



数据科学与计算机学院
School of Data and Computer Science

自然语言处理

Natural Language Processing

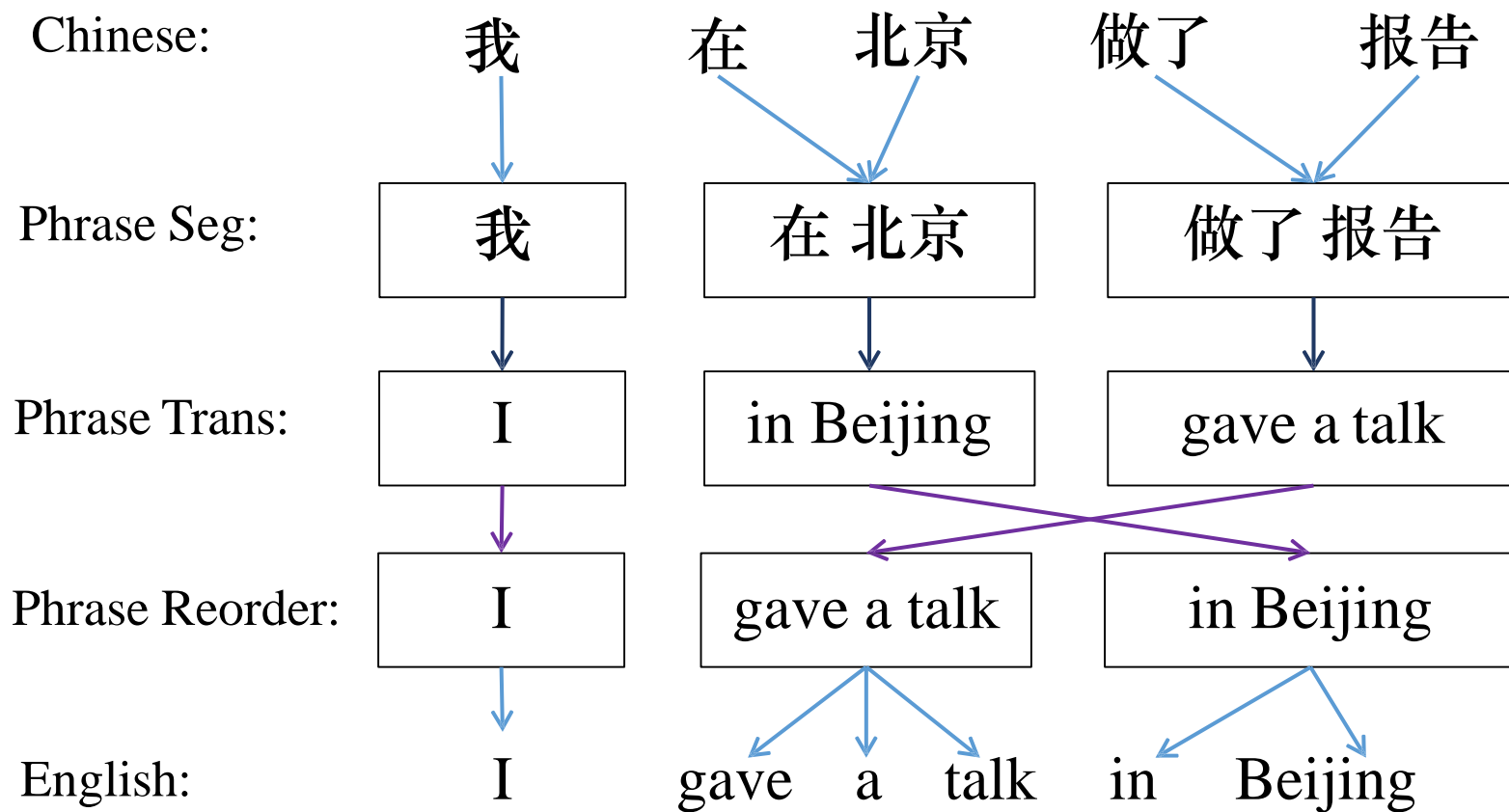
权小军 教授

中山大学数据科学与计算机学院

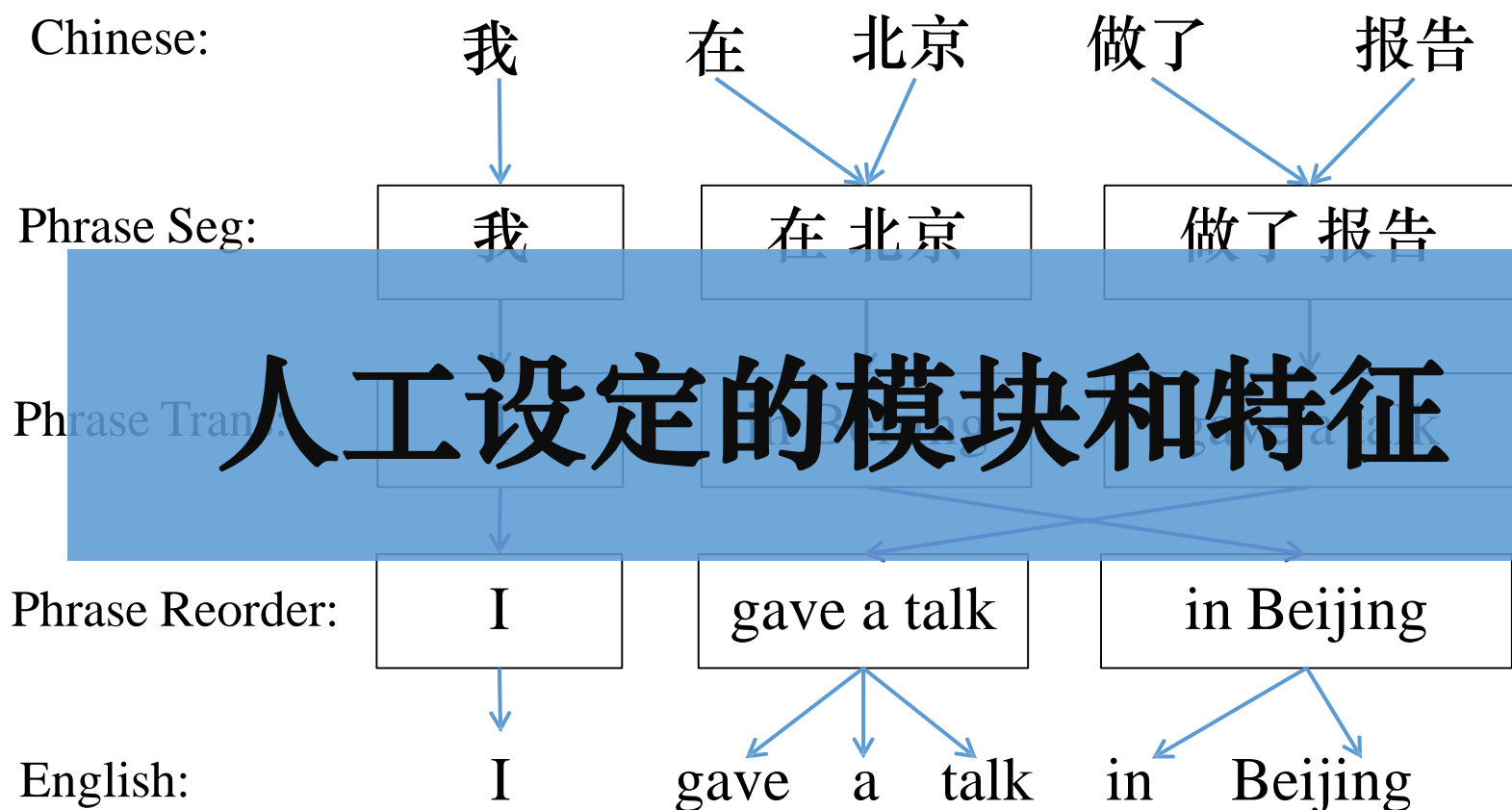
quanxj3@mail.sysu.edu.cn

神经机器翻译

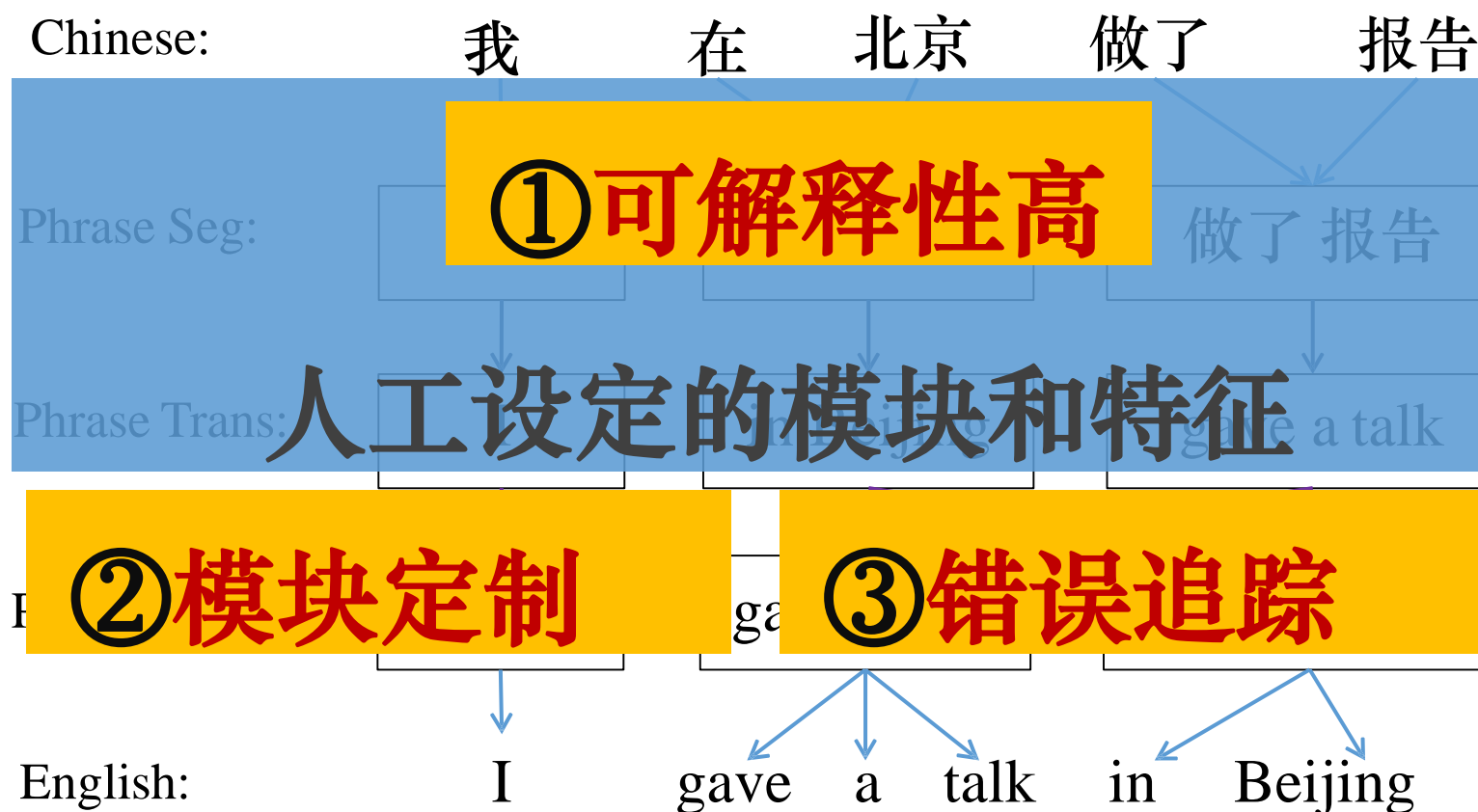
统计机器翻译



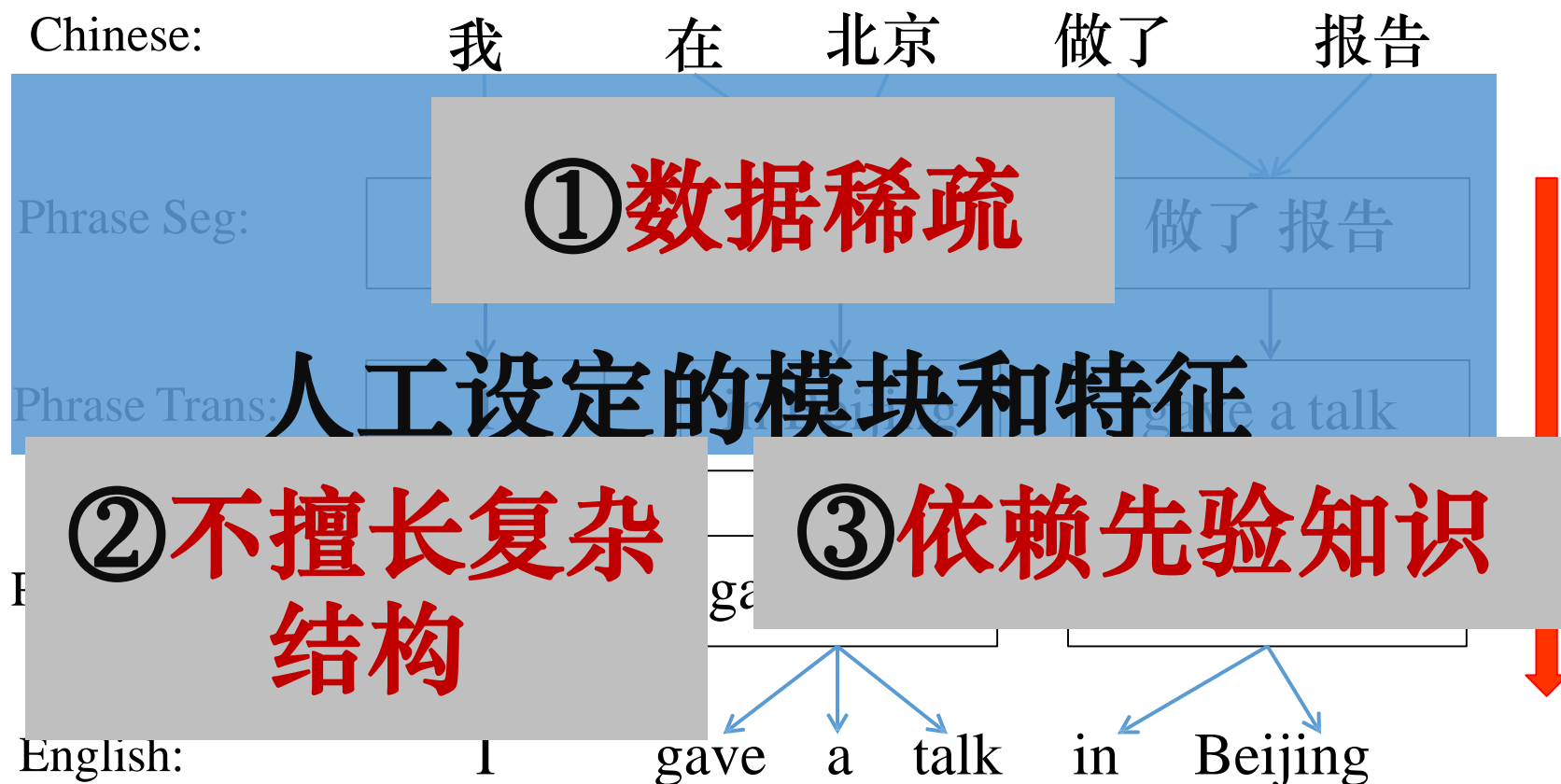
统计机器翻译



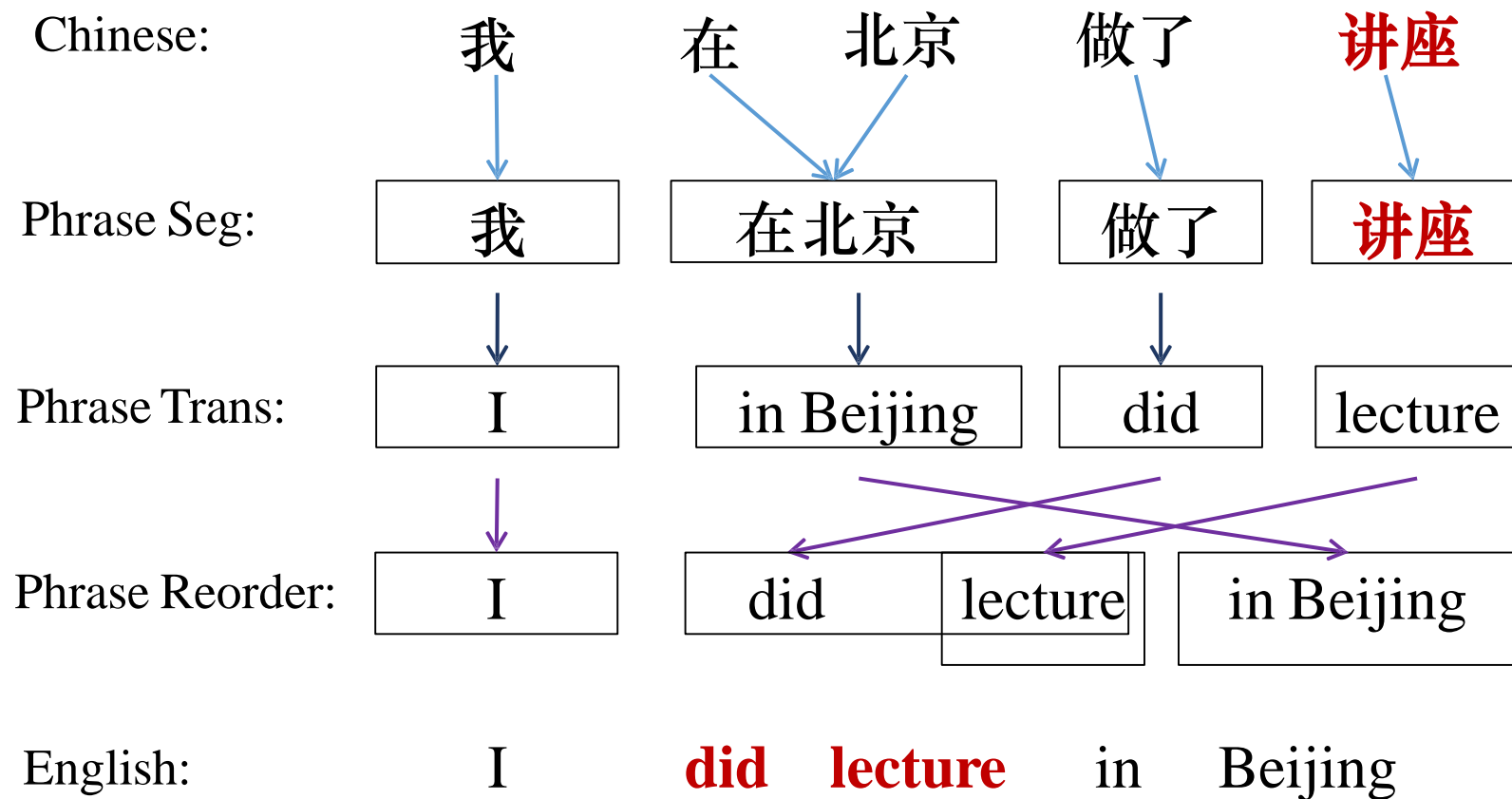
统计机器翻译



统计机器翻译



统计机器翻译



①数据稀疏

统计机器翻译

Chinese

美国总统布什昨天在白宫与以色列总理沙龙就中东局势 举行了一个小时的会谈。 ✕

English

Yesterday, U.S. President George W. Bush at the White House with Israeli Prime Minister Ariel Sharon on the situation in the Middle East held a one-hour talks.

②不擅长复杂结构

统计机器翻译→神经机器翻译

离散符号表示方法

讲座 \otimes 报告 = 0

统计机器翻译→神经机器翻译

离散符号表示方法 \Rightarrow 连续分布式表示方法

$$\text{讲座} \otimes \text{报告} = 0$$

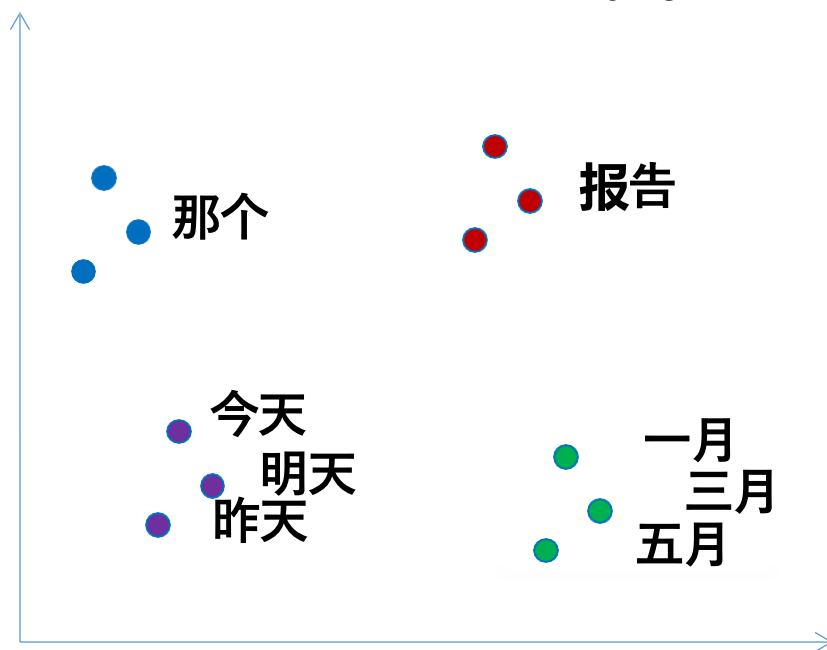
$$\begin{matrix} \text{讲座} \\ \begin{bmatrix} 0.48 \\ 0.46 \\ 0.26 \end{bmatrix} \end{matrix} \otimes \begin{matrix} \text{报告} \\ \begin{bmatrix} 0.42 \\ 0.51 \\ 0.21 \end{bmatrix} \end{matrix} \approx 1$$

统计机器翻译→神经机器翻译

离散符号表示方法 \Rightarrow 连续分布式表示方法

讲座 \otimes 报告 = 0

讲座 \otimes 报告 ≈ 1

$$\begin{bmatrix} 0.48 \\ 0.46 \\ 0.26 \end{bmatrix} \otimes \begin{bmatrix} 0.42 \\ 0.51 \\ 0.21 \end{bmatrix} \approx 1$$


低维、稠密的连续实数空间

统计机器翻译→神经机器翻译

离散符号表示方法 \Rightarrow 连续分布式表示方法

讲座 \otimes 报告 = 0

讲座 \otimes 报告 ≈ 1

$$\begin{bmatrix} 0.48 \\ 0.46 \\ 0.26 \end{bmatrix} \otimes \begin{bmatrix} 0.42 \\ 0.51 \\ 0.21 \end{bmatrix} \approx 1$$

分布式的语义表示是统计机器翻译到神经机器翻译的核心

明天
昨天
三月
五月

低维、稠密的连续实数空间

神经机器翻译

Chinese:

我 在 北京 做了 报告



编码网络



分布式语义表示



解码网络

English:

I gave a talk in Beijing

神经机器翻译

Chinese:

我 在 北京 做了 报告

编码网络

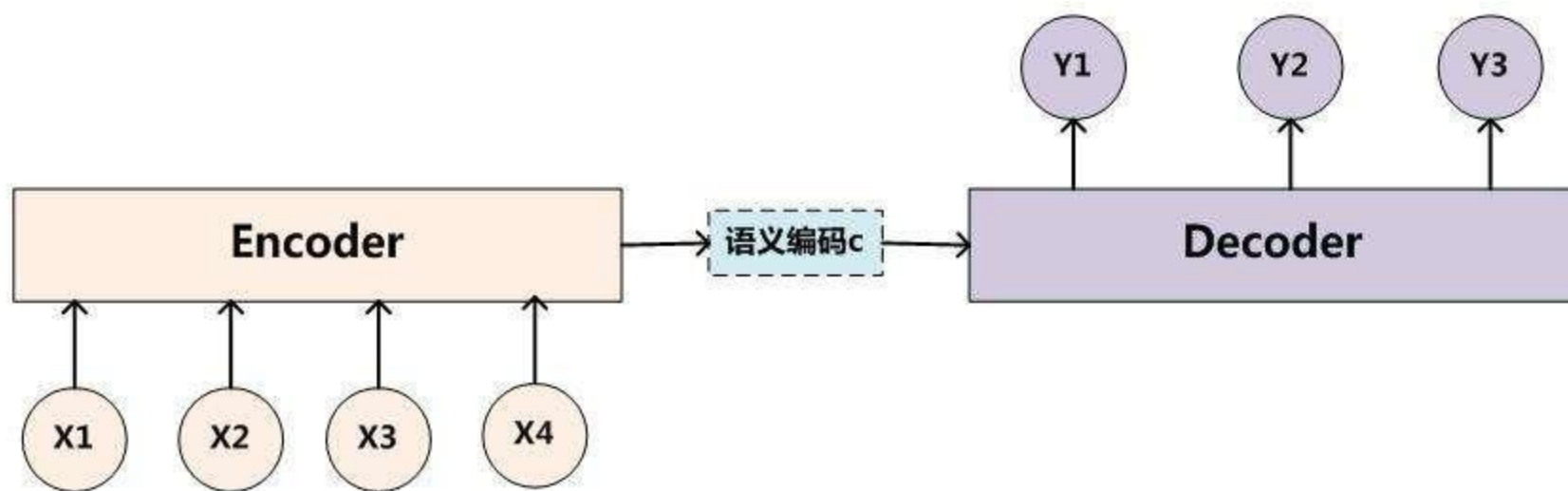
仅需要两个神经网络

解码网络

English:

I gave a talk in Beijing

神经机器翻译



神经机器翻译



Translate

ChineseEnglishSpanishDetect language ▾

↔

美国总统布什昨天在白宫与以色列总理沙龙就中东局势举行了一个小时的会谈。✕

Ä🎤🔊拼 ▾

EnglishChinese (Simplified)Spanish ▾

Translate

US President George W. Bush held an hour-long meeting with Israeli Prime Minister Ariel Sharon on the situation in the Middle East yesterday at the White House.

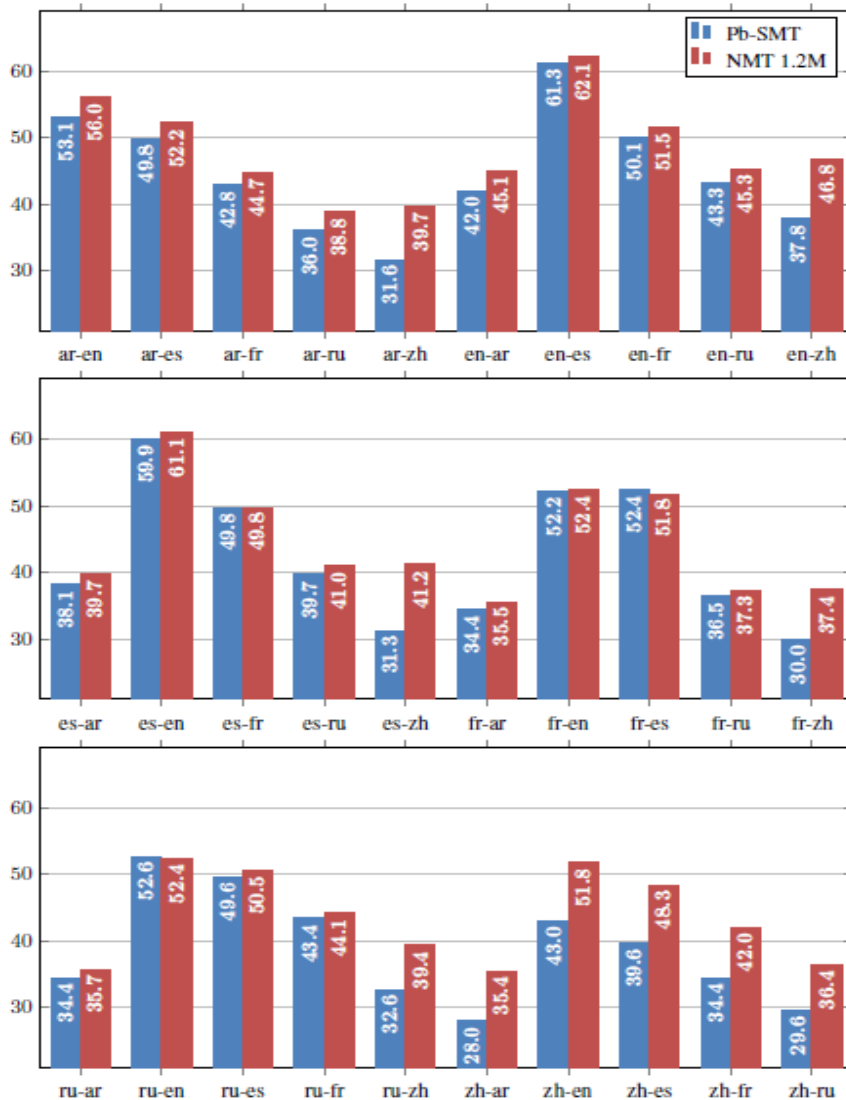
☆📄🔊🔗

✎ Suggest an edit

神经机器翻译

神经机器翻译
大获全胜！

[Junczys-Dowmunt et al, 2016]



统计机器翻译→神经机器翻译

离散符号表示方法 \Rightarrow 连续分布式表示方法

讲座

$$\begin{bmatrix} 0.48 \\ 0.46 \\ 0.26 \end{bmatrix}$$

\otimes

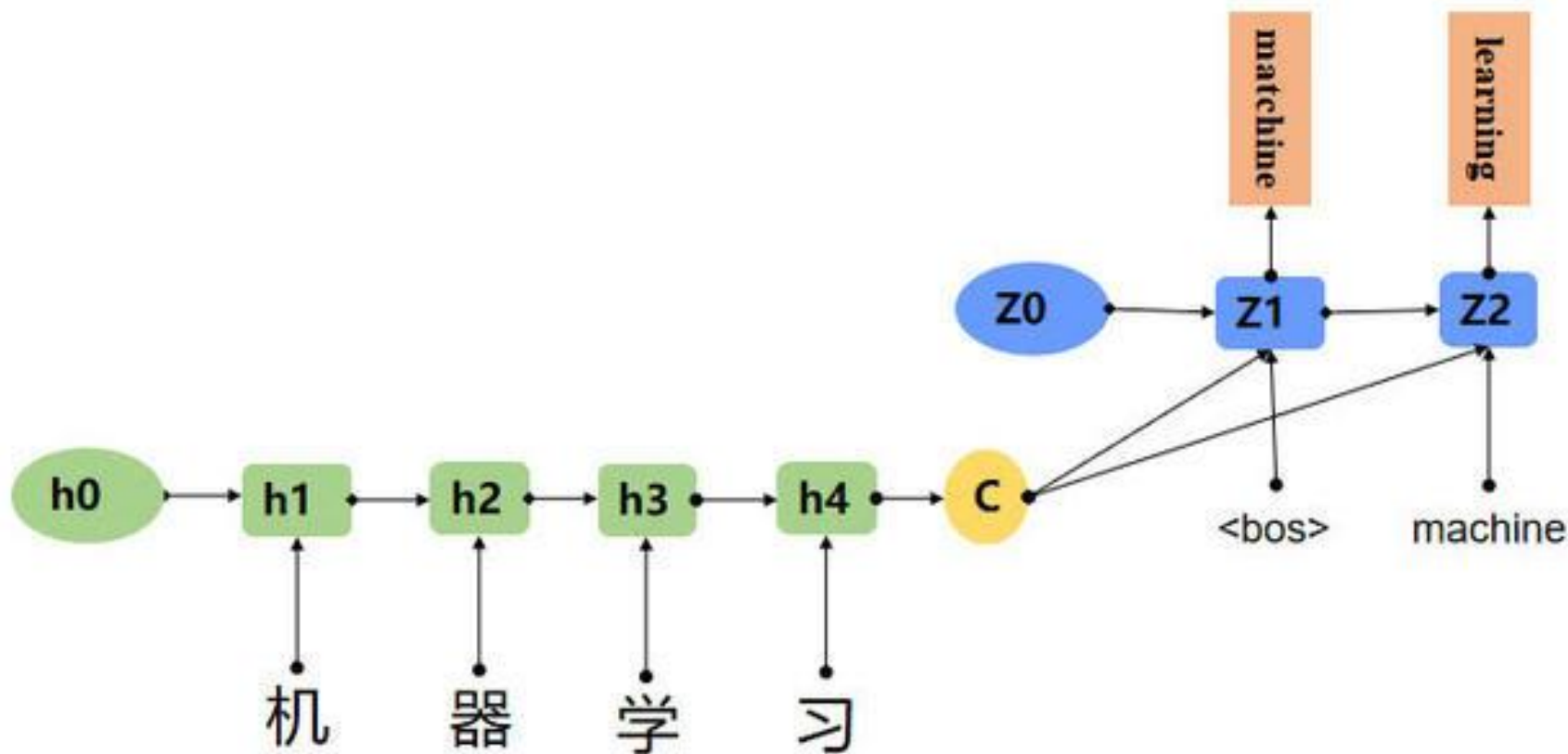
报告

$$\begin{bmatrix} 0.42 \\ 0.51 \\ 0.21 \end{bmatrix}$$

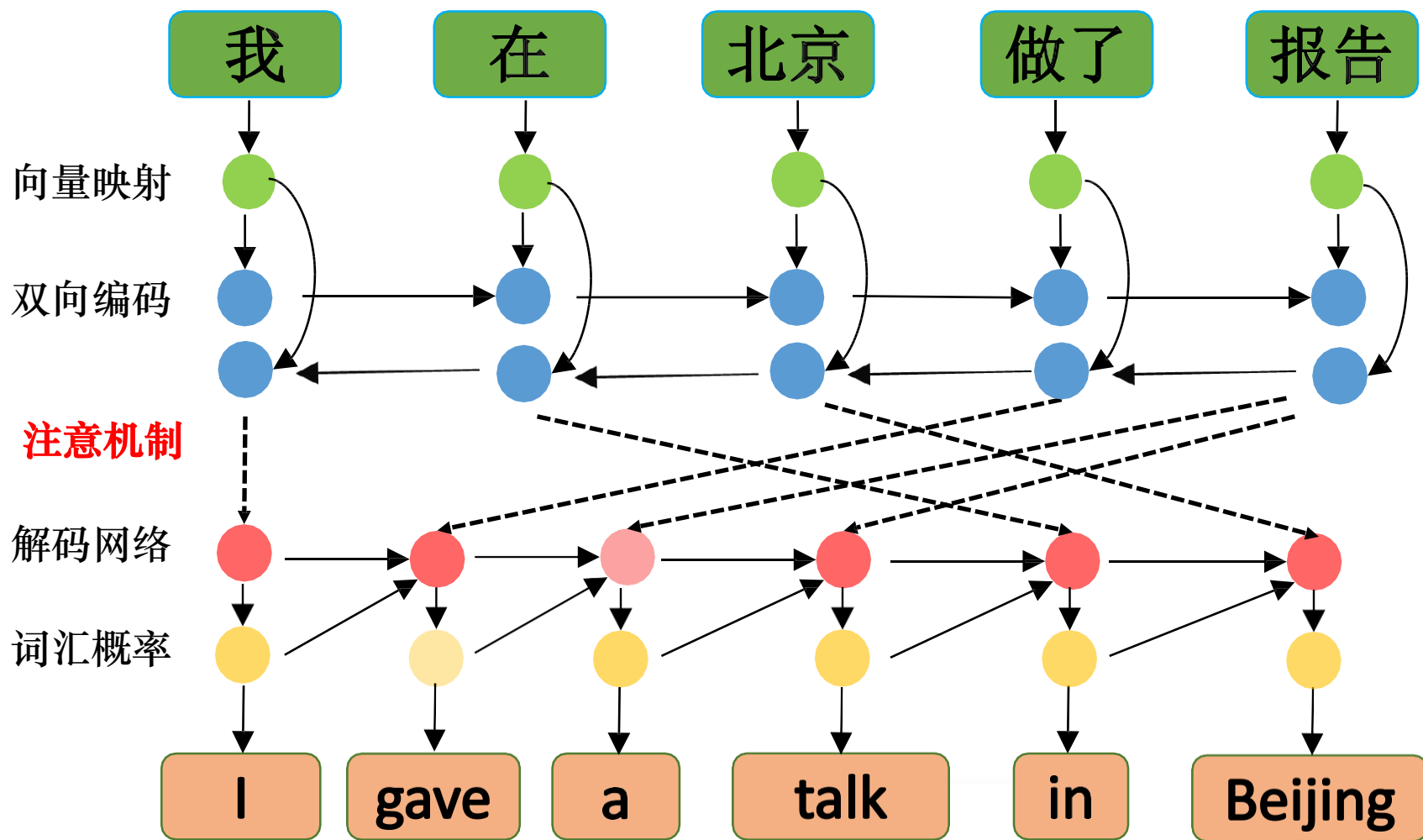
≈ 1

表示是核心
运算是关键

统计机器翻译→神经机器翻译

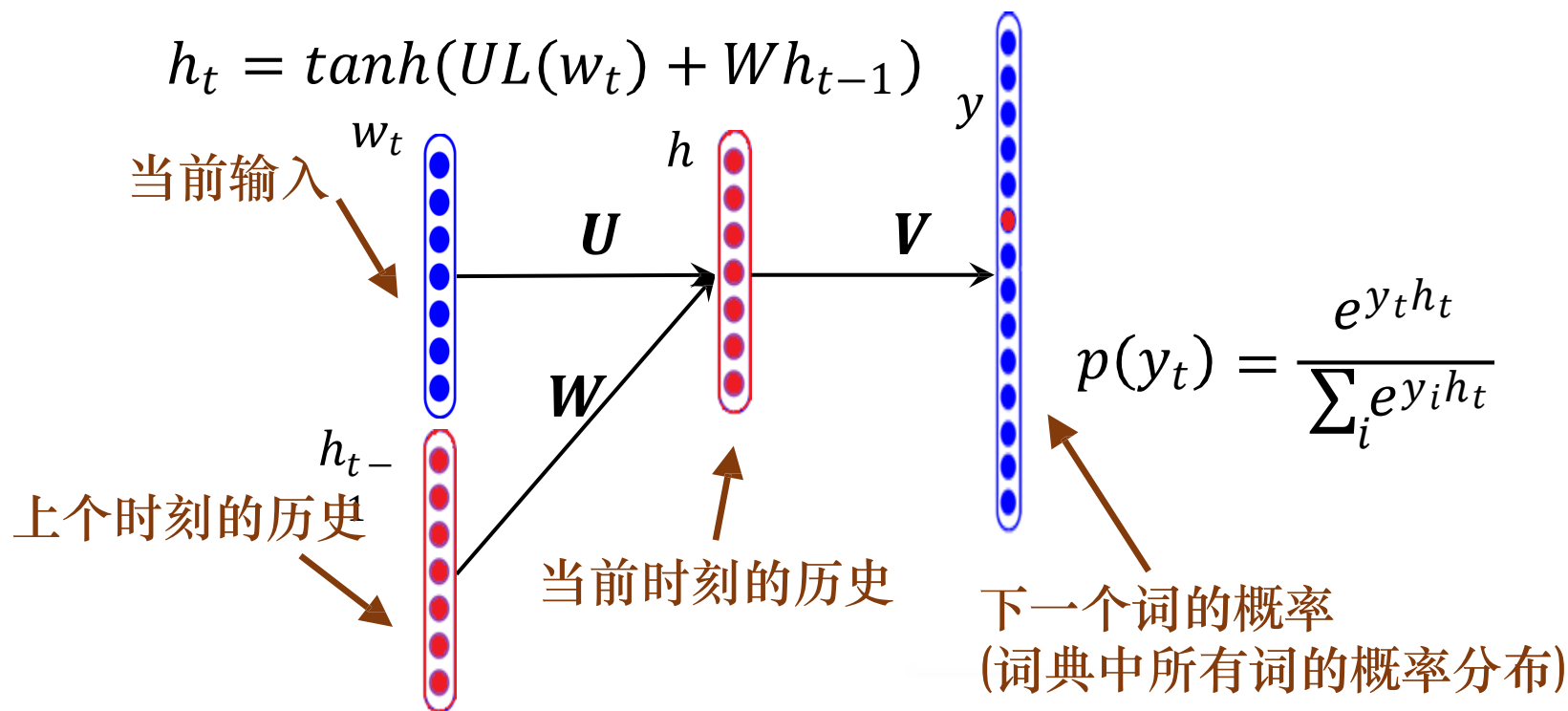


神经机器翻译

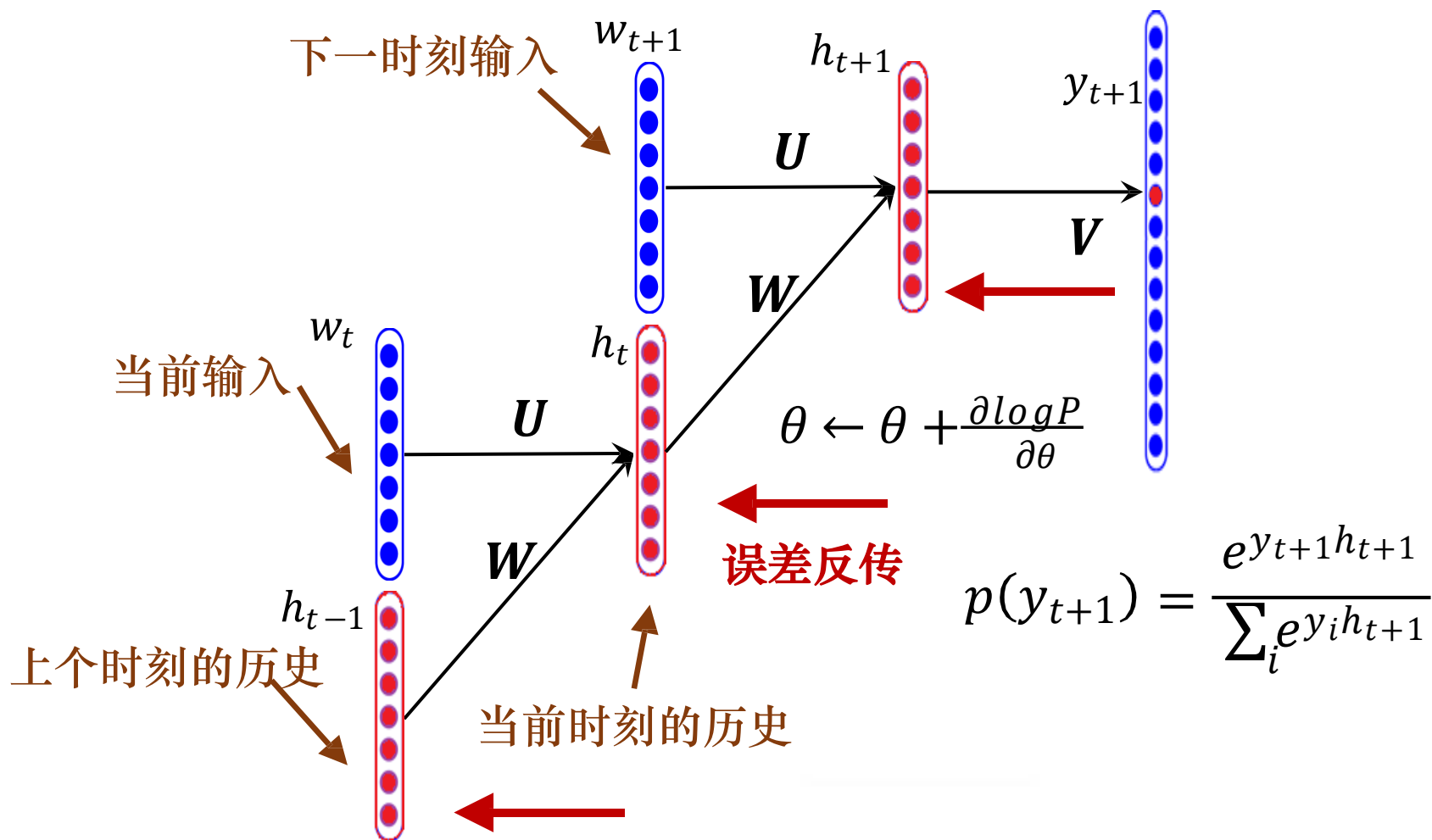


循环神经网络

- 输入: $t - 1$ 时刻历史 h_{t-1} 与 t 时刻输入 w_t
- 输出: t 时刻历史 h_t 与 下个时刻 $t + 1$ 输入 y_t 的概率



循环神经网络



神经机器翻译

$$h_s = \tanh(UL(w_s) + Wh_{s-1})$$

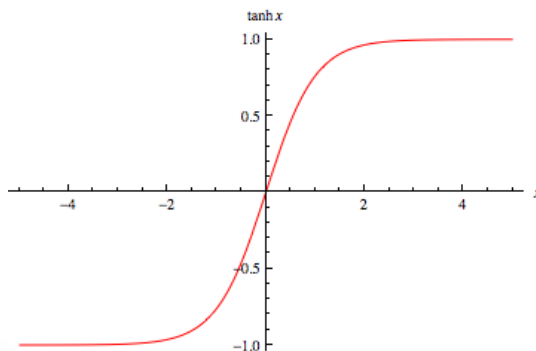
$L(w_s):$ $w_s \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in R^3$ 我 $\longrightarrow \begin{bmatrix} 0.1 \\ 0.9 \\ 0.6 \end{bmatrix}$ 随机初始化

$h_{s-1}: \quad$ 上一时刻的历史信息 $h_0 = \begin{bmatrix} 0.0 \\ 0.0 \\ 0.0 \end{bmatrix}$

$$U = \begin{bmatrix} 0.1 & 0.2 & 0.1 \\ 0.3 & 0.2 & 0.0 \\ 0.4 & 0.0 & 0.2 \end{bmatrix} \in R^{3 \times 3} \quad W = \begin{bmatrix} 0.0 & 0.3 & 0.2 \\ 0.1 & 0.1 & 0.3 \\ 0.0 & 0.4 & 0.1 \end{bmatrix} \in R^{3 \times 3}$$

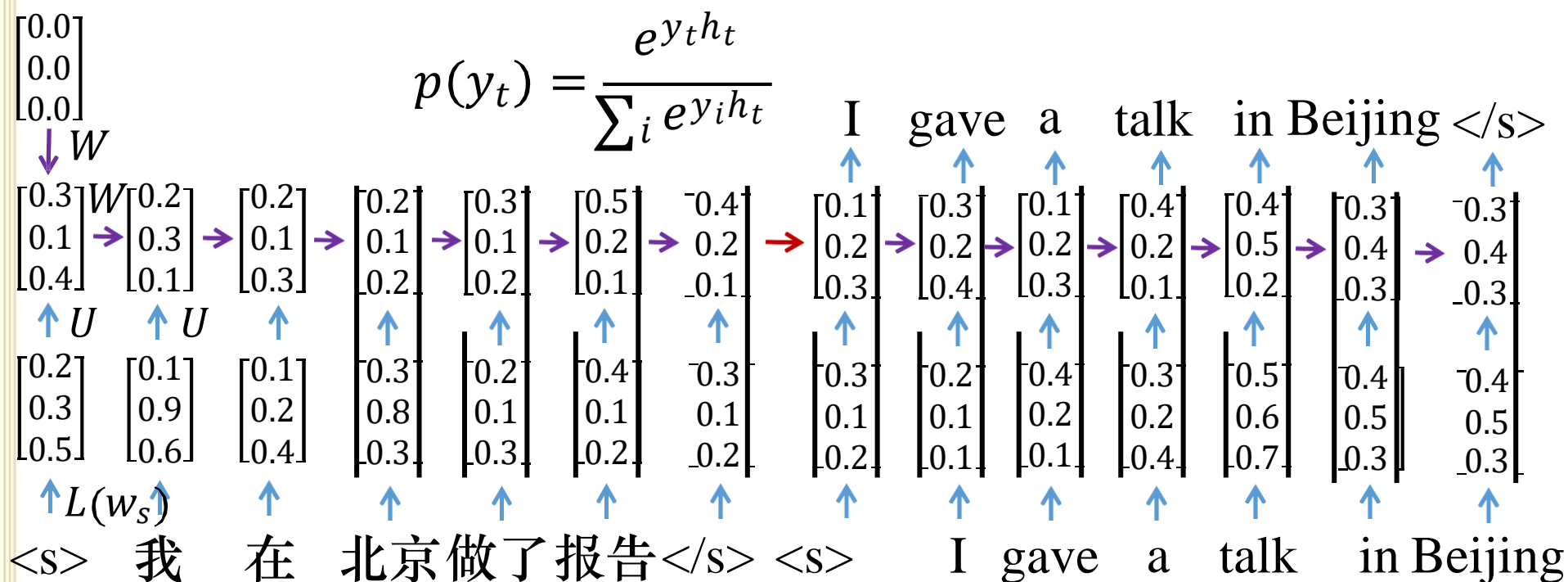
$$z = UL(w_s) + Wh_{s-1} \in R^3$$

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \longrightarrow$$



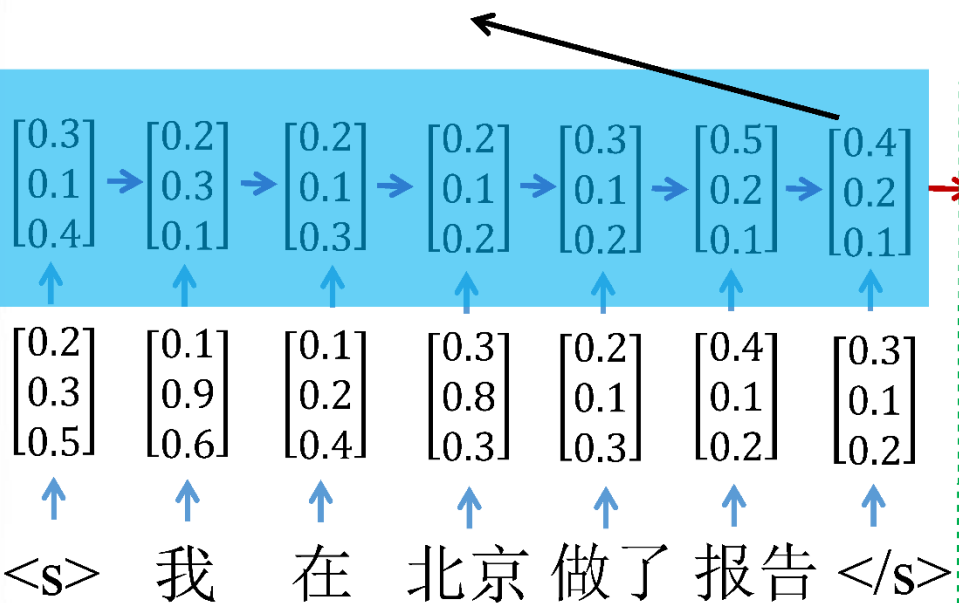
神经机器翻译

$$h_s = \tanh(UL(w_s) + Wh_{s-1}) \quad h_t = \tanh(UL(w_t) + Wh_{t-1})$$



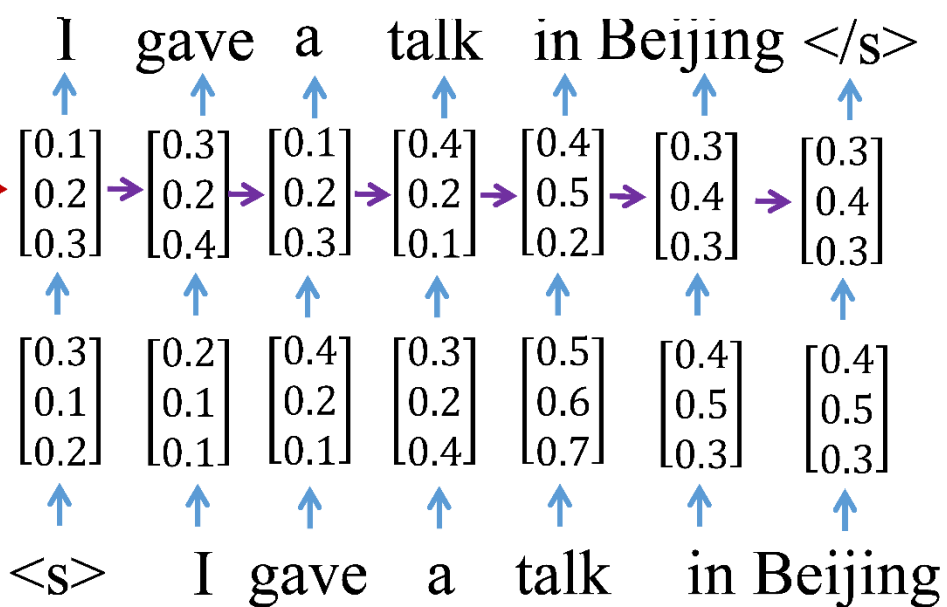
神经机器翻译

将源语言句子编码成一个
实数向量语义表示



编码器

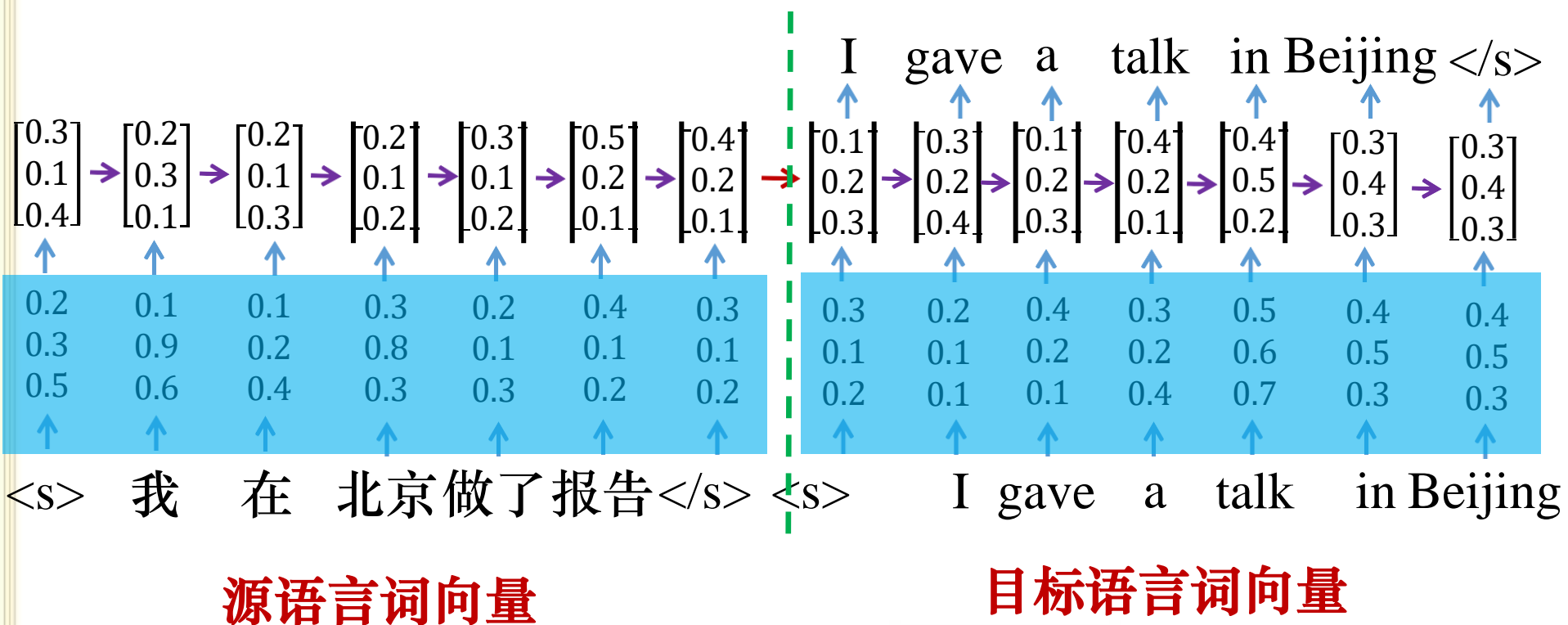
将源语言句子的语义表示
解码生成目标语言句子



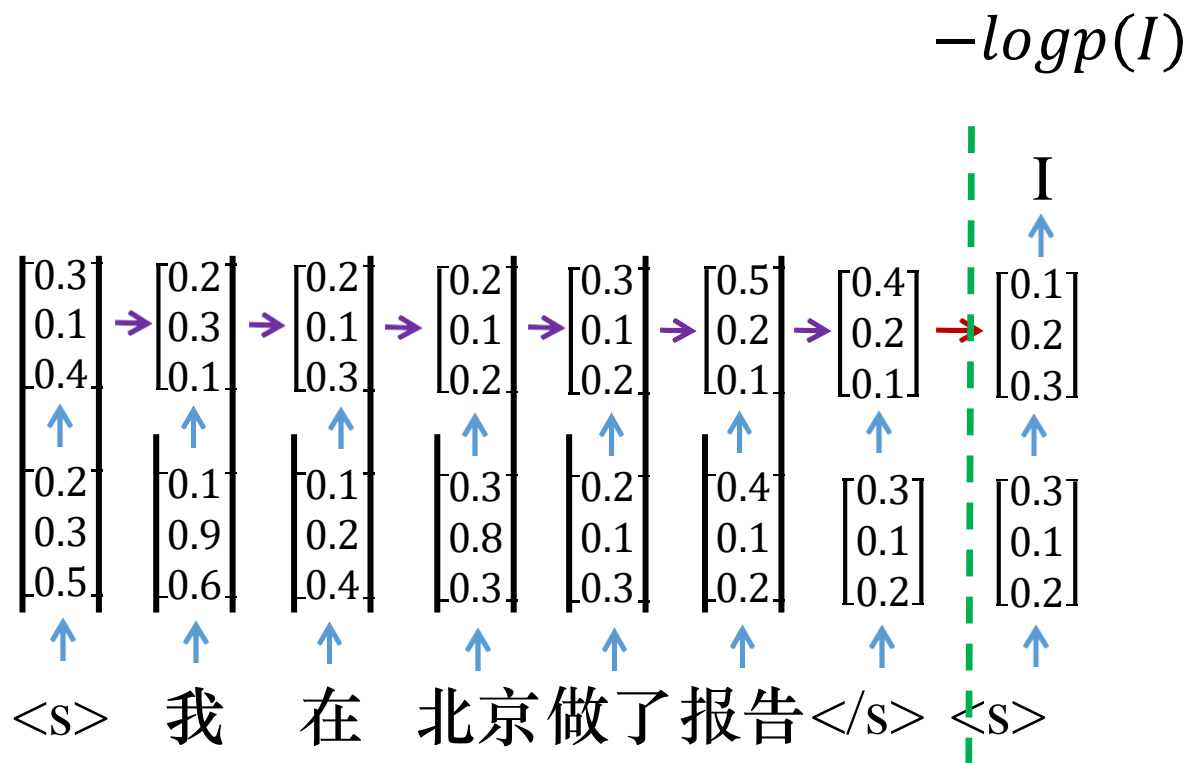
解码器

神经机器翻译

词向量可以随机初始化，在训练过程中进行优化！

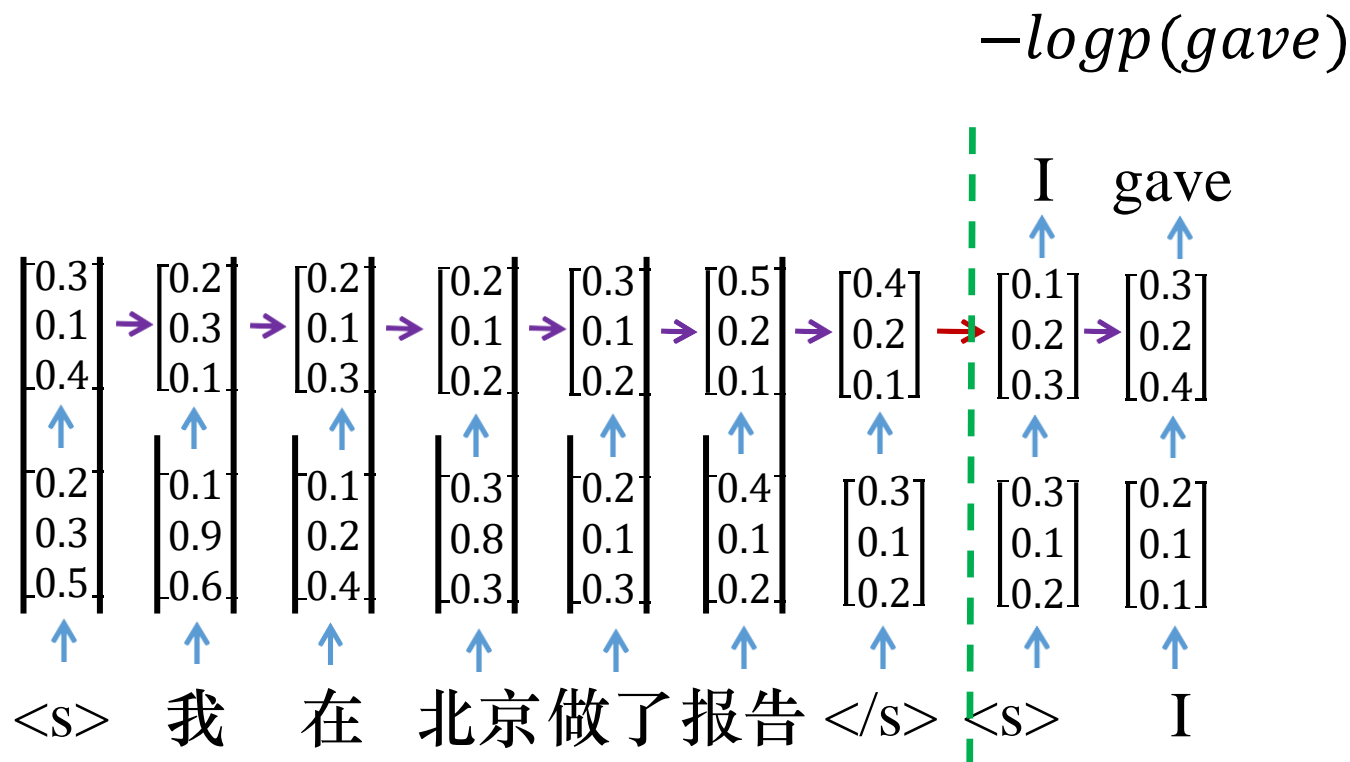


神经机器翻译



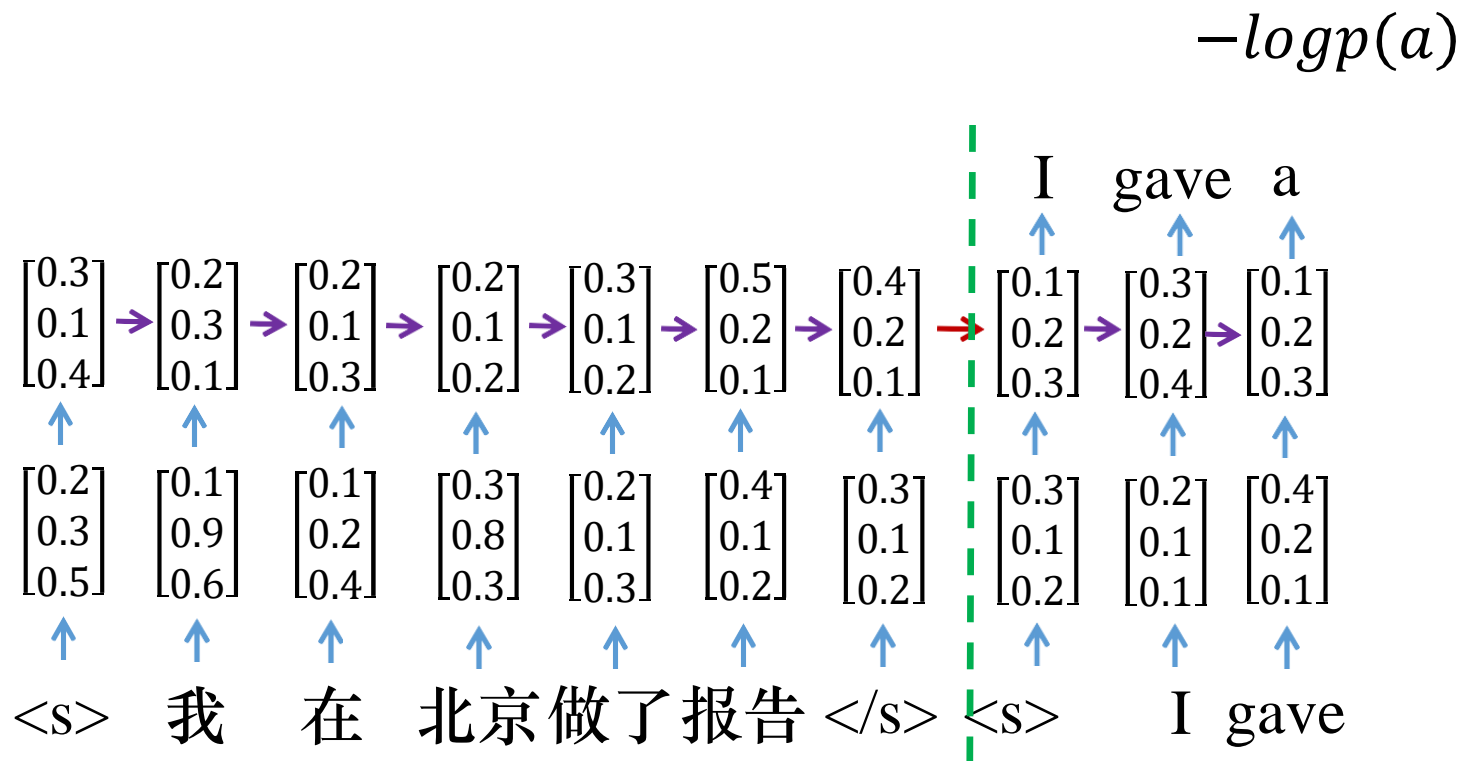
最大化 $P(\textit{target}|\textit{source})$

神经机器翻译



最大化 $P(\text{target}|\text{source})$

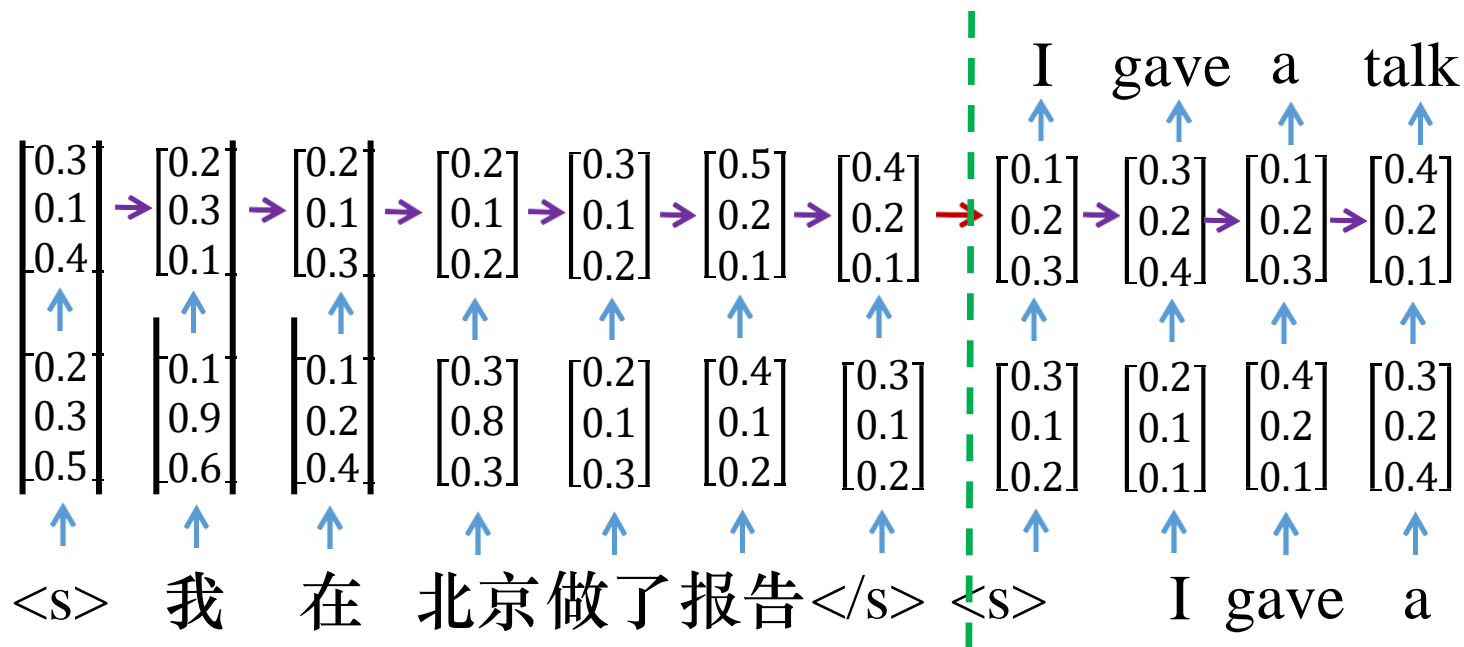
神经机器翻译



最大化 $P(\textit{target}|\textit{source})$

神经机器翻译

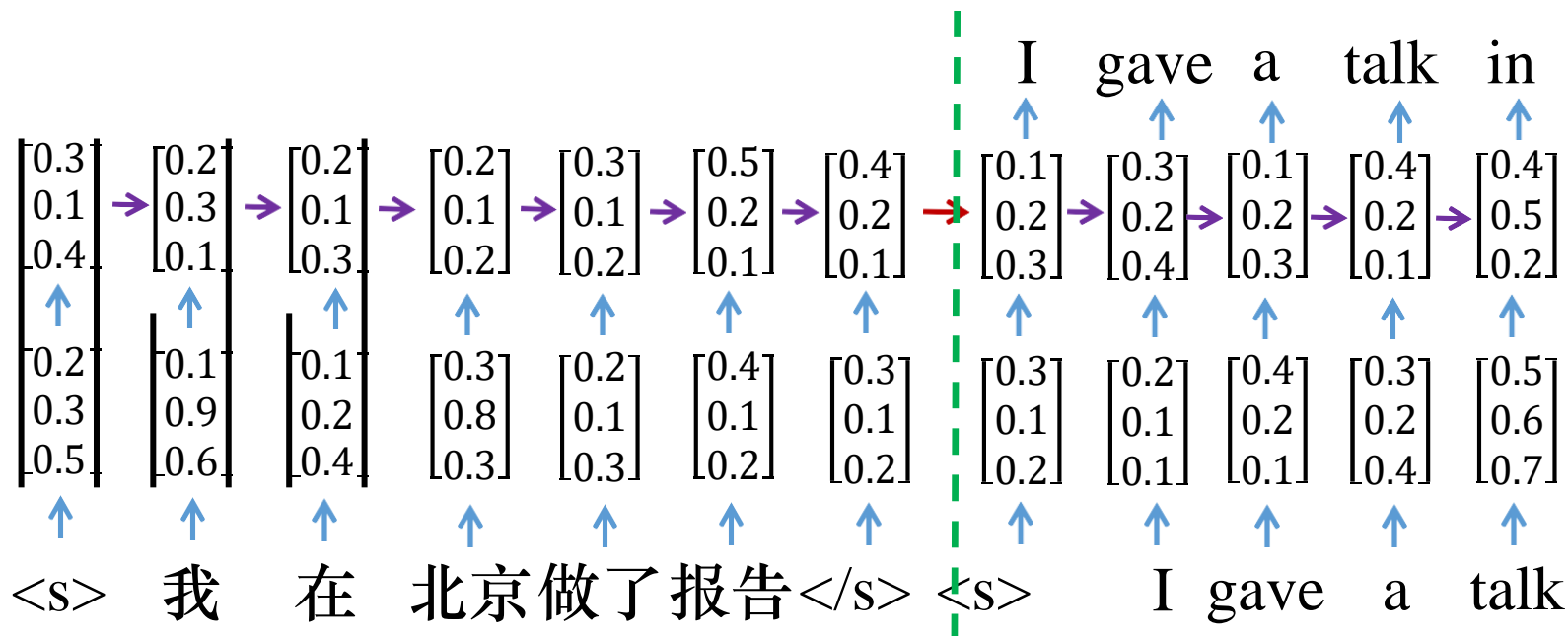
$$-\log p(\text{talk})$$



最大化 $P(\text{target}|\text{source})$

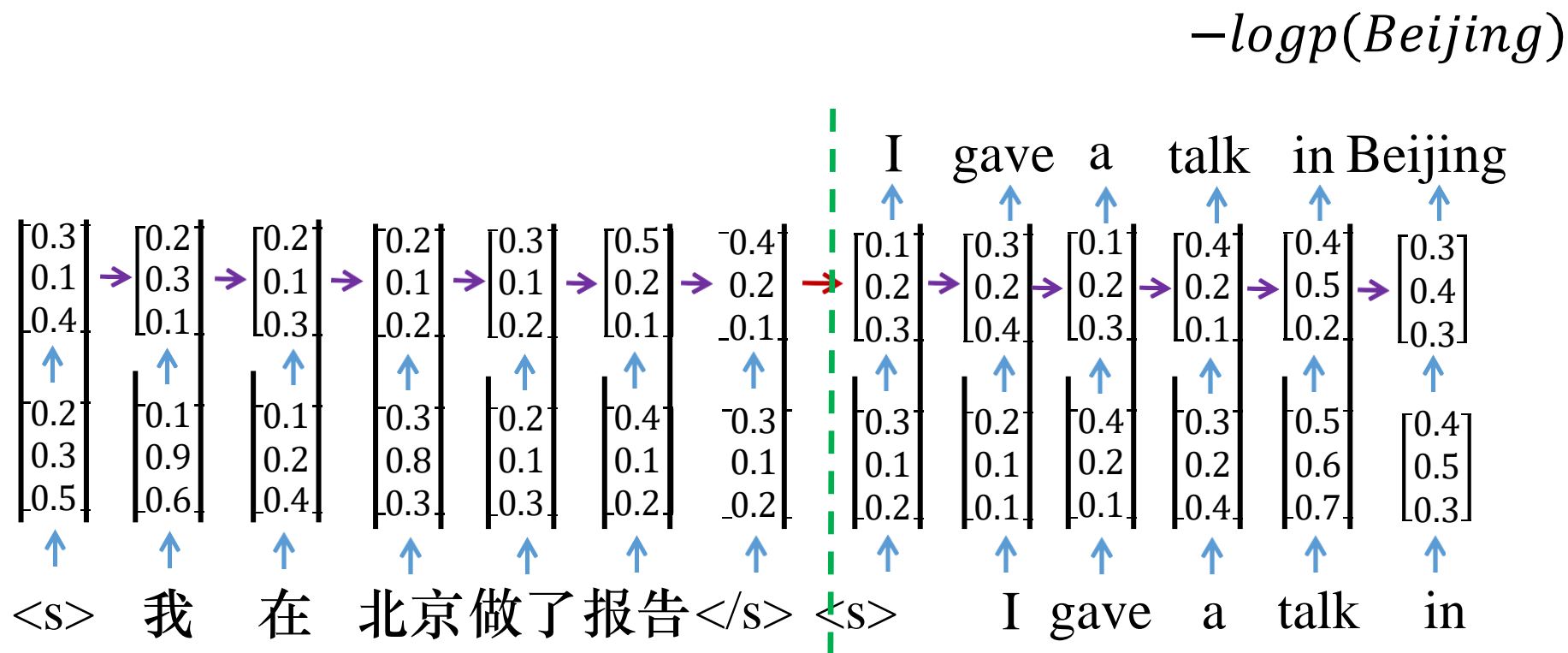
神经机器翻译

$-\log p(in)$



最大化 $P(\text{target}|\text{source})$

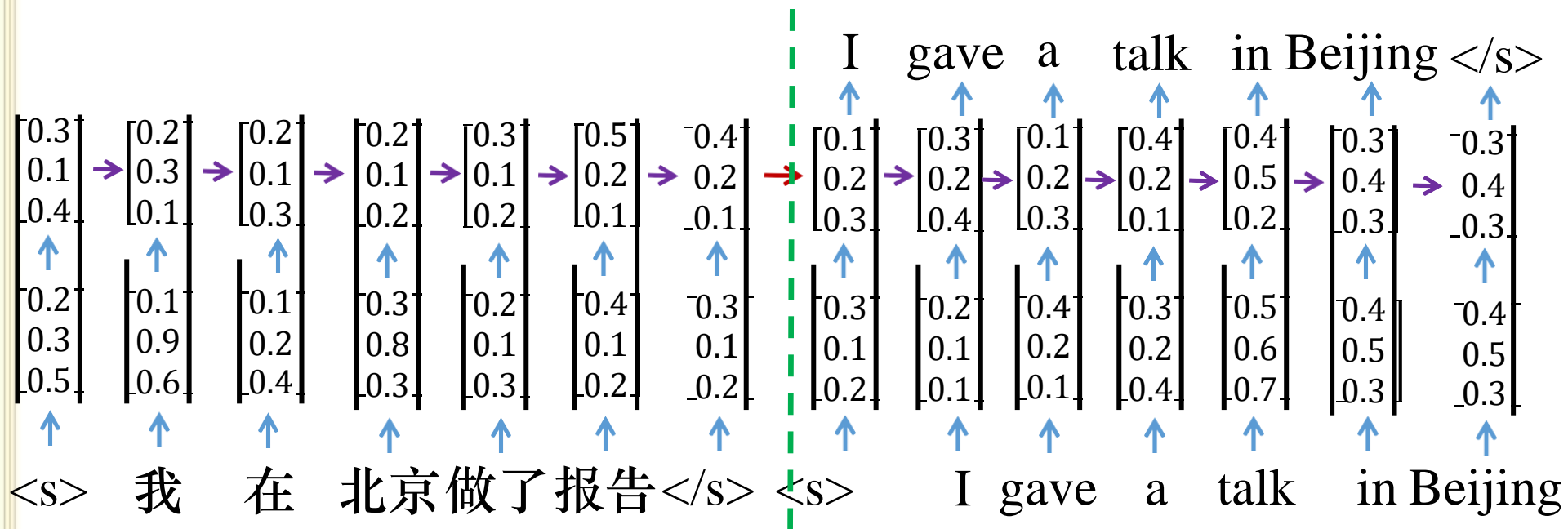
神经机器翻译



最大化 $P(\text{target}|\text{source})$

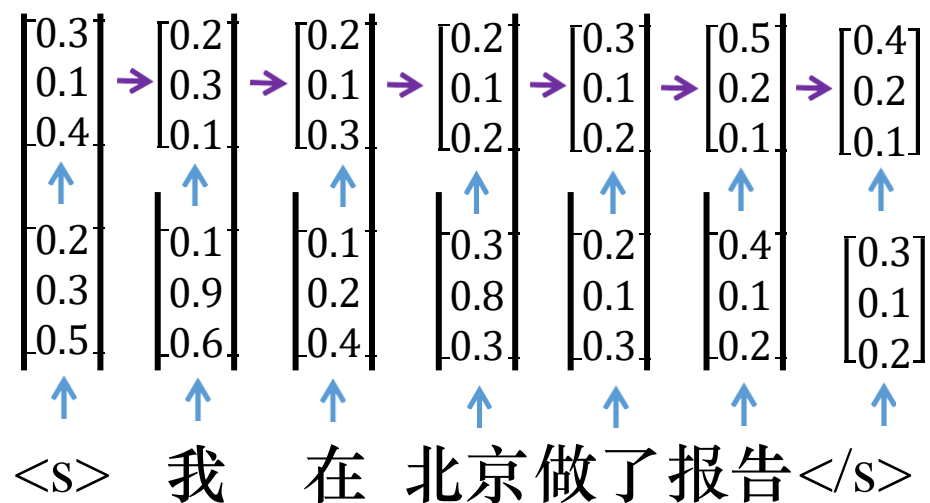
神经机器翻译

$$-\log p(\langle /s \rangle)$$

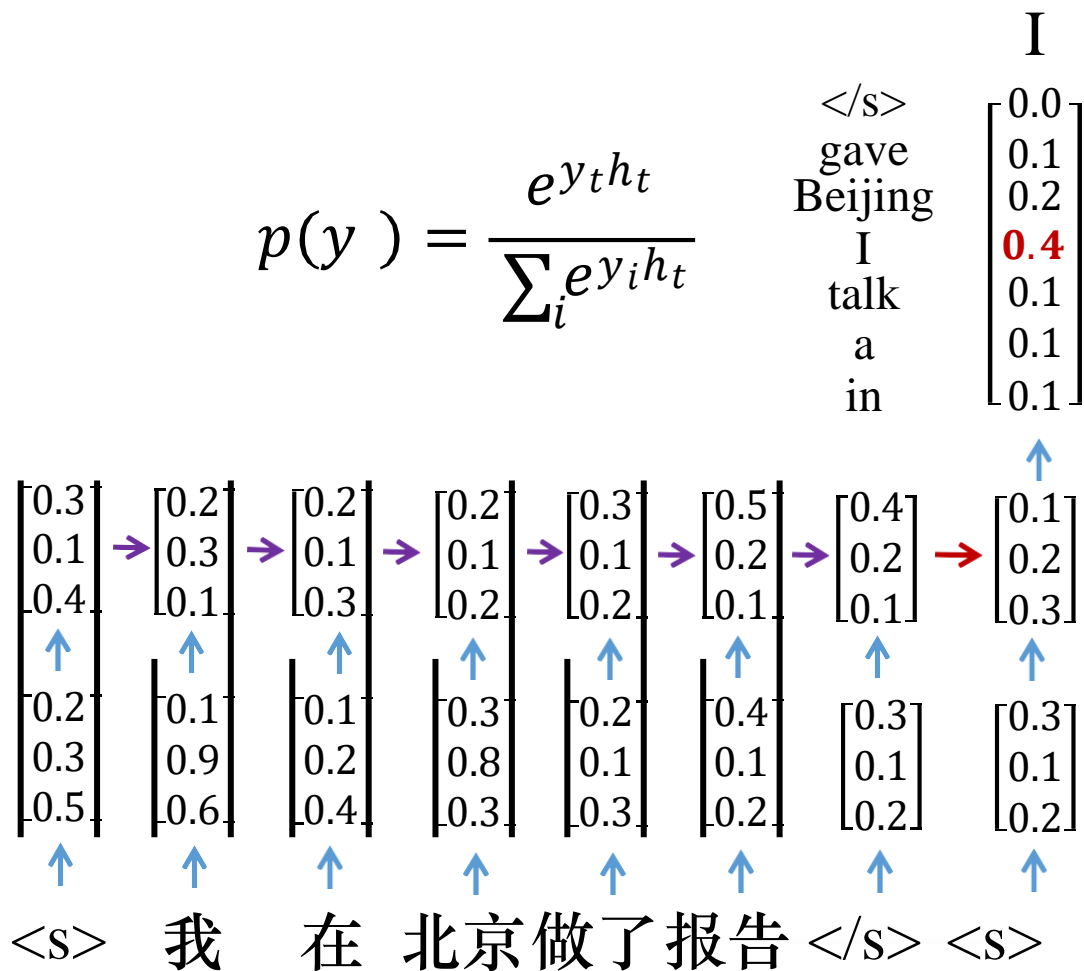


最大化 $P(\textit{target}|\textit{source})$

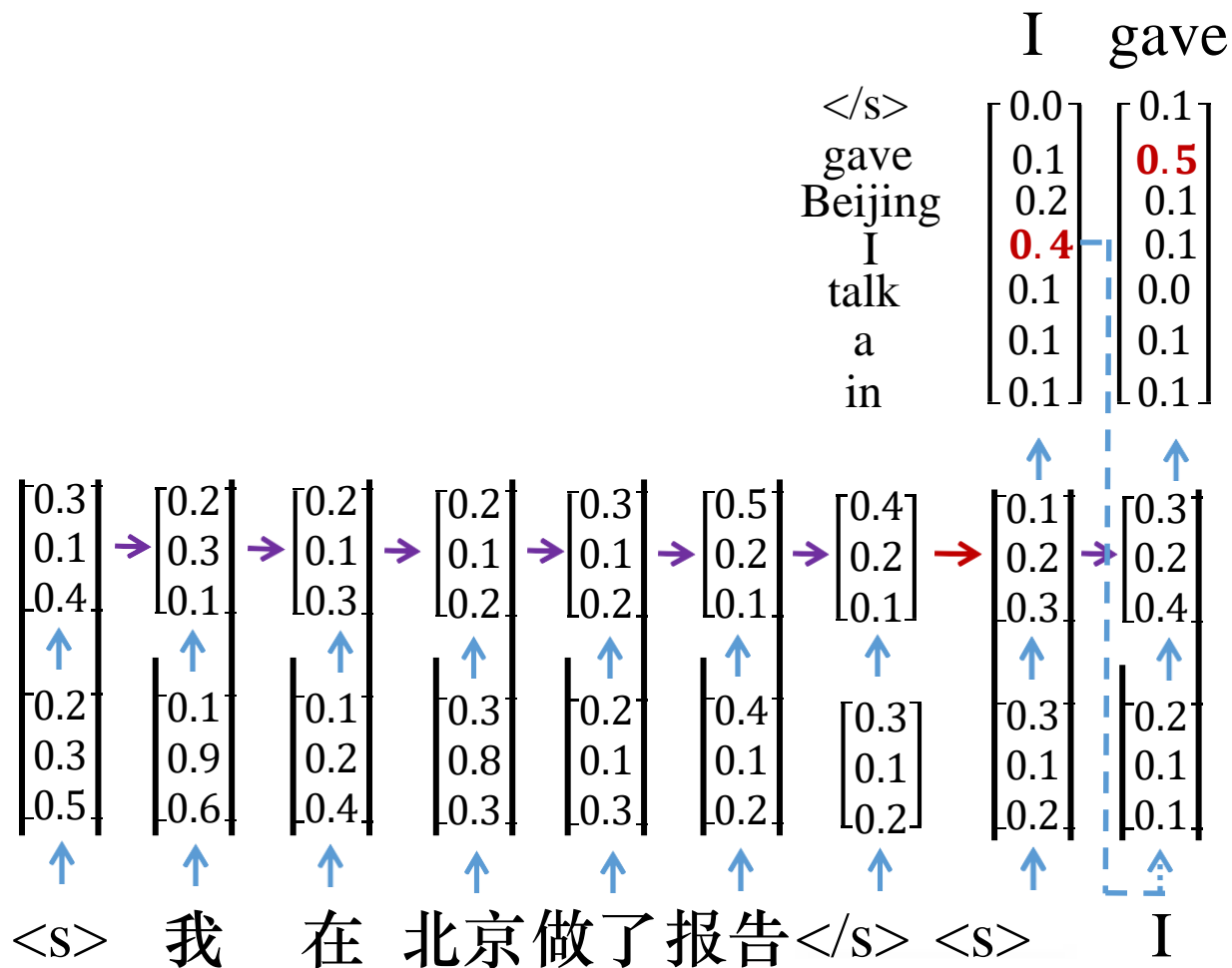
神经机器翻译-测试



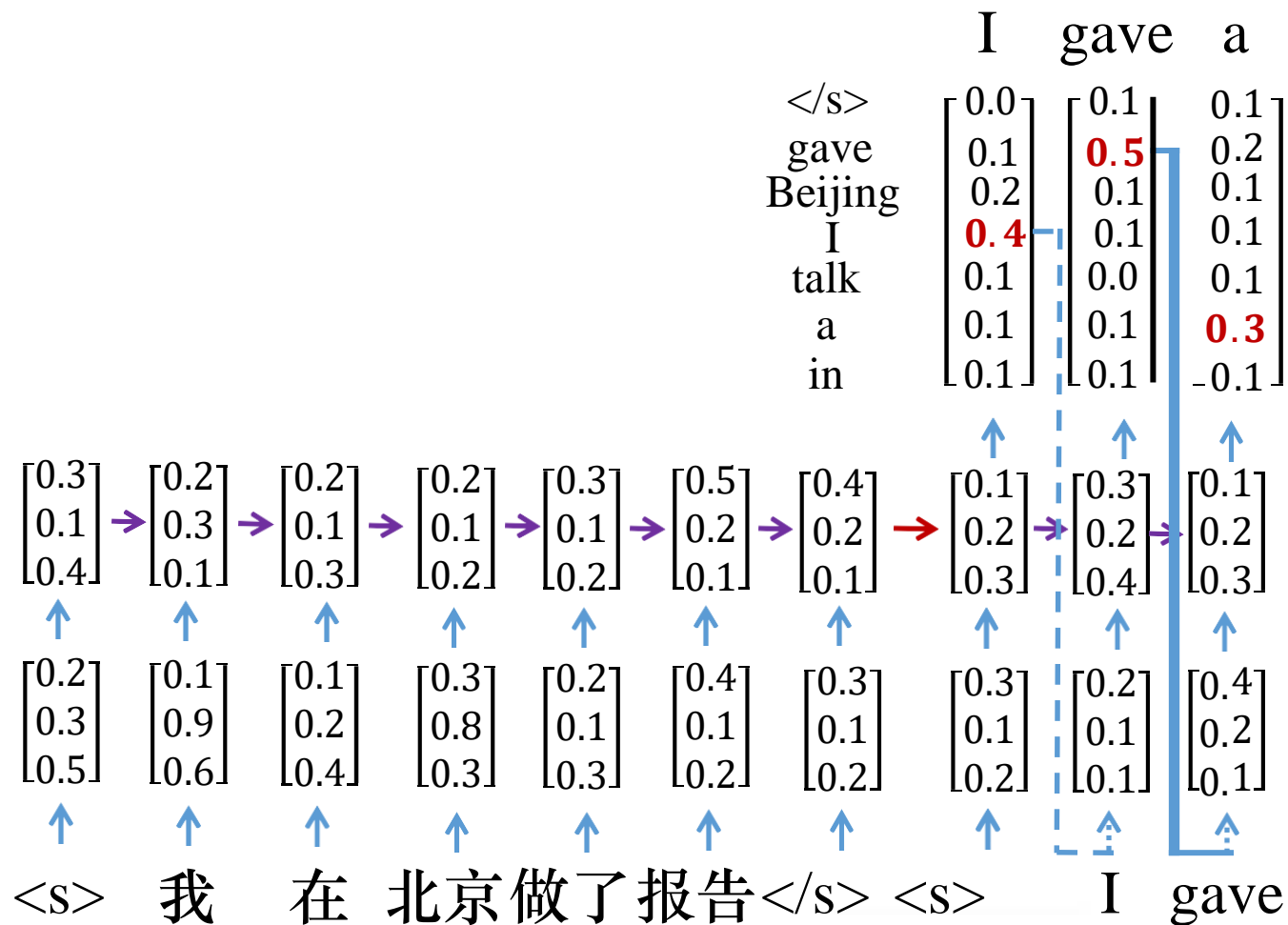
神经机器翻译-测试



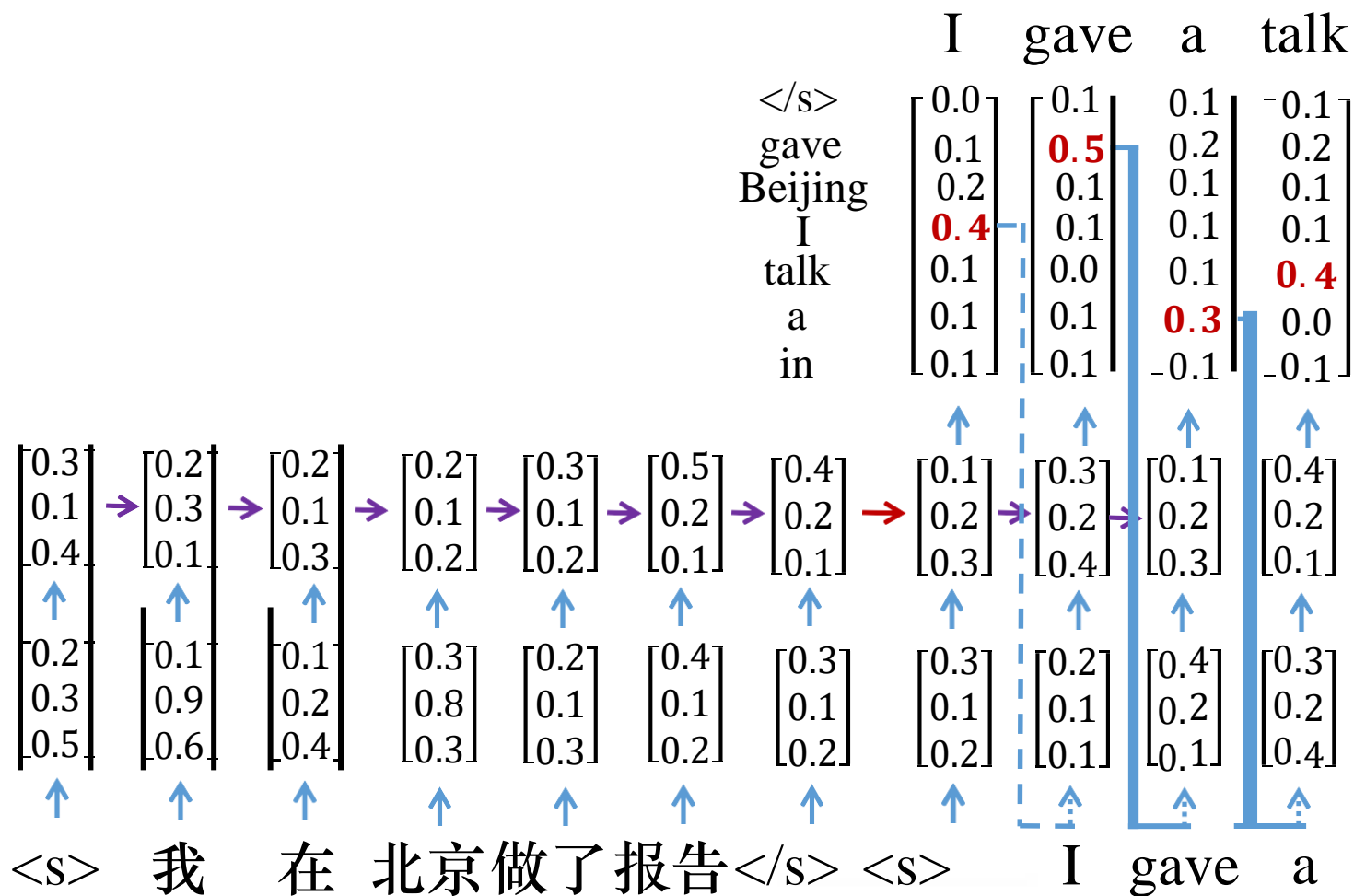
神经机器翻译-测试



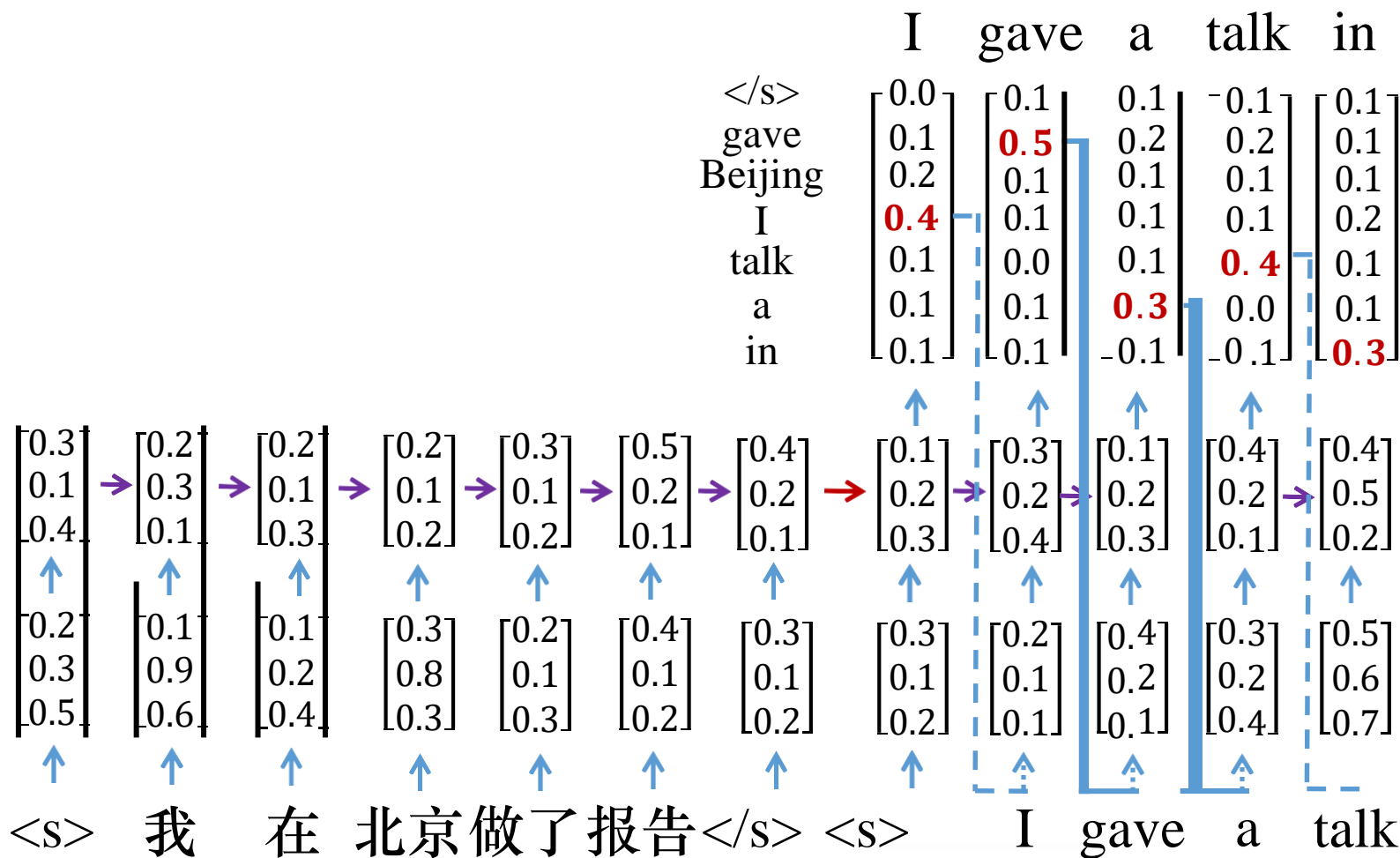
神经机器翻译-测试



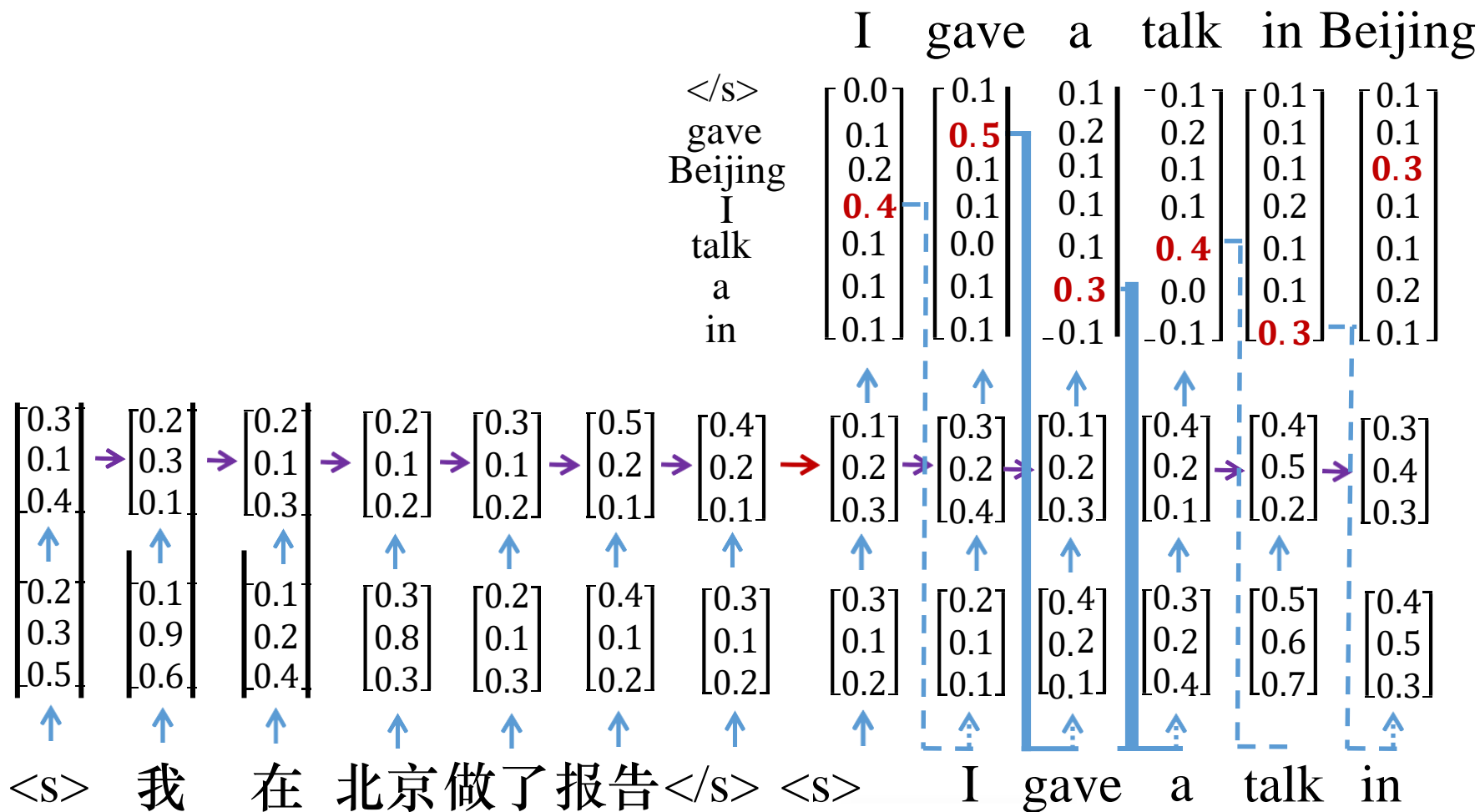
神经机器翻译-测试



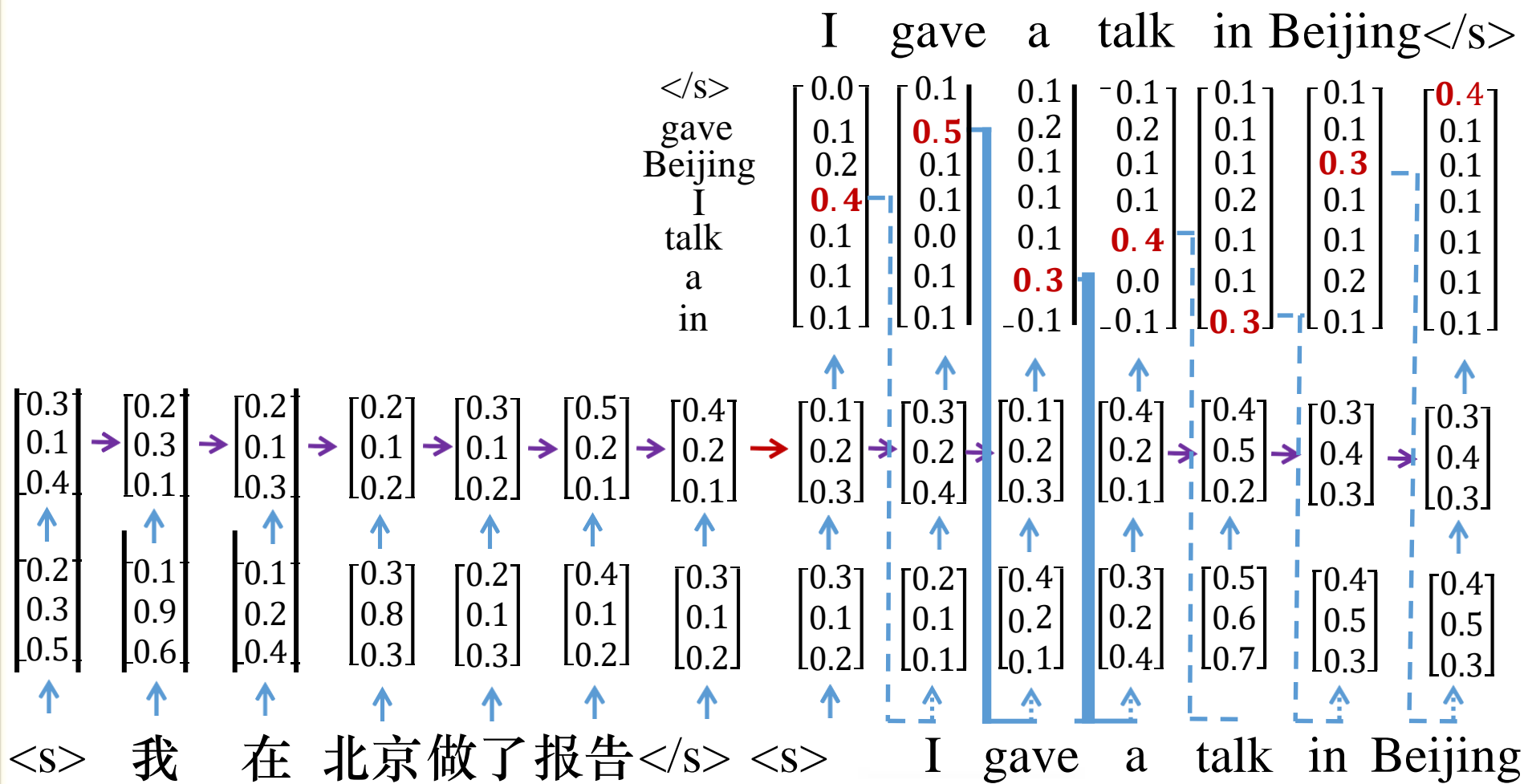
神经机器翻译-测试



神经机器翻译-测试

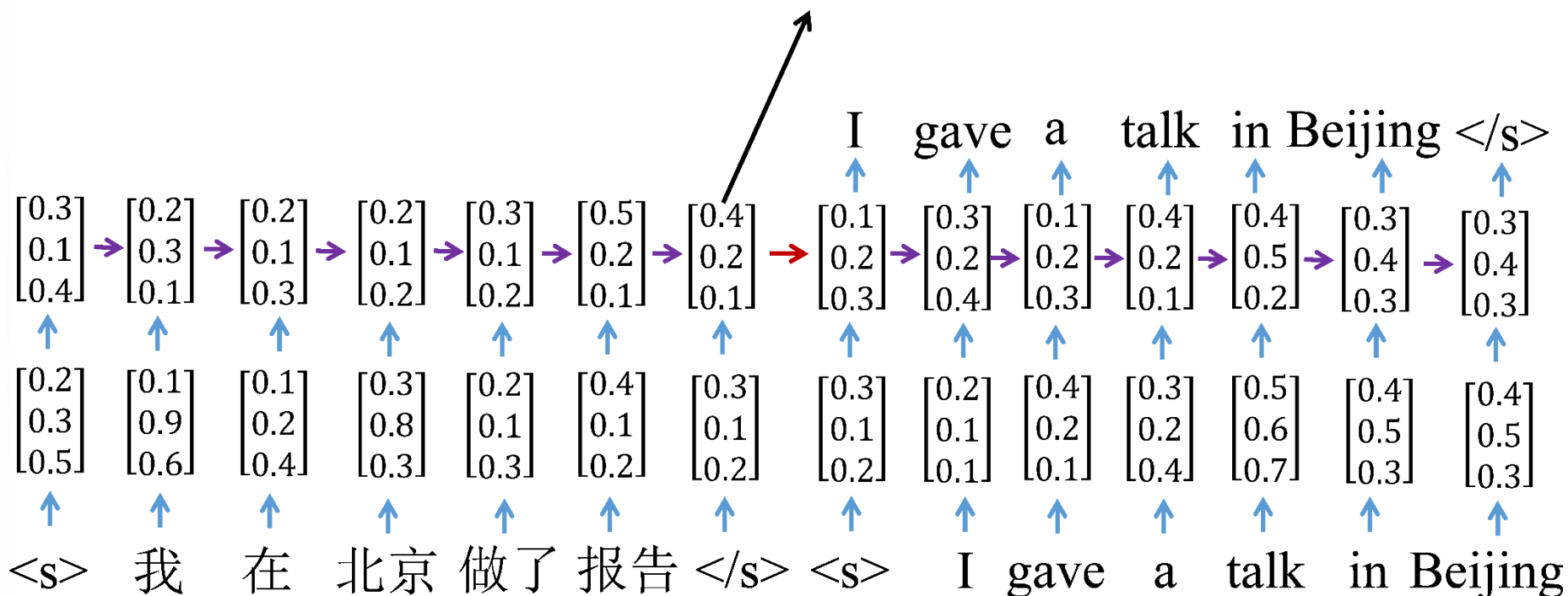


神经机器翻译-测试



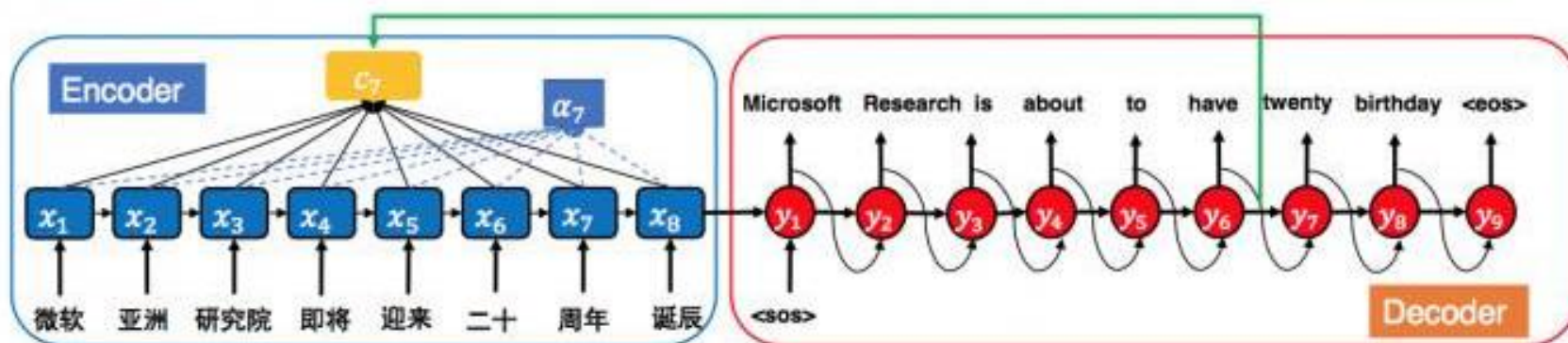
神经机器翻译-计算单元

一个实数向量无法表示
源语言句子的完整语义



统计机器翻译→神经机器翻译

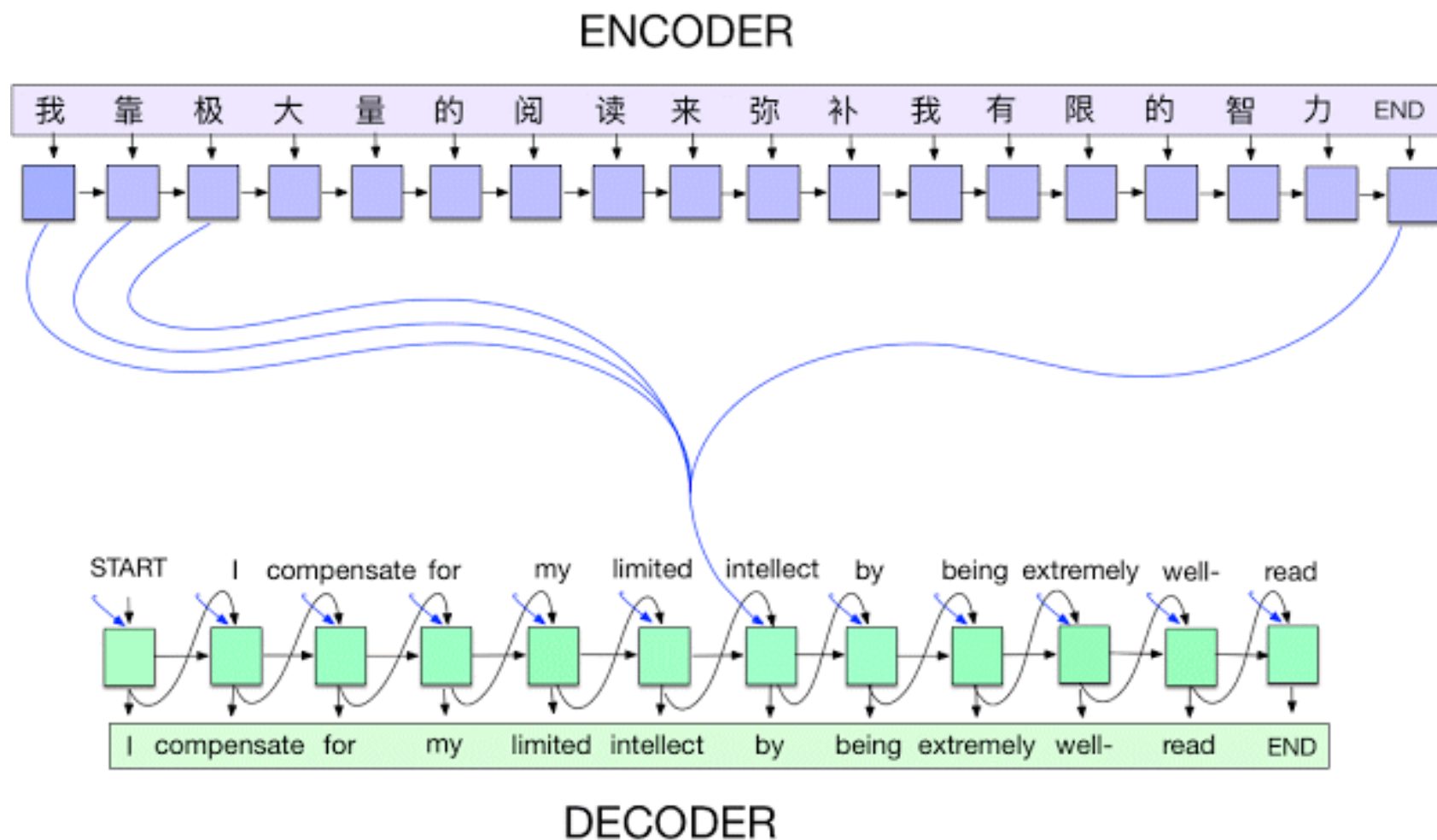
Attention 注意力机制



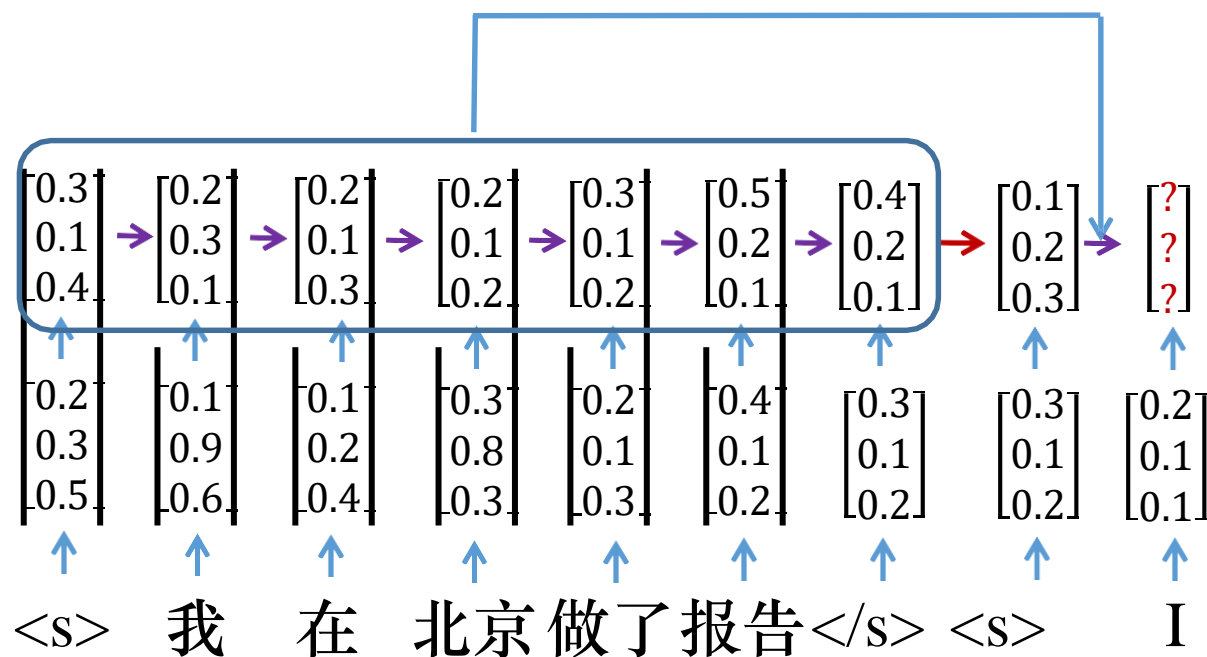
- $c_j = \sum_{i=1}^{T_x} \alpha_{ji} h_i$
- $\alpha_{ji} = \frac{\exp(e_{ji})}{\sum_{k=1}^{T_x} \exp(e_{jk})}$
- $e_{ji} = A(s_{j-1}, h_i)$

$$s_j = f(y_{j-1}, s_{j-1}; c_j): \text{LSTM/GRU}$$

神经机器翻译-注意机制

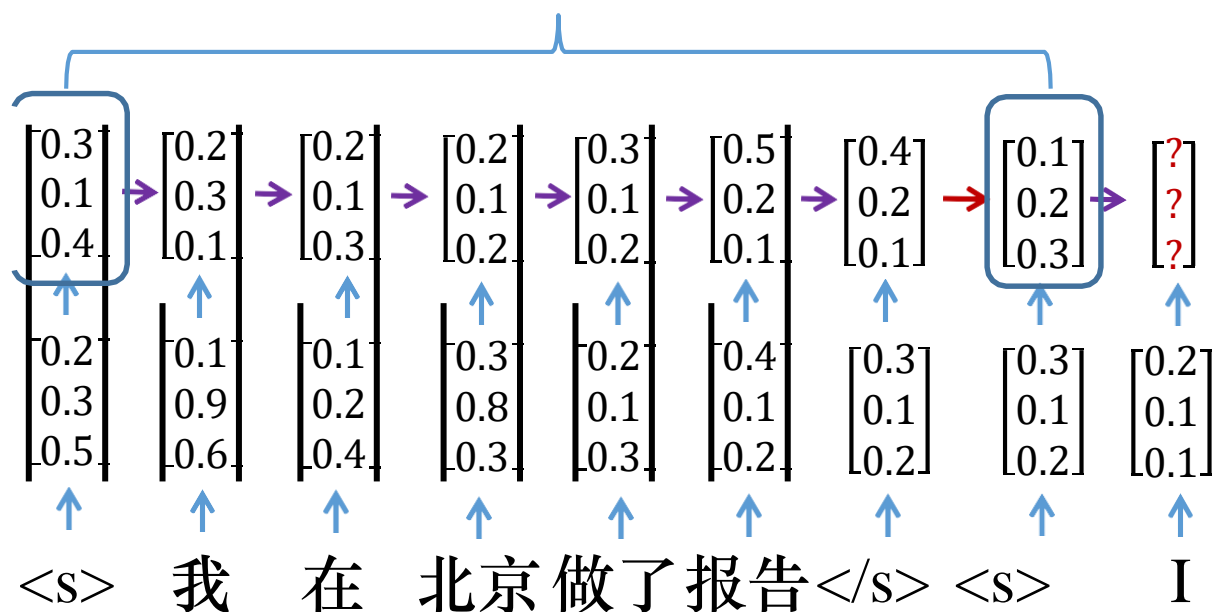


神经机器翻译-注意机制



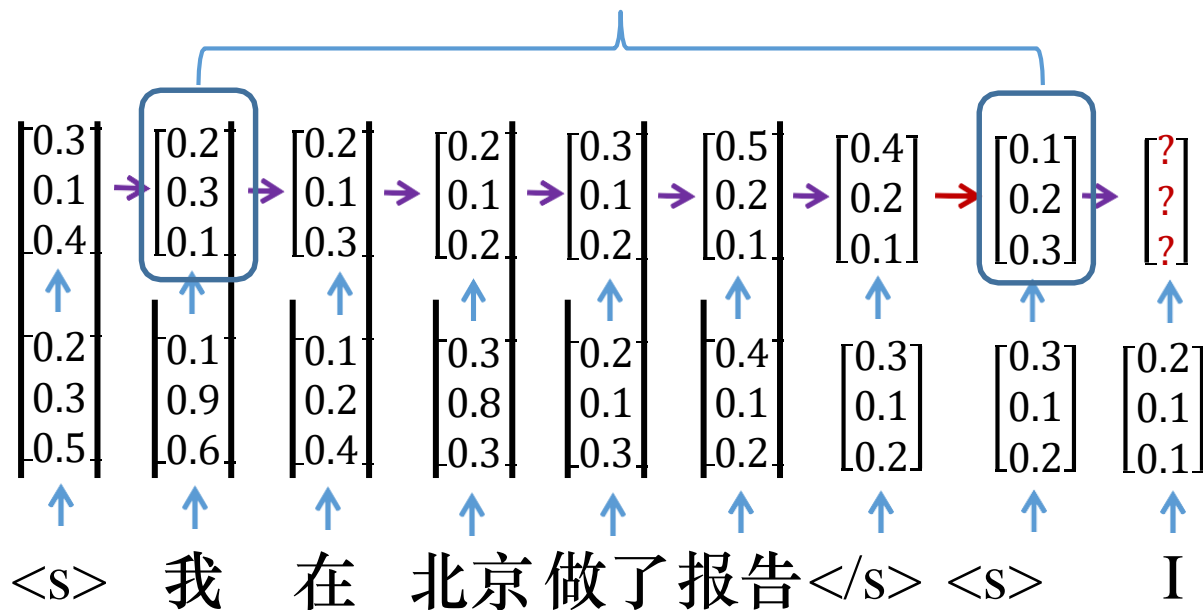
神经机器翻译-注意机制

$$\text{score}(h_s, h_t) = 1$$

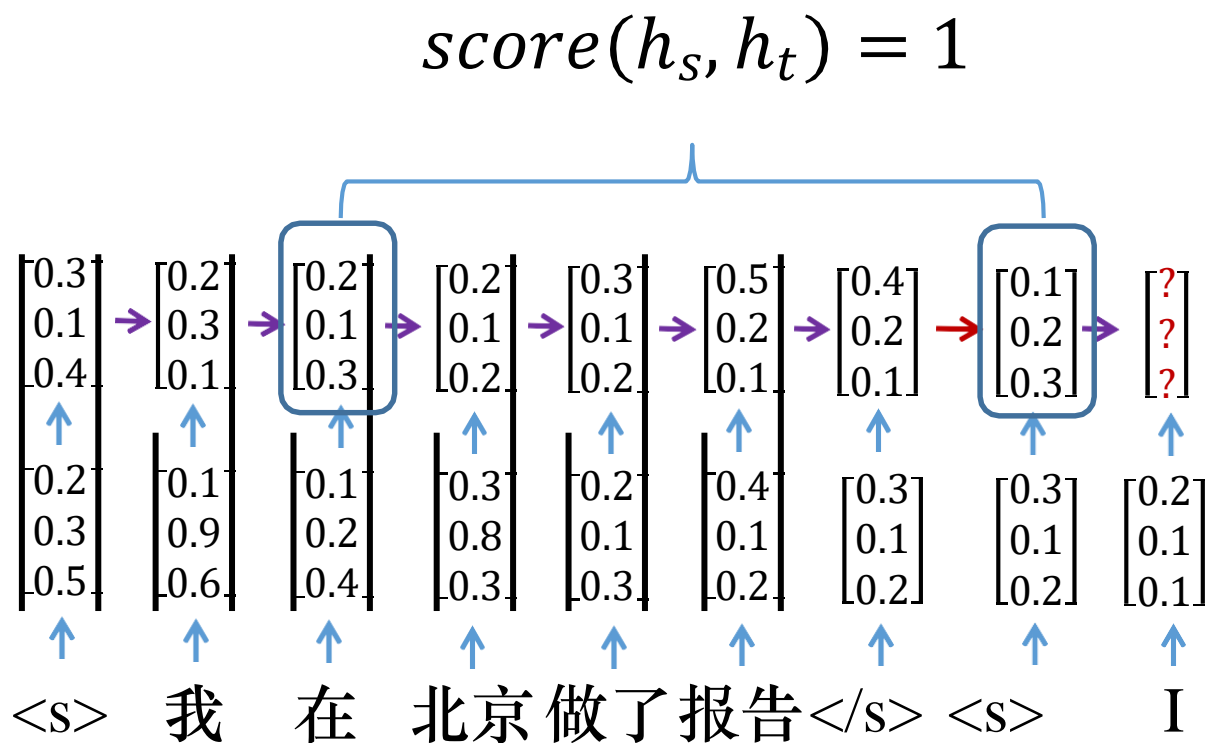


神经机器翻译-注意机制

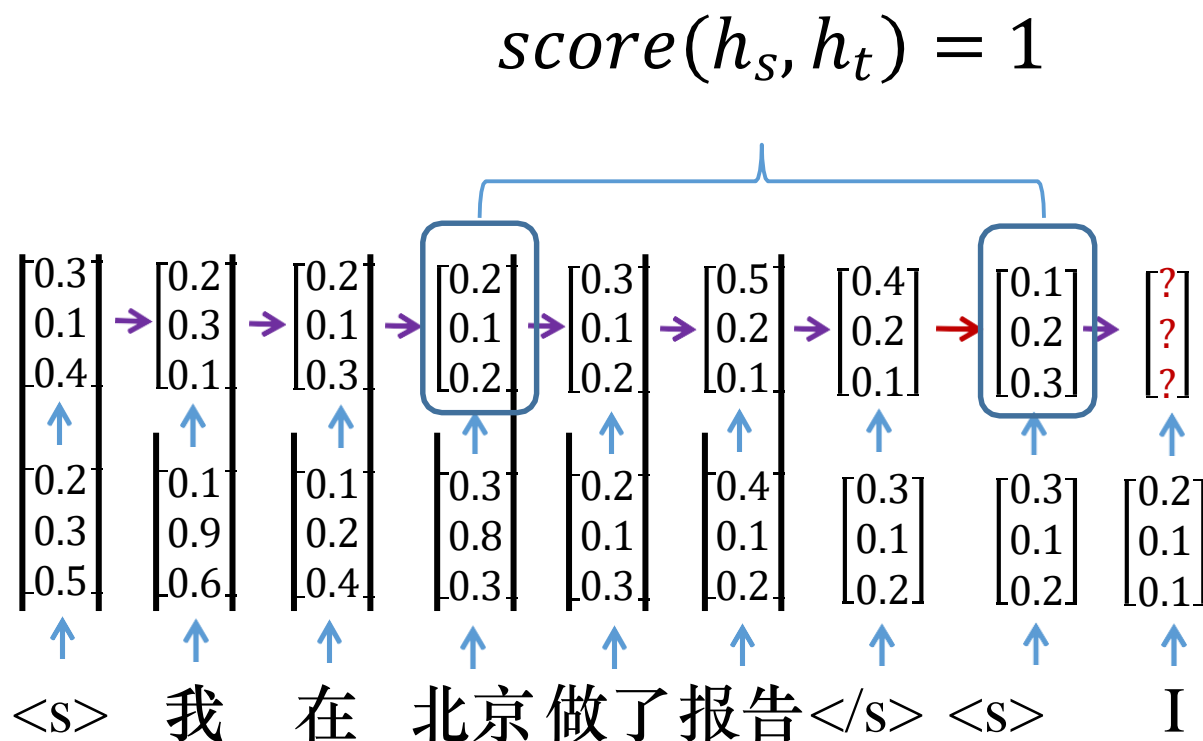
$$\text{score}(h_s, h_t) = 1$$



神经机器翻译-注意机制

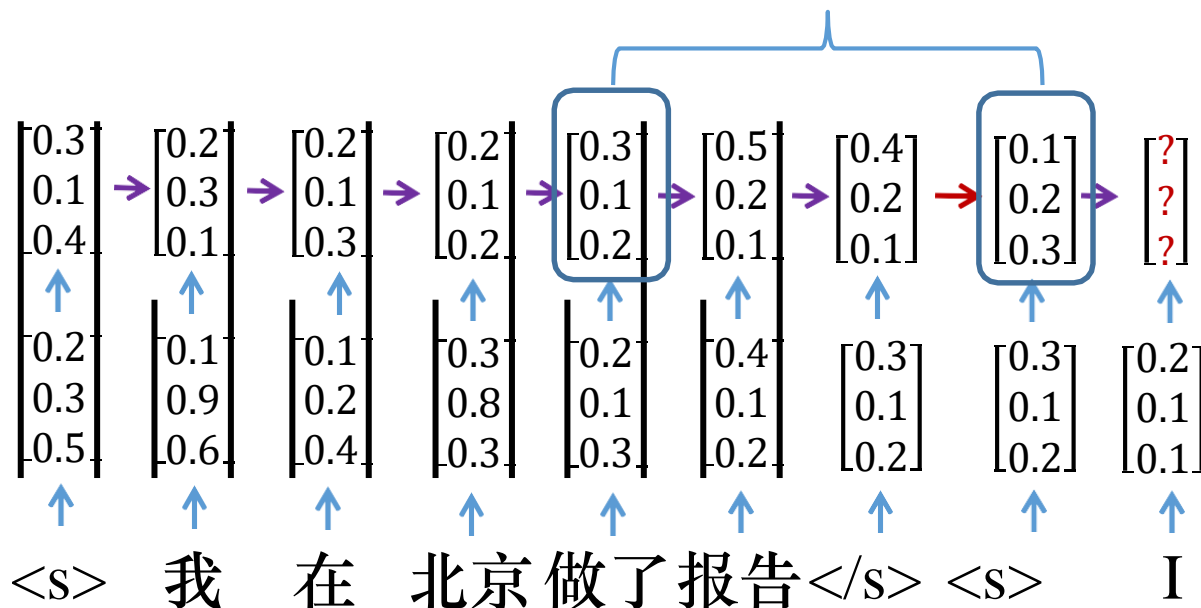


神经机器翻译-注意机制



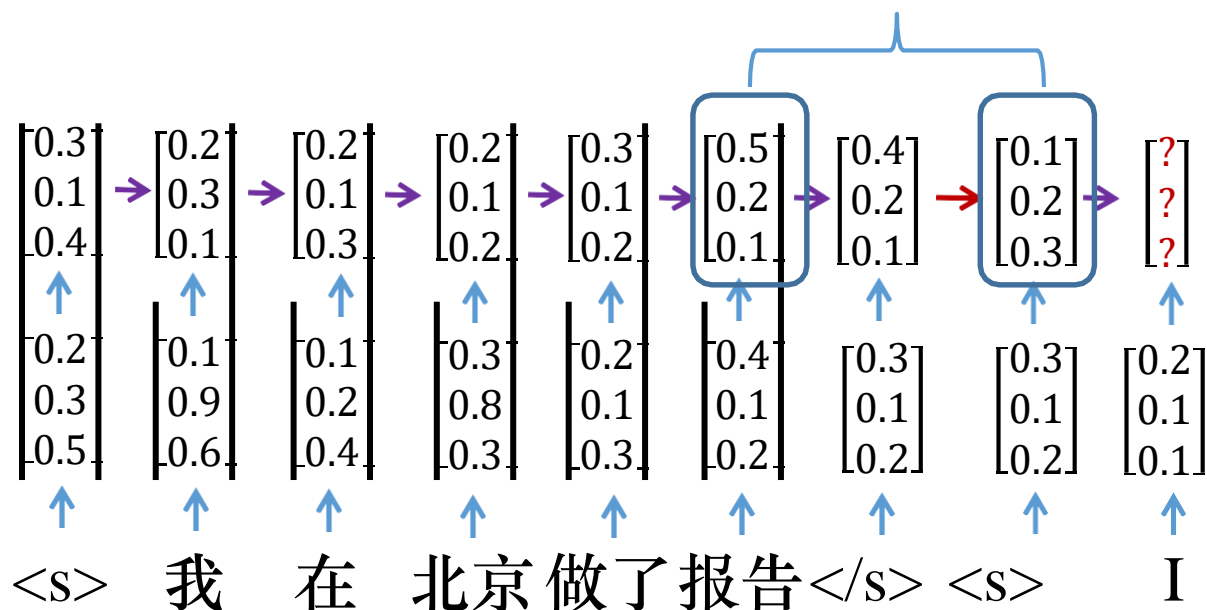
神经机器翻译-注意机制

$$\text{score}(h_s, h_t) = 4$$



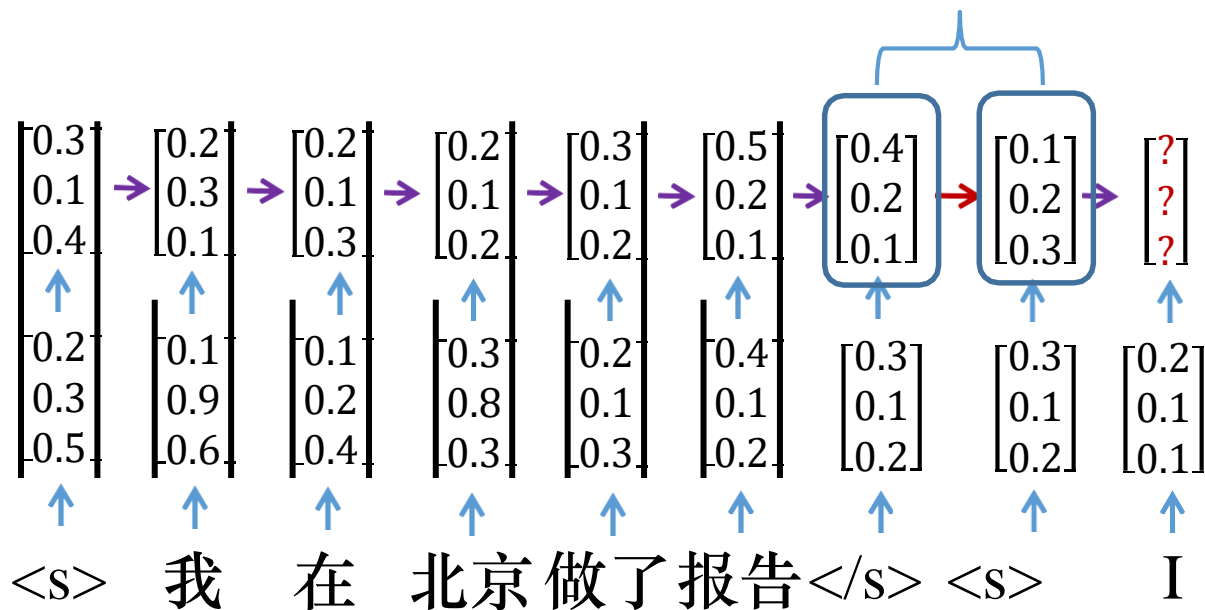
神经机器翻译-注意机制

$$\text{score}(h_s, h_t) = 2$$

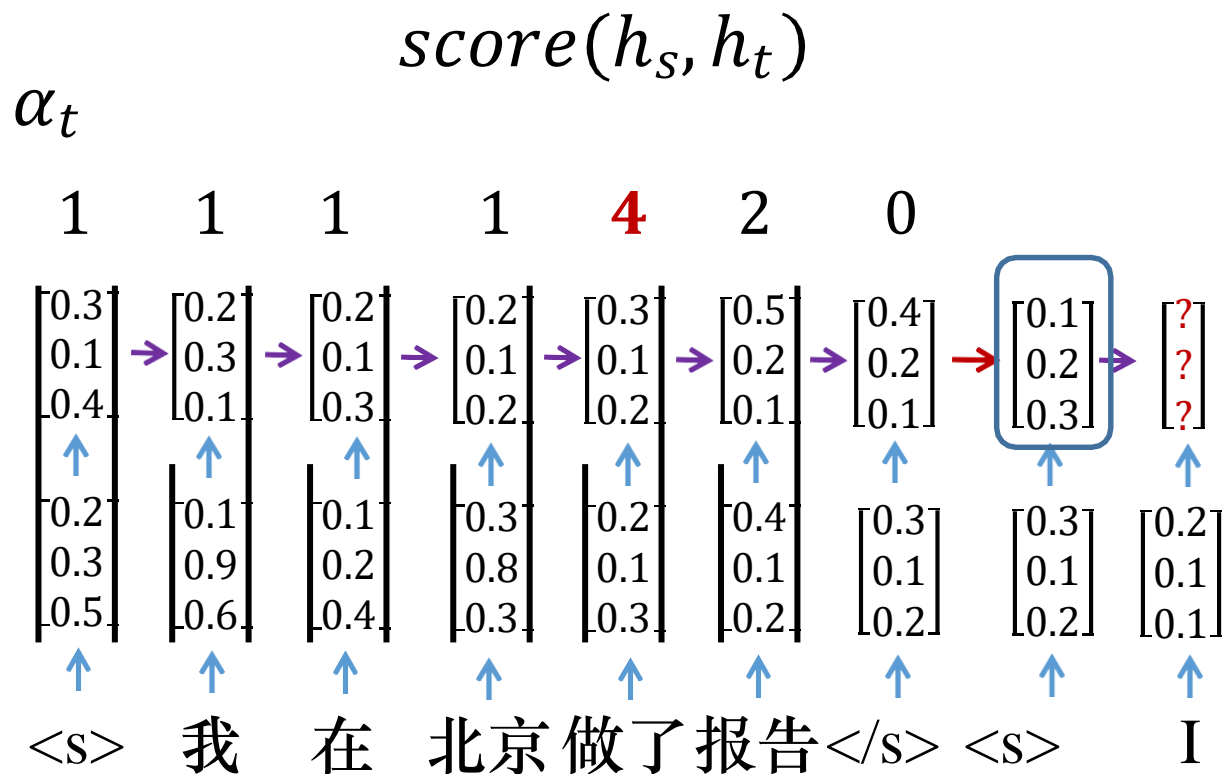


神经机器翻译-注意机制

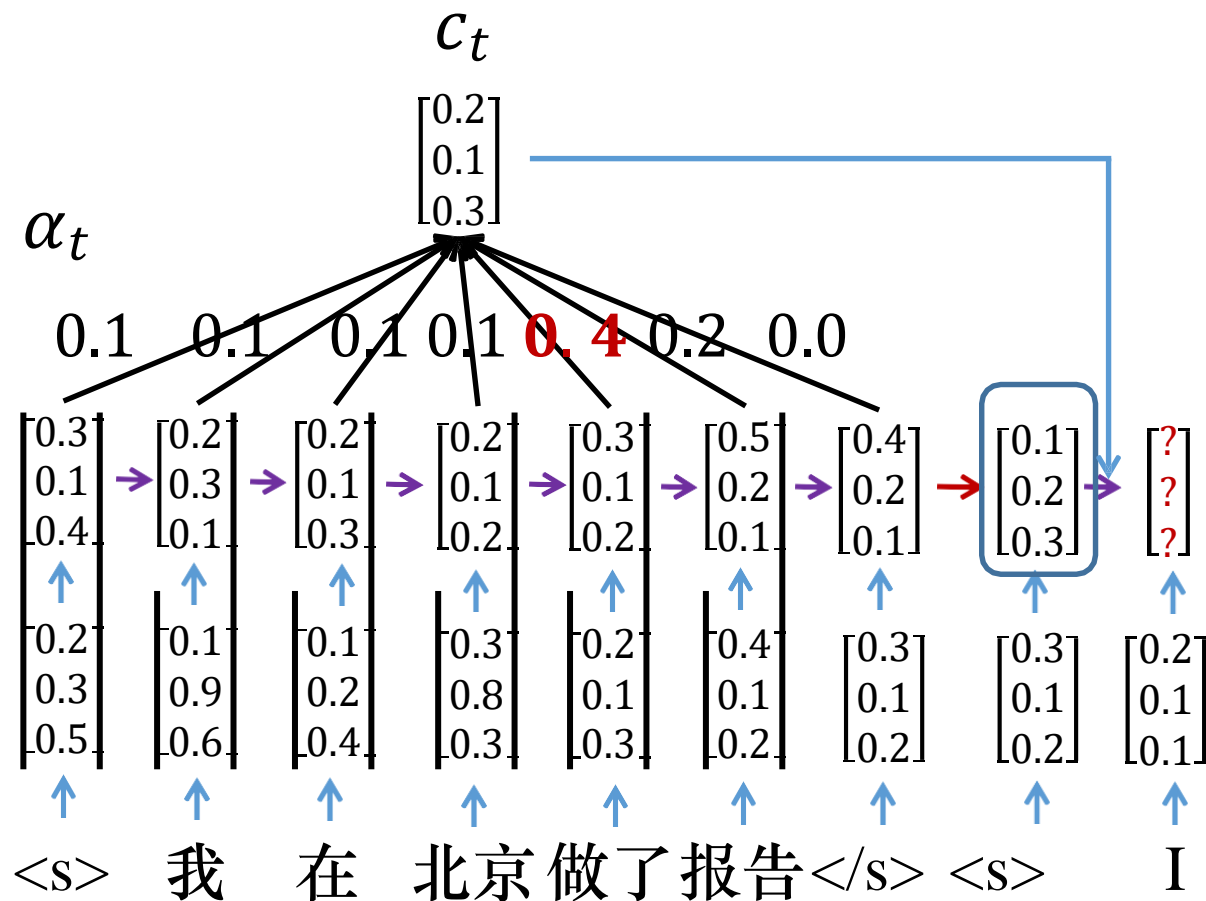
$$\text{score}(h_s, h_t) = 0$$



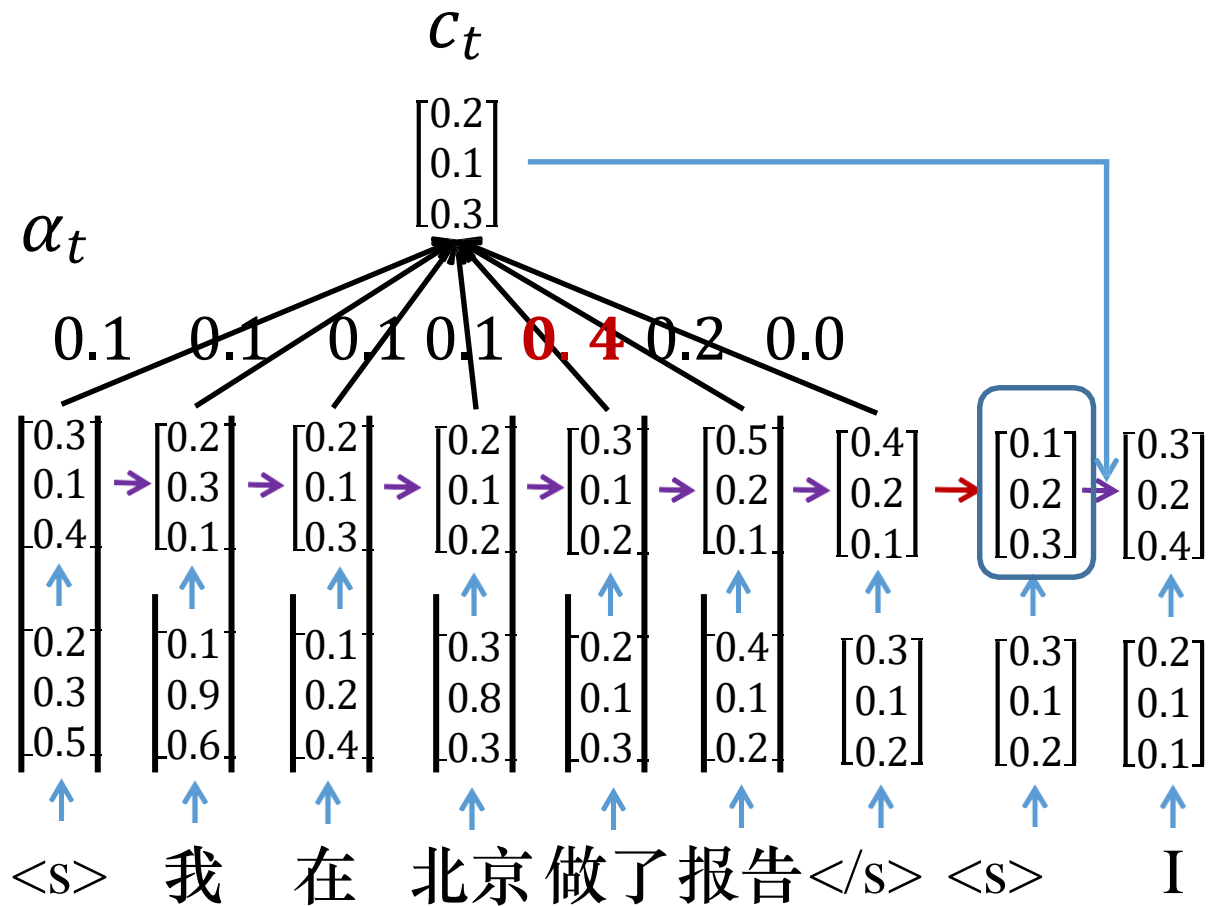
神经机器翻译-注意机制



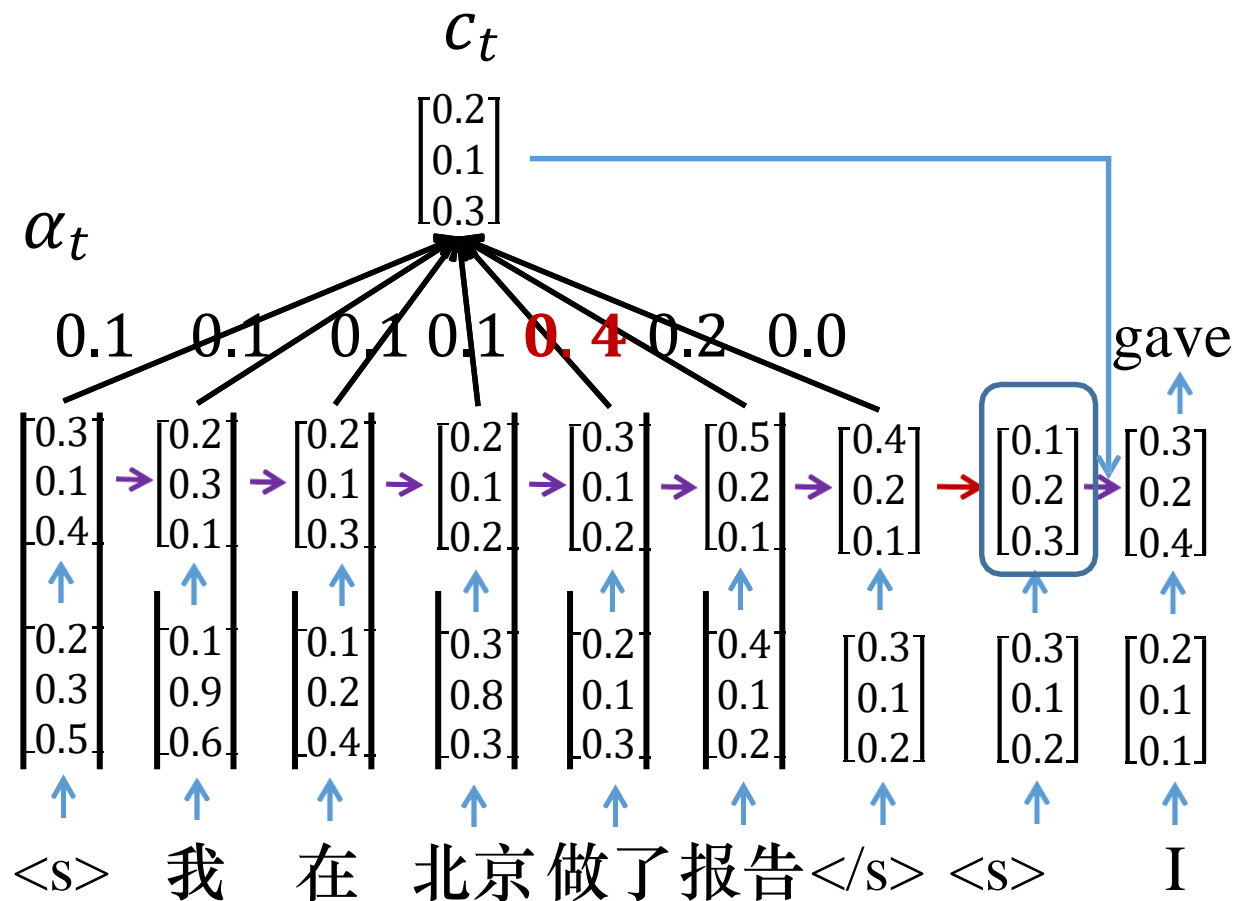
神经机器翻译-注意机制



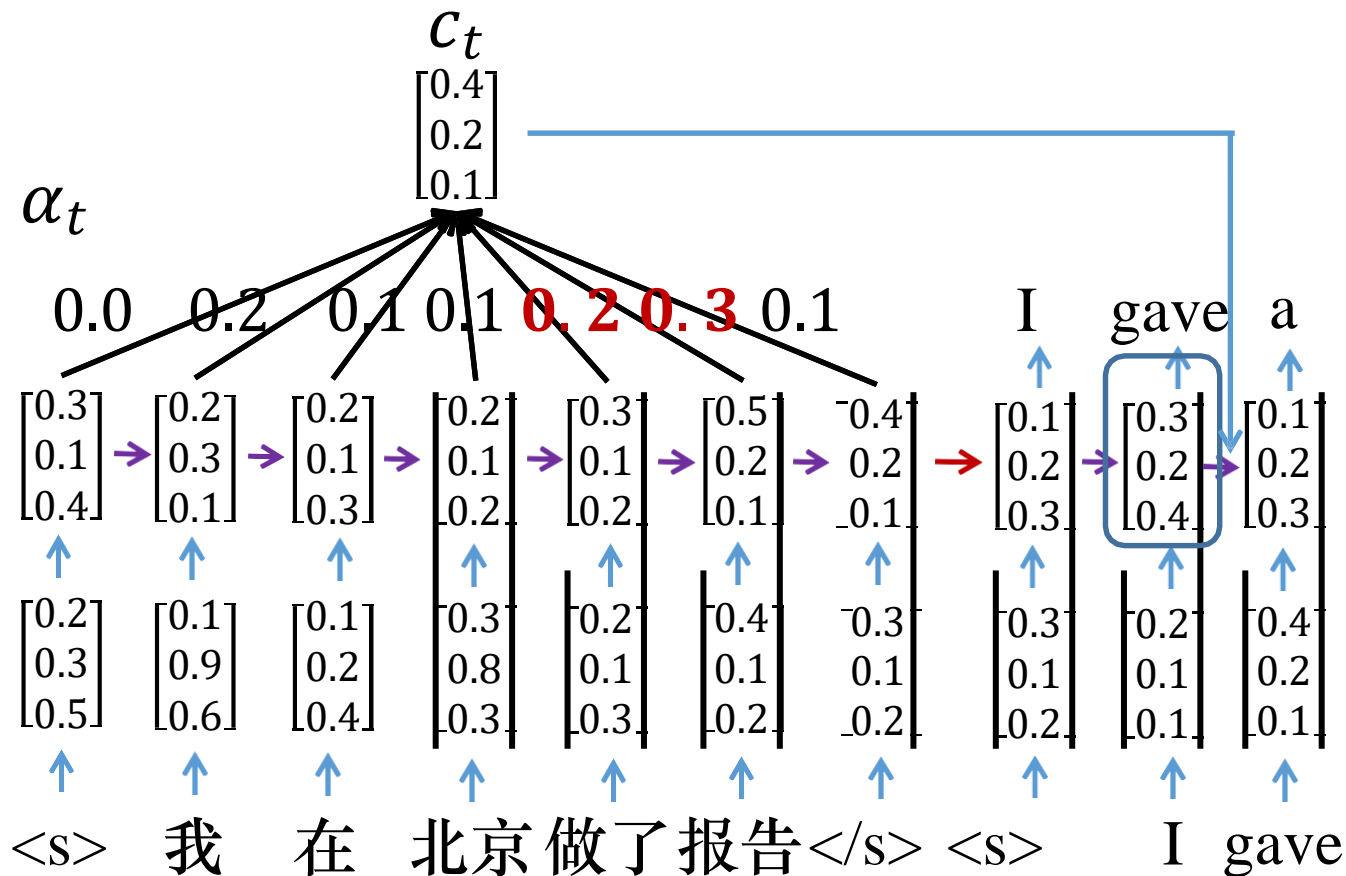
神经机器翻译-注意机制



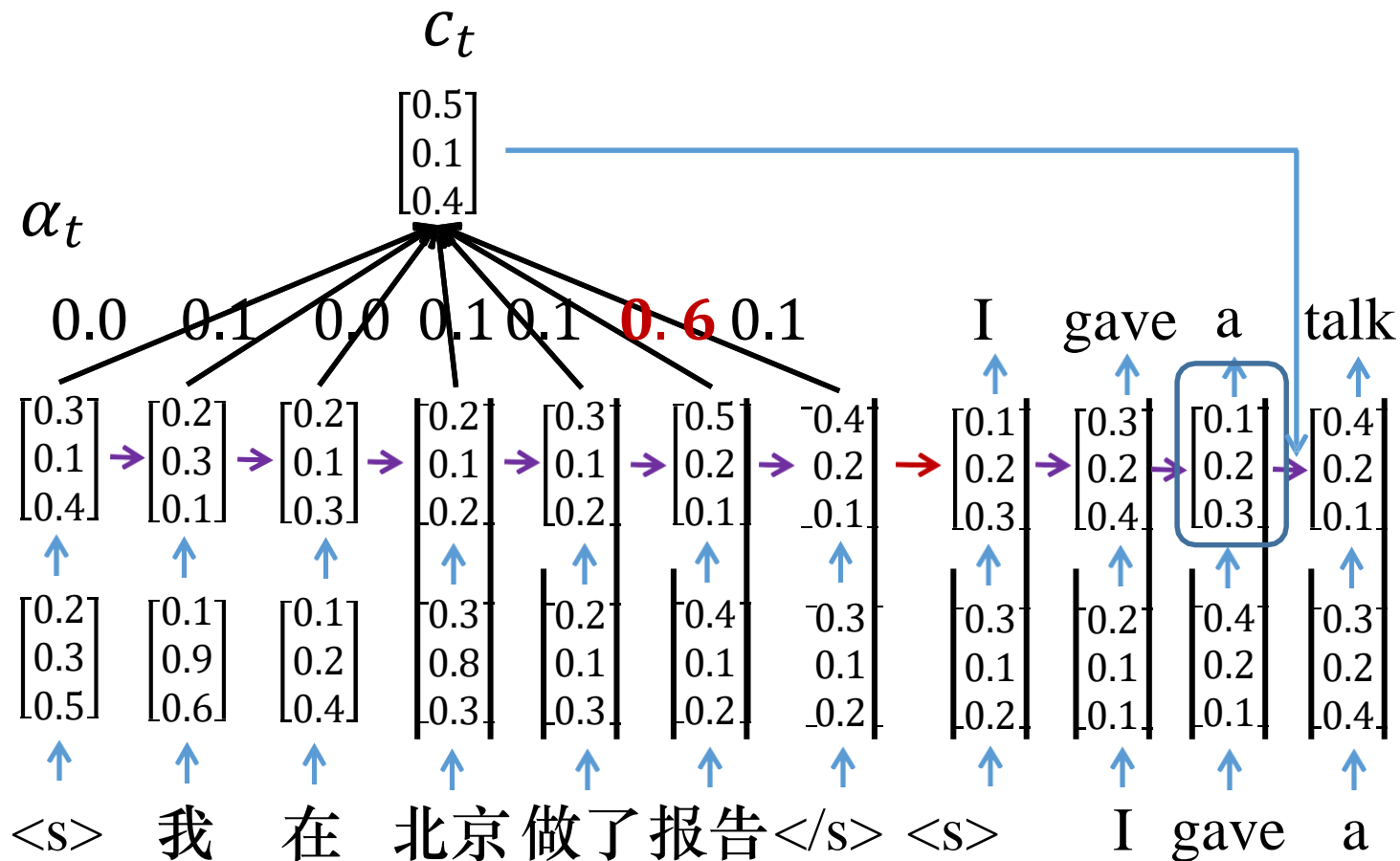
神经机器翻译-注意机制



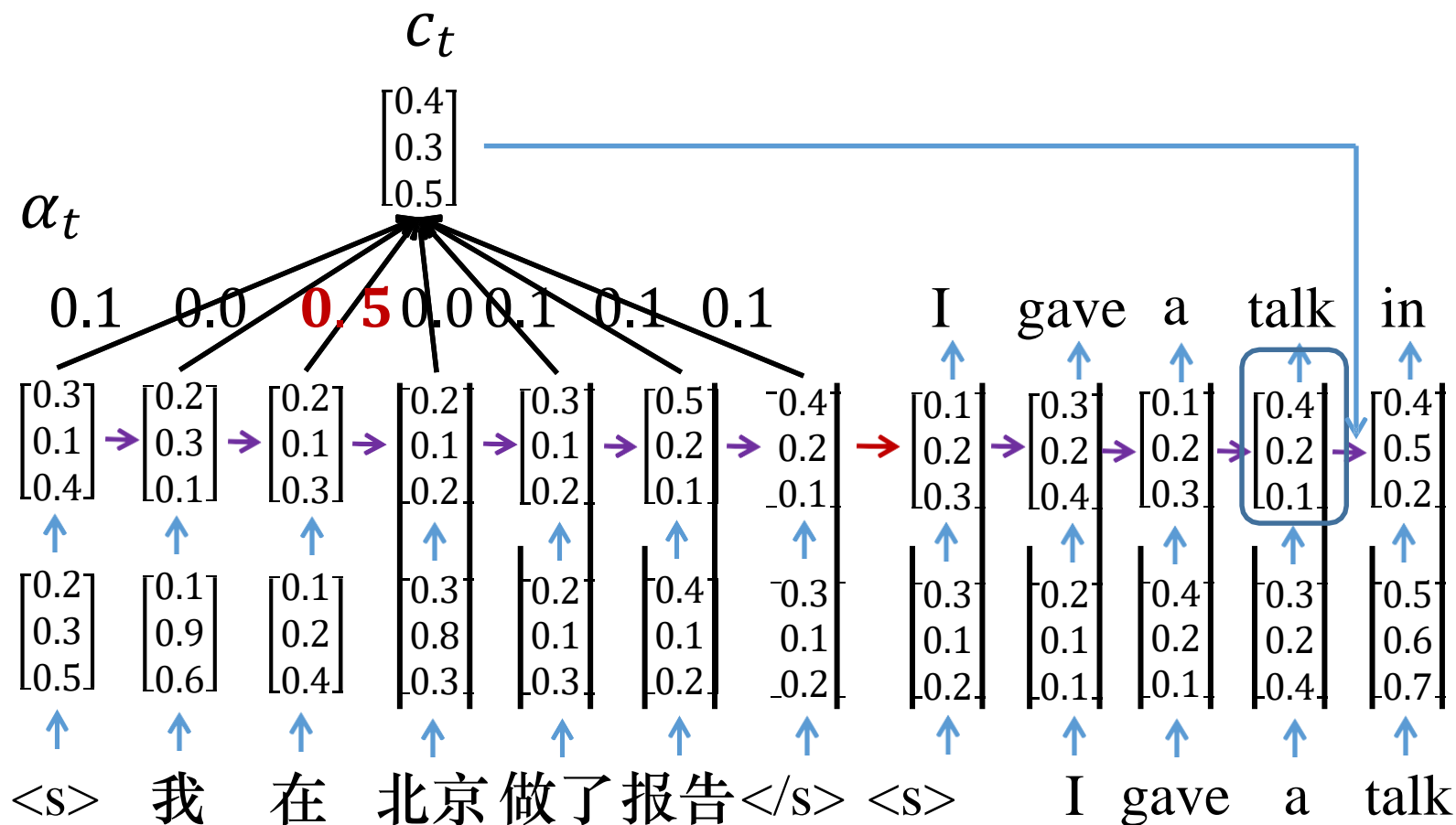
神经机器翻译-注意机制



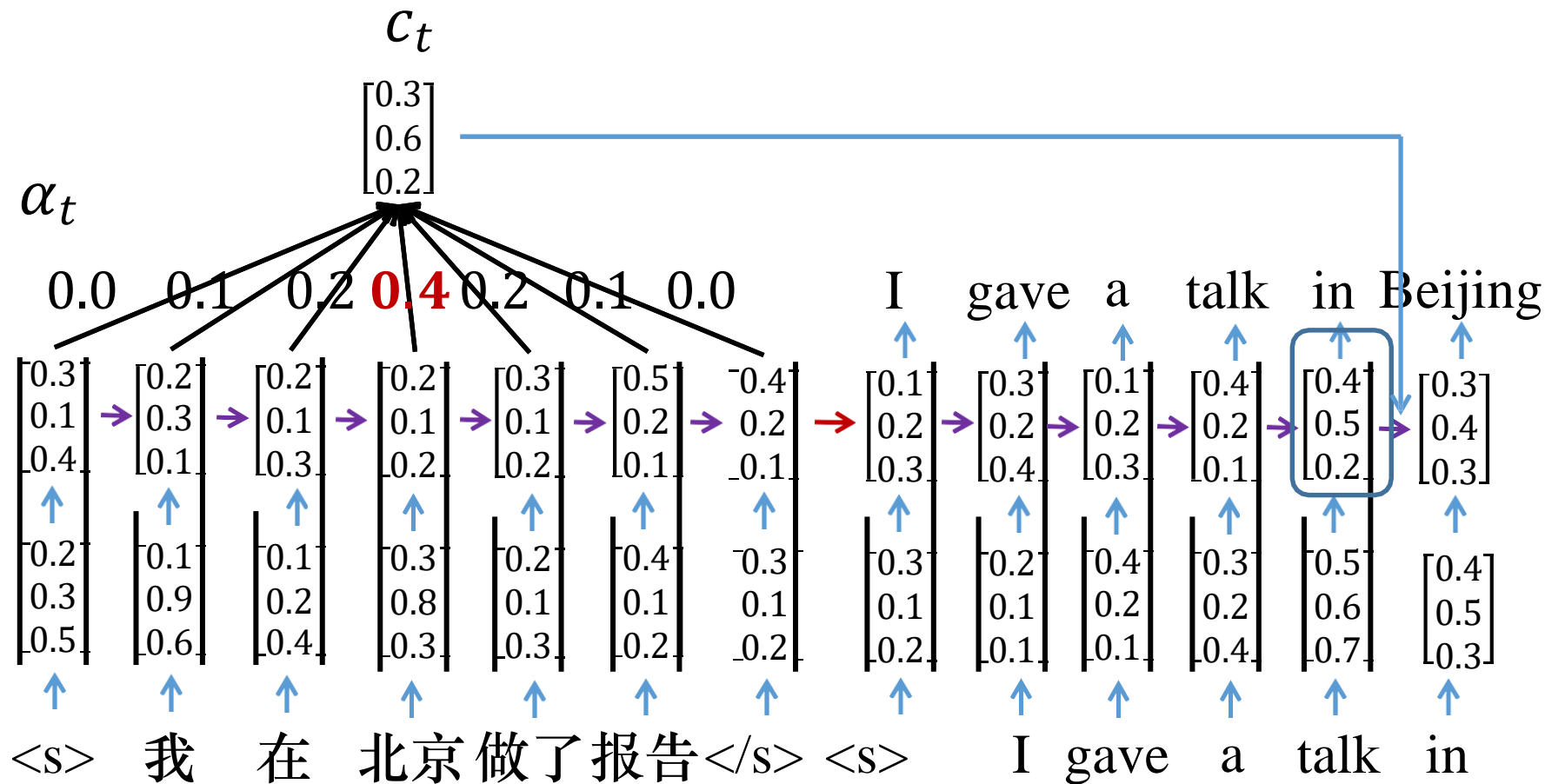
神经机器翻译-注意机制



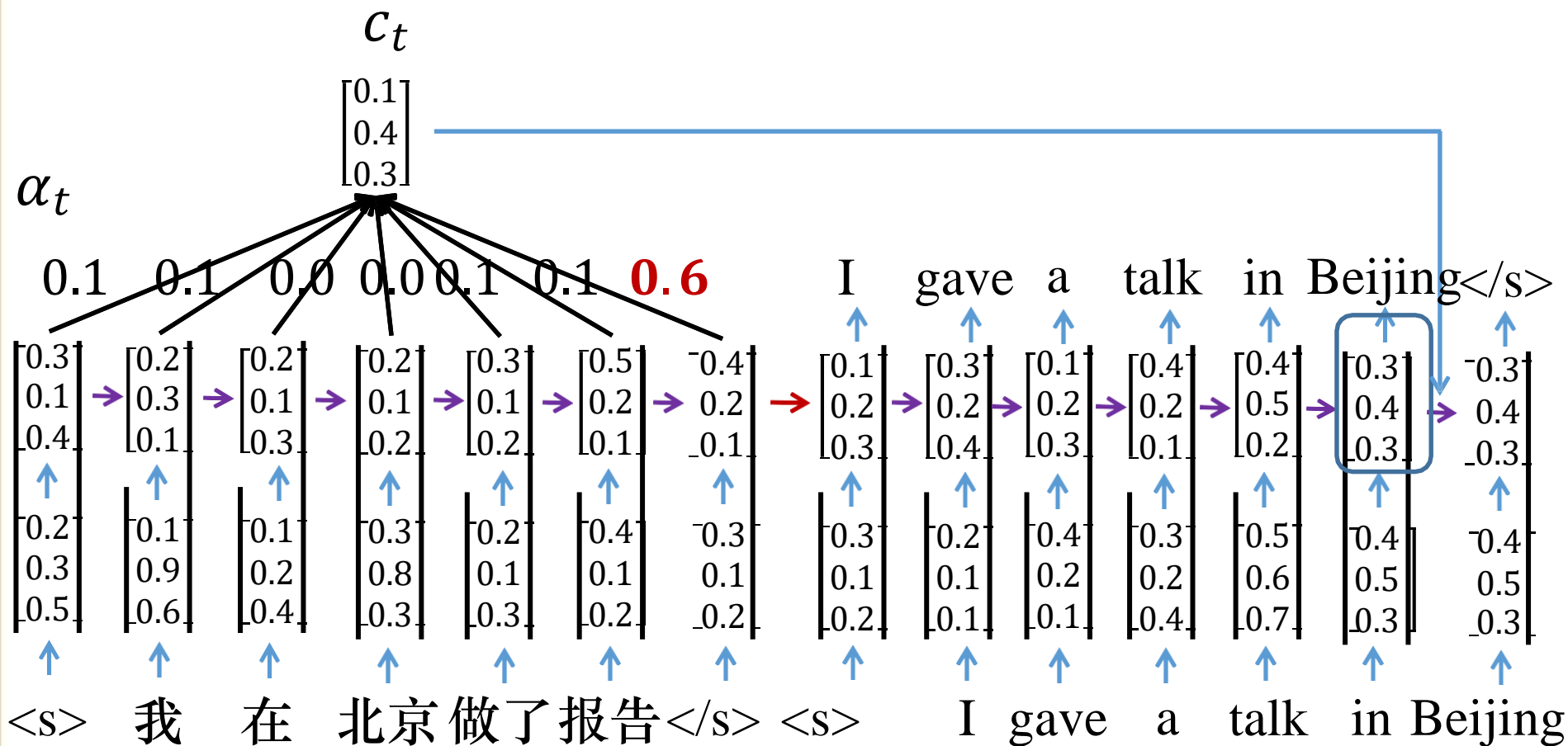
神经机器翻译-注意机制



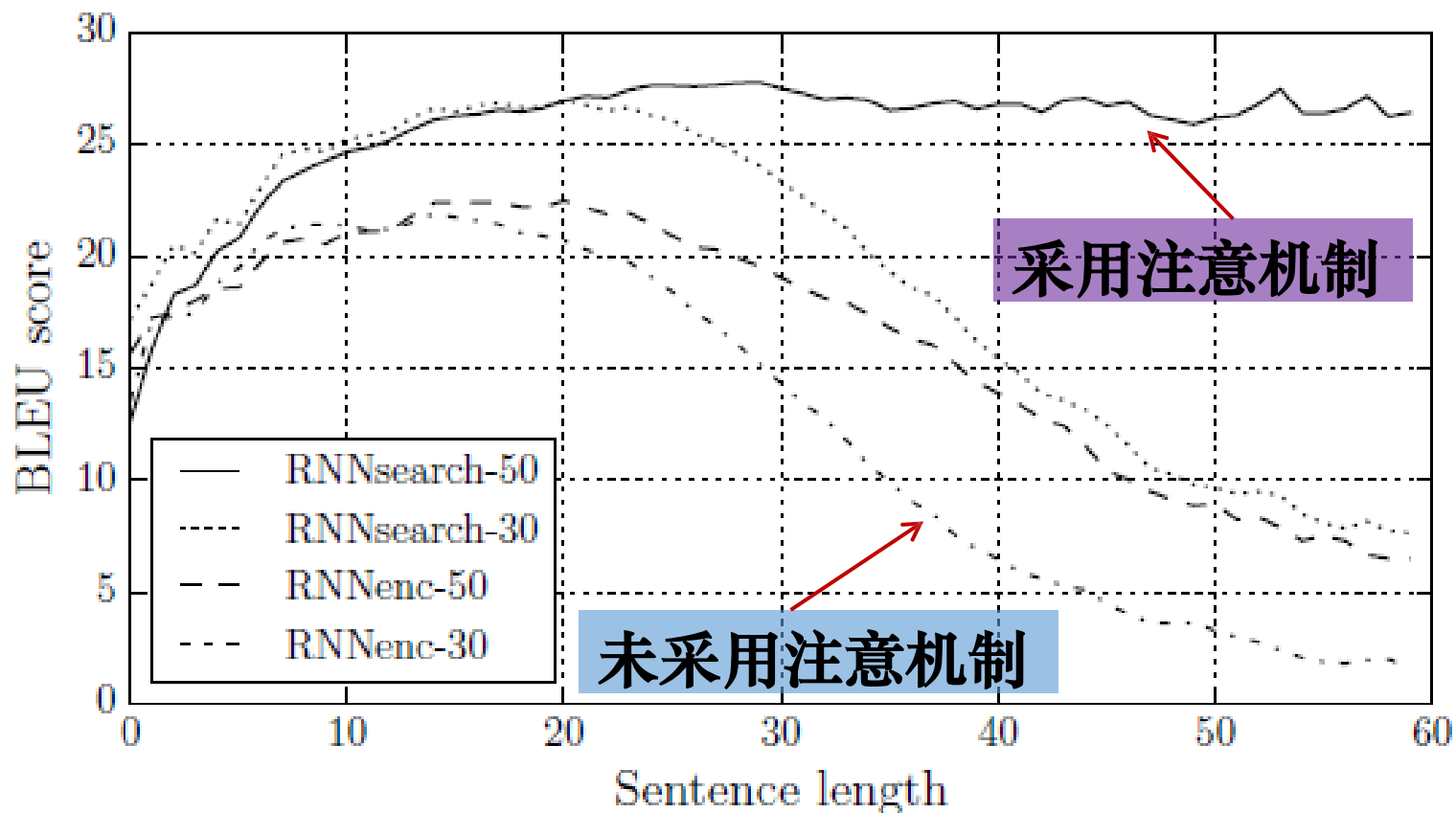
神经机器翻译-注意机制



神经机器翻译-注意机制



神经机器翻译-计算单元



RNNenc: 无注意机制, RNNsearch: 采用注意机制

翻译实例

south korean envoy calls for dialogue between the united states and north korea .

南韩
特使
呼吁
美国
与
北韩
对话



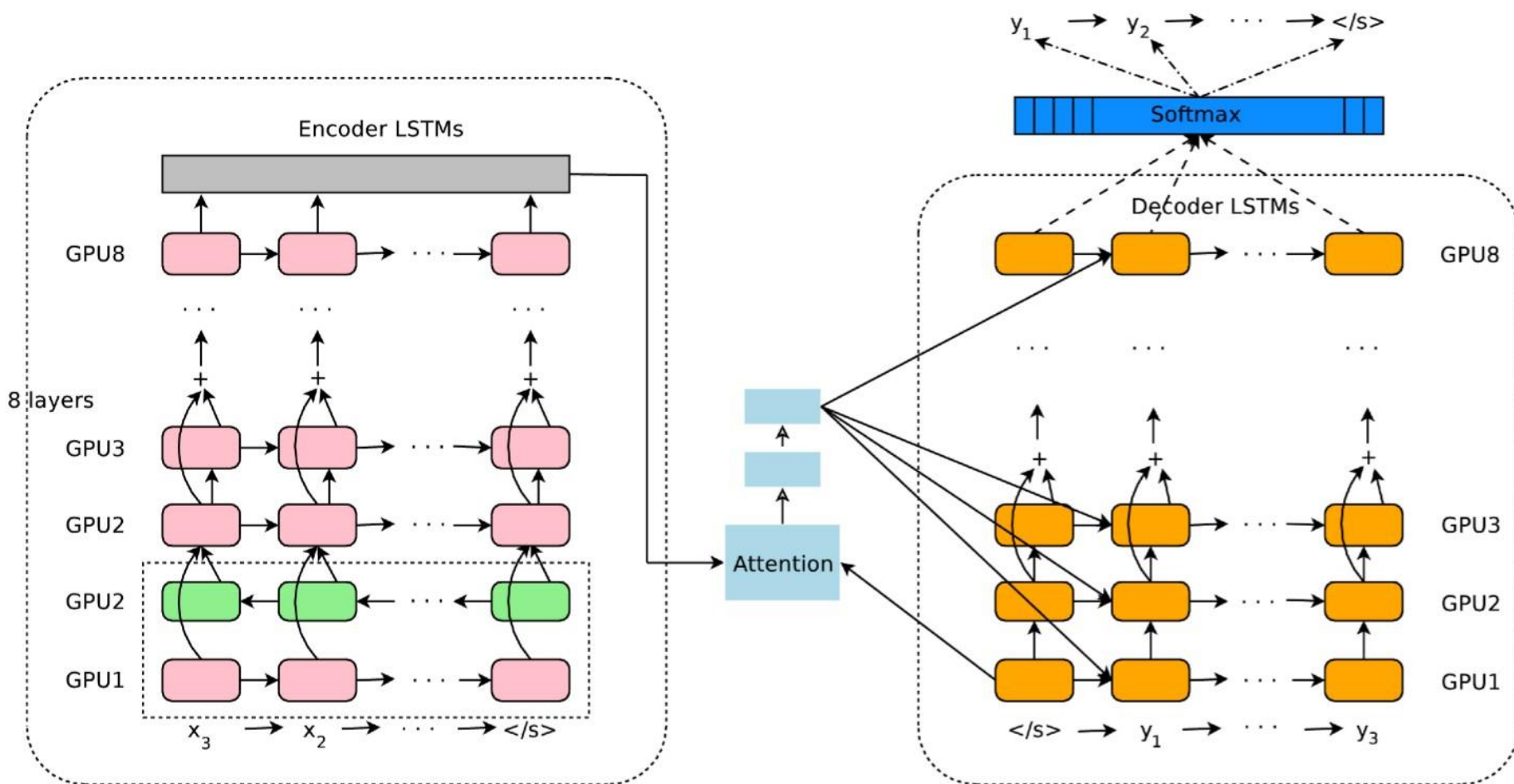
工业界线上产品



工业界线上产品



工业界线上产品



GNMT：谷歌神经翻译系统

工业界线上产品

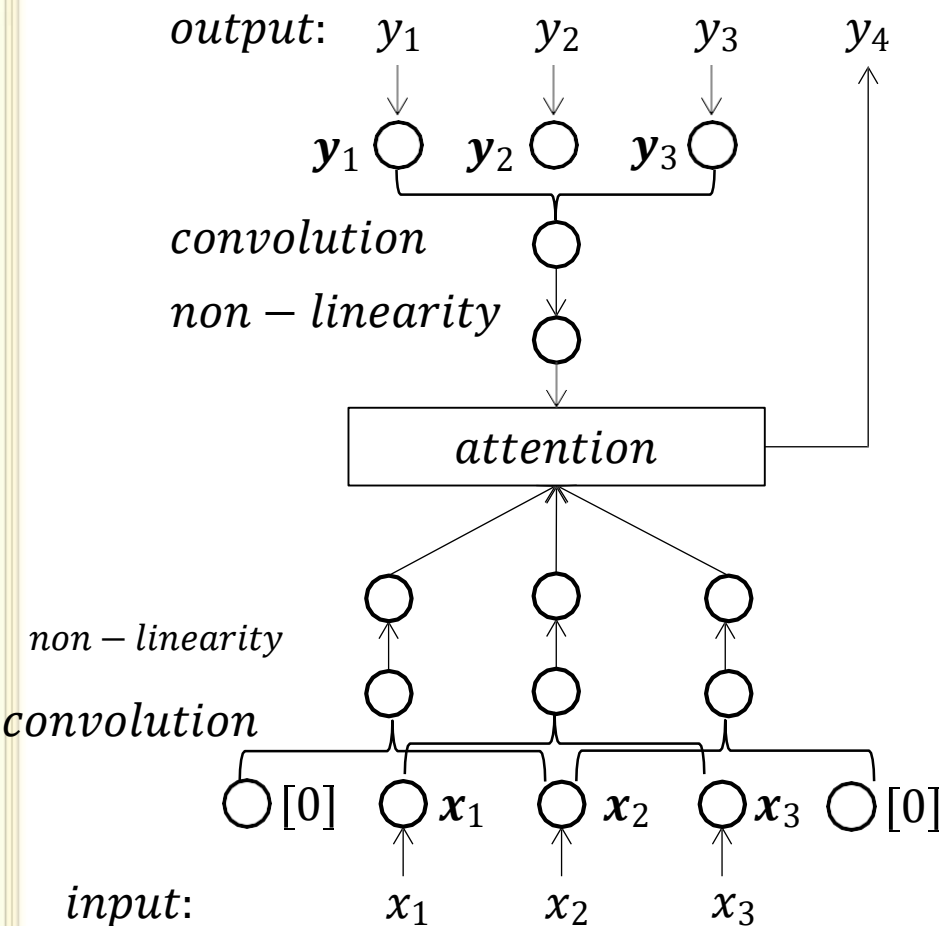
	PBMT	GNMT	Human	Relative Improvement
English → Spanish	4.885	5.428	5.550	87%
English → French	4.932	5.295	5.496	64%
English → Chinese	4.035	4.594	4.987	58%
Spanish → English	4.875	5.155	5.455	63%
French → English	5.046	5.343	5.404	83%
Chinese → English	3.694	4.263	4.636	60%

人工评测提升显著!

GNMT：谷歌神经翻译系统

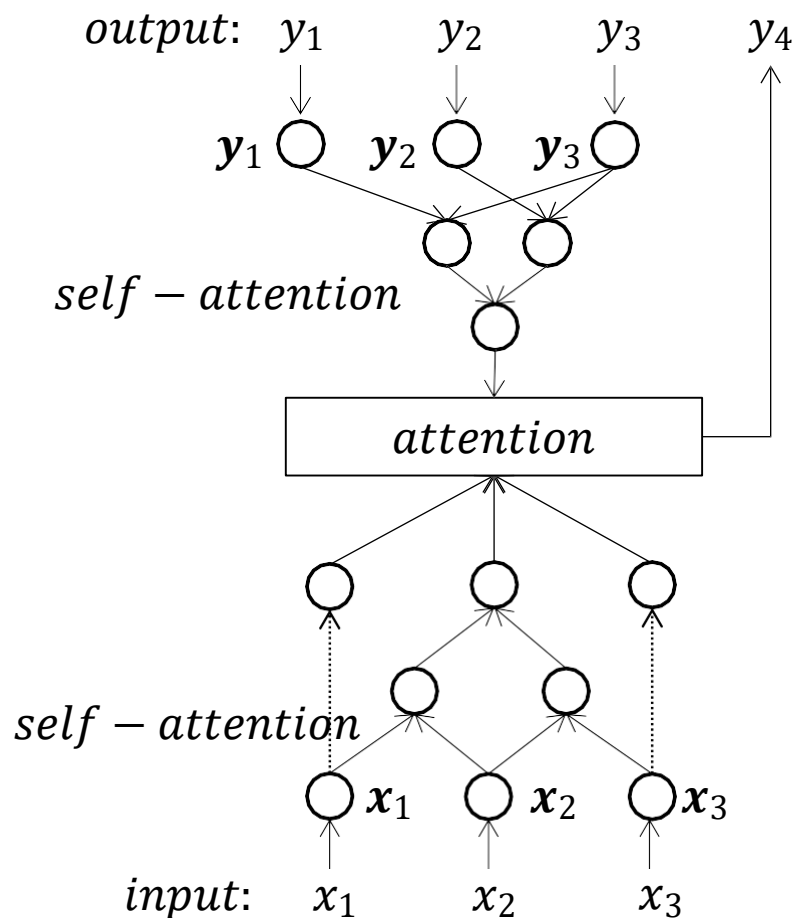
[Wu et al, 2016]

模型变革



(a) 基于卷积神经网络的翻译模型

CNMT : Facebook神经翻译系统



(b) 基于纯注意机制的翻译模型

Transformer : Google第二代

开源工具

1. [TensorFlow](#) (Transformer): 谷歌, python, C++/GPU
2. [ConvolutionalNMT](#): Facebook, Torch/GPU
3. [OpenNMT](#): Systran+哈佛, Torch/GPU
4. [GroundHog](#): 加拿大蒙特利尔大学, python/GPU
5. [dl4mt](#): 美国纽约大学, python/GPU
6. [Paddle](#): 百度, C++/GPU
7. [Zoph_RNN](#): 美国南加州大学, C++/GPU
8. [EUREKA-MangoNMT](#): 中科院自动化所, C++/CPU
9. [Nematus](#): 爱丁堡大学, C++/GPU
-

机器翻译技术落地

- 在线翻译（谷歌、微软、百度、有道、搜狗等）
- 翻译机（科大讯飞、准儿、百度、搜狗等）
- 同传机器翻译（微软、讯飞、腾讯、搜狗等）
 - 基于PowerPoint的语音同传（微软，TAUS 3.22-23）
 - 面向自由说话人的语音同传（腾讯，博鳌亚洲论坛 4.8-11）

未来展望

- 神经机器翻译采用编码解码网络，简单有效，已逐渐取代统计机器翻译，成为主流研究范式
- 神经机器翻译仍面临诸多问题
 - 缺乏可解释性
 - 难利用先验知识、语言相关知识
 - 训练、测试复杂度高（需GPU、甚至TPU）
 - 领域、场景迁移性能差

未来展望

➤ 未来发展

- 神经机器翻译的可解释性研究
- 与专家知识、常识知识的融合研究
- 场景、领域的迁移和定制化研究
- 面向资源稀缺语言的机器翻译建模
- 多模态机器翻译（语音和文本的一体化）研究
- 与硬件的一体化研究

Thank you!

权小军 中山大学数据科学与计算机学院