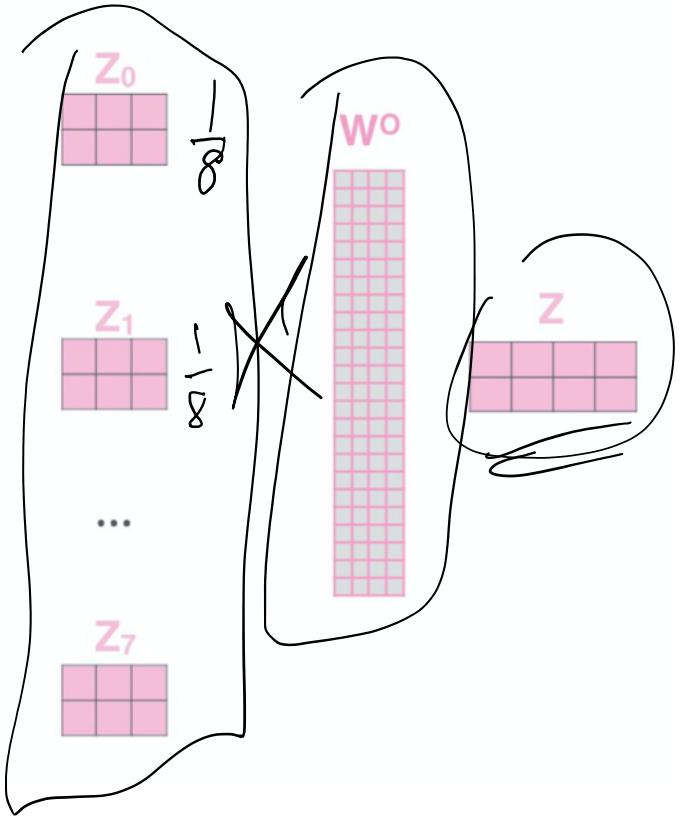
5) Concatenate the resulting Z matrices, then multiply with weight matrix W^O to produce the output of the layer

Thinking Machines







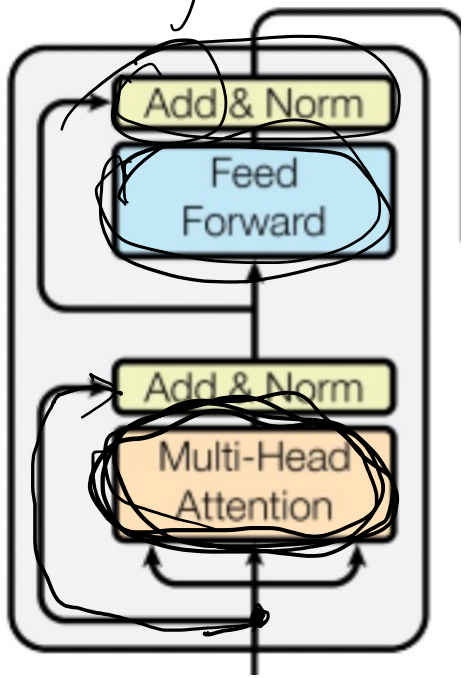


* In all encoders other than #0, we don't need embedding. We start directly with the output of the encoder right below this one

Resnet

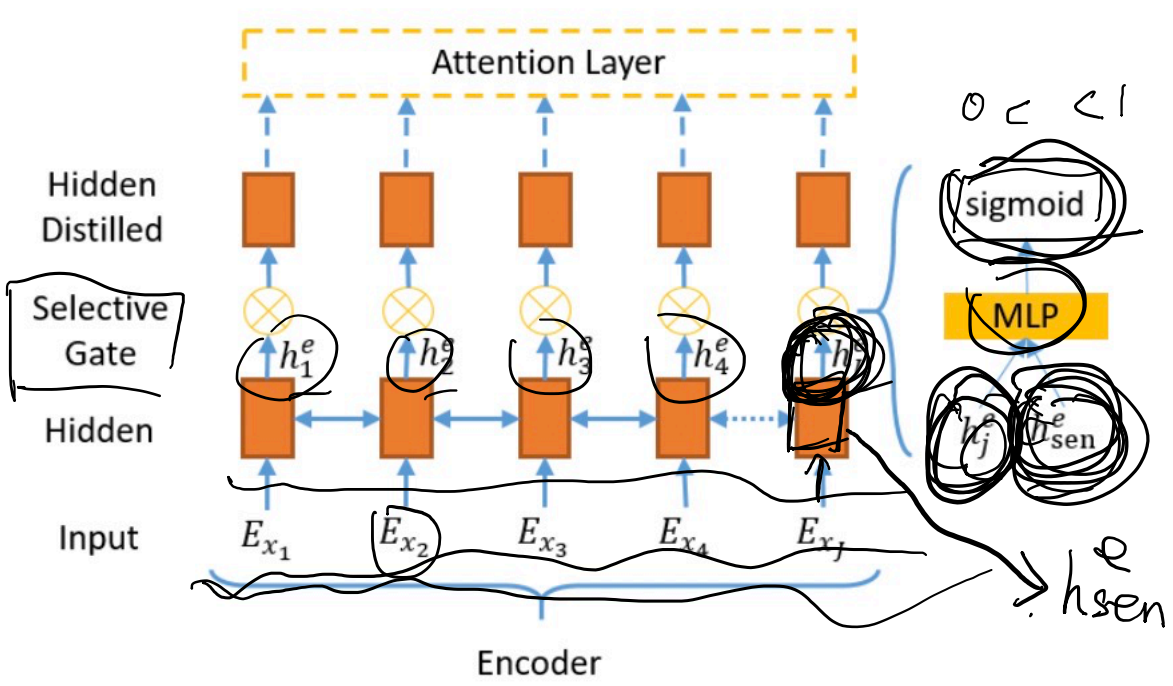
skip

Layer Norm

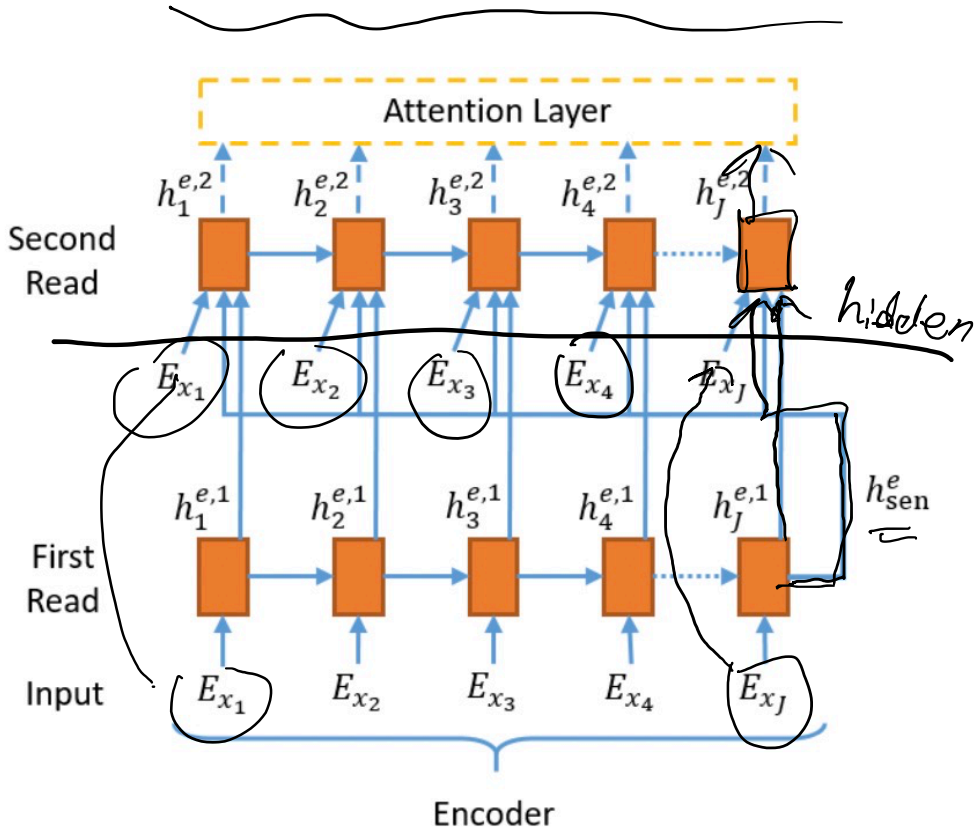


Improving Encoded Representations

Selective Encoding



Read-Again Encoding



Improving Decoder

Embedding Weight Sharing

SUMMARY GENERATION

Diverse Beam Decoding

the top-B hypotheses may differ by just a couple tokens at the end of sequences, which not only affects the quality of generated sequences but also wastes computational resources

Outline

- OOV 和Word-repetition解决
- Training Strategies
- 抽提式文本摘要基本方法
- 相关代码实践

Text summarization

Extractive text summarization

Source Text: Peter and Elizabeth took a taxi to attend the night party in the city.

While in the party, Elizabeth collapsed and was rushed to the hospital.

Summary: Peter and Elizabeth attend party city. Elizabeth rushed hospital.

Abstractive text summarization

Source Text: Peter and Elizabeth took a taxi to attend the night party in the city.

While in the party, Elizabeth collapsed and was rushed to the hospital.

Summary: Elizabeth was hospitalized after attending a party with Peter.

Extractive Text Summarization

tf-idf

输入文本表征的构建

根据表征给句子打分

根据一定数量的句子组合

topic representation

Topic Words
Frequency-driven Approaches
Latent Semantic Analysis
Bayesian Topic Models

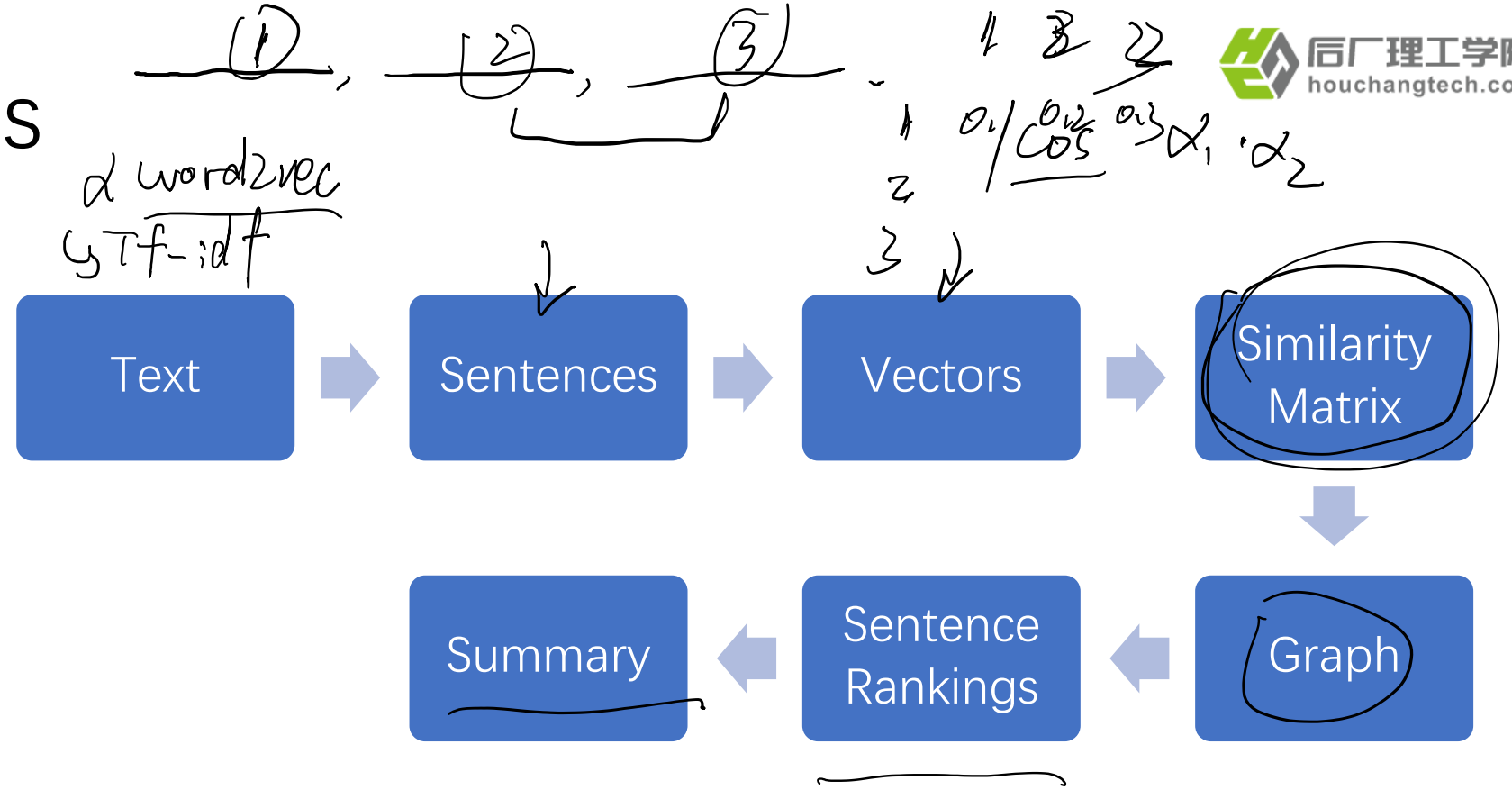
indicator representation

Graph Methods ✓
Machine Learning

Text Summarization Techniques: A Brief Survey

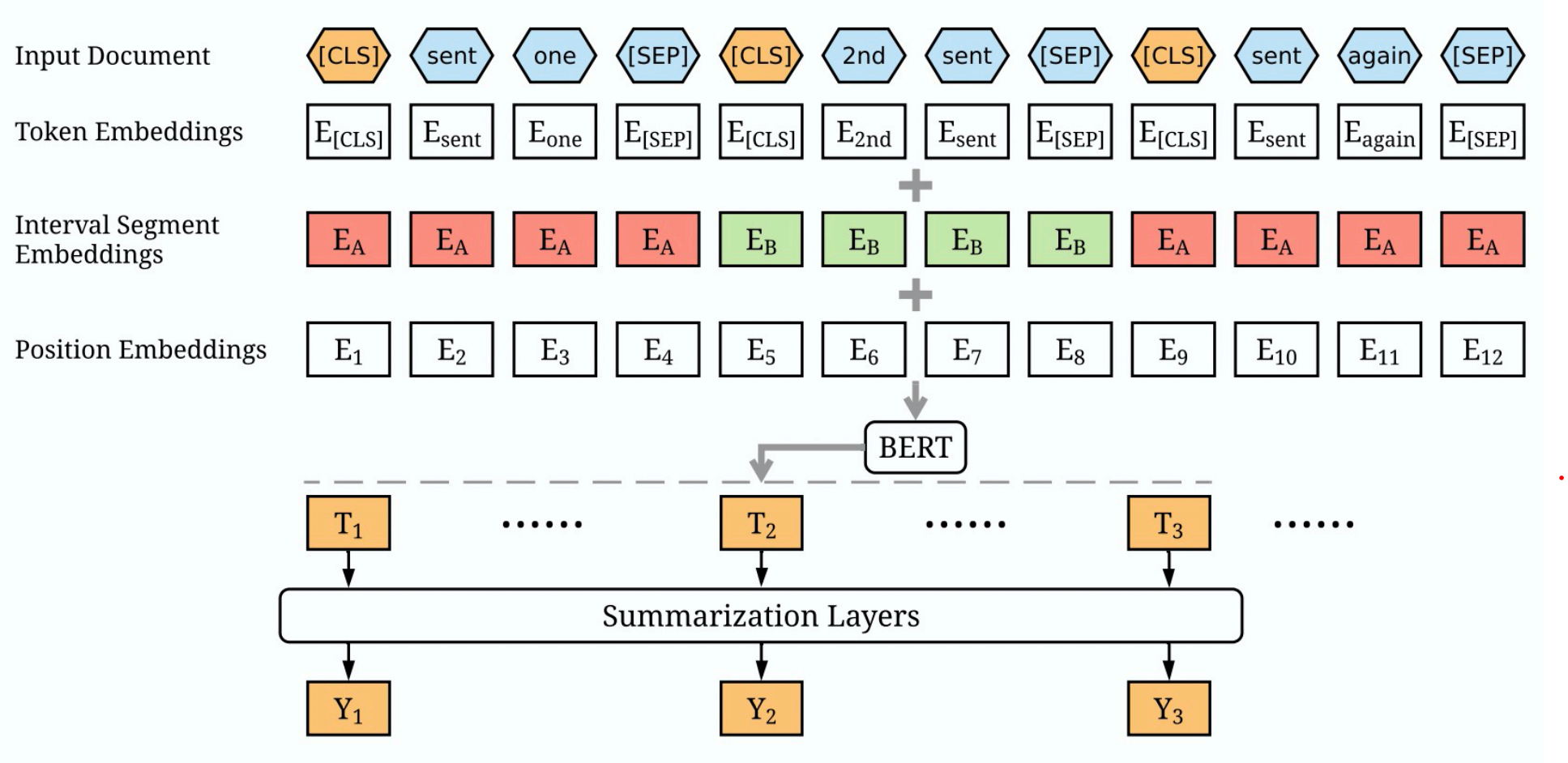
Graph Methods

TextRank算法



1. 第一步是把所有文章整合成文本数据
2. 接下来把文本分割成单个句子
3. 然后，我们将为每个句子找到向量表示（词向量）
4. 计算句子向量间的相似性并存放在矩阵中
5. 然后将相似矩阵转换为以句子为节点、相似性得分为边的图结构，用于句子TextRank计算

BertSum



Fine-tune BERT for Extractive Summarization

Outline

- OOV 和Word-repetition解决
- Training Strategies
- 抽提式文本摘要基本方法
- 相关代码实践

作业

Bye ！