机器翻译研究热点与前沿趋势分析

苏劲松

厦门大学

2018.10.26 CWMT2018

17-18 NMT长文论文分布

会议/期刊	2017	2018
ACL	14	12
EMNLP	15	22
NAACL	---	14
COLING	---	18
TACL	4	3
TASLP	2	5
AAAI	3	10
IJCAI	4	2
ICLR	3	6
NIPS	3	---
ICML	1	1
总计	49	93

研究方向

- 词汇表受限
- 资源受限NMT
- 模型
  - 模型架构（RNMT，Transformer，ConvS2S）
  - 模型对比和分析
  - 模型训练

1.词汇表受限

- 未登录词处理
- 词汇表采样
- 细粒度单元

细粒度单元

字词混合

- Attention-via-Attention Neural Machine Translation. Zhao and Zhang, AAAI2018.
- Improving Character-based Decoding Using Target-Side Morphological Information for Neural Machine Translation. Passban et al., NAACL2018.
- Combining Character and Word Information in Neural Machine Translation Using a Multi-Level Attention. Chen et al., NAACL2018.
- Revisiting Character-Based Neural Machine Translation with Capacity and Compression. Ling et al., EMNLP2018.

细粒度单元

- 子词
  - Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates. Kudo, ACL2018.
  - Improving Neural Machine Translation by Incorporating Hierarchical Subword Features. Morishita et al., COLING2018.
- 词干+后缀
  - Improved English to Russian Translation by Neural Suffix Prediction. Song et al., AAAI2018.
  - Tailoring Neural Architectures for Translating from Morphologically Rich Languages. Passban et al., COLING2018.

2.资源受限NMT

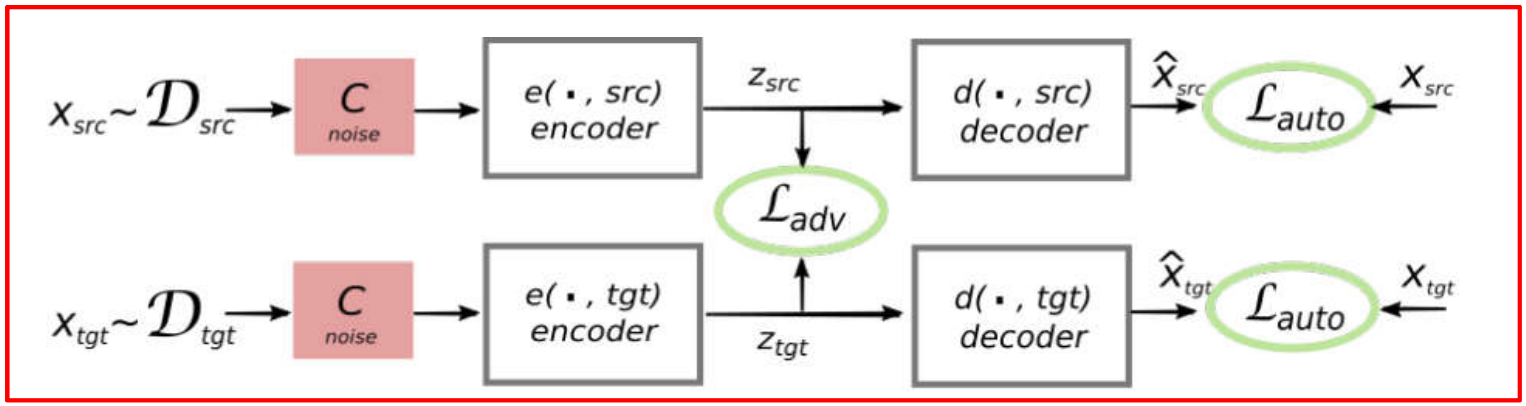
- 无监督 NMT
- 半监督 NMT
- 基于枢轴的 NMT
- 领域自适应
- 多模态
- 多语言多任务

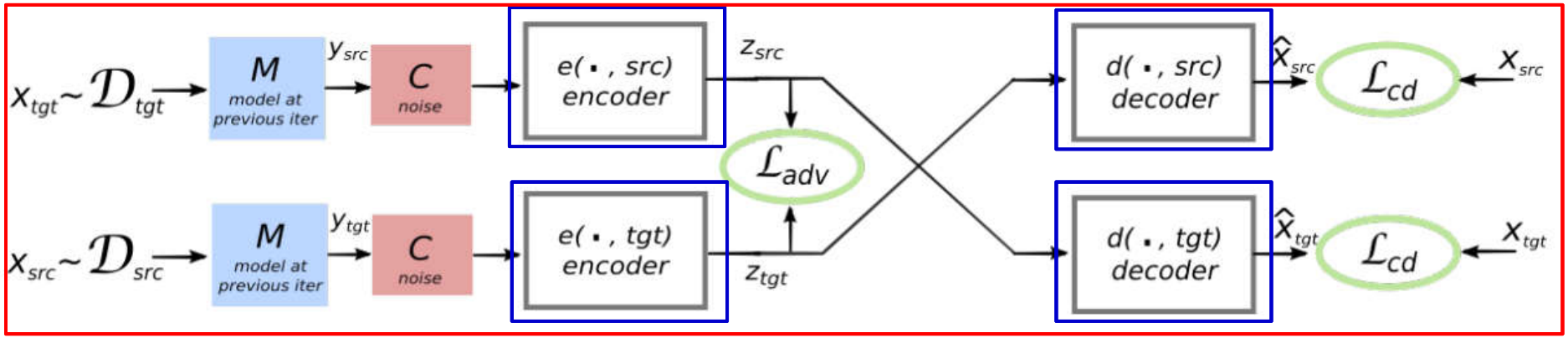
无监督NMT

- Unsupervised Machine Translation Using Monolingual Corpora Only. Lample et al., ICLR2018.
- Unsupervised Neural Machine Translation. Artetxe et al., ICLR2018.
- Unsupervised Neural Machine Translation with Weight Sharing. Yang et al., ACL2018
- Phrase-Based & Neural Unsupervised Machine Translation. Lample et al., EMNLP2018.

无监督NMT

单语自编码器





Back-Translation

Unsupervised Machine Translation Using Monolingual Corpora Only. Guillaume Lample, Alexis Conneau, Ludovic Denoyer, Marc'Aurelio Ranzato, *ICLR2018*.

无监督NMT

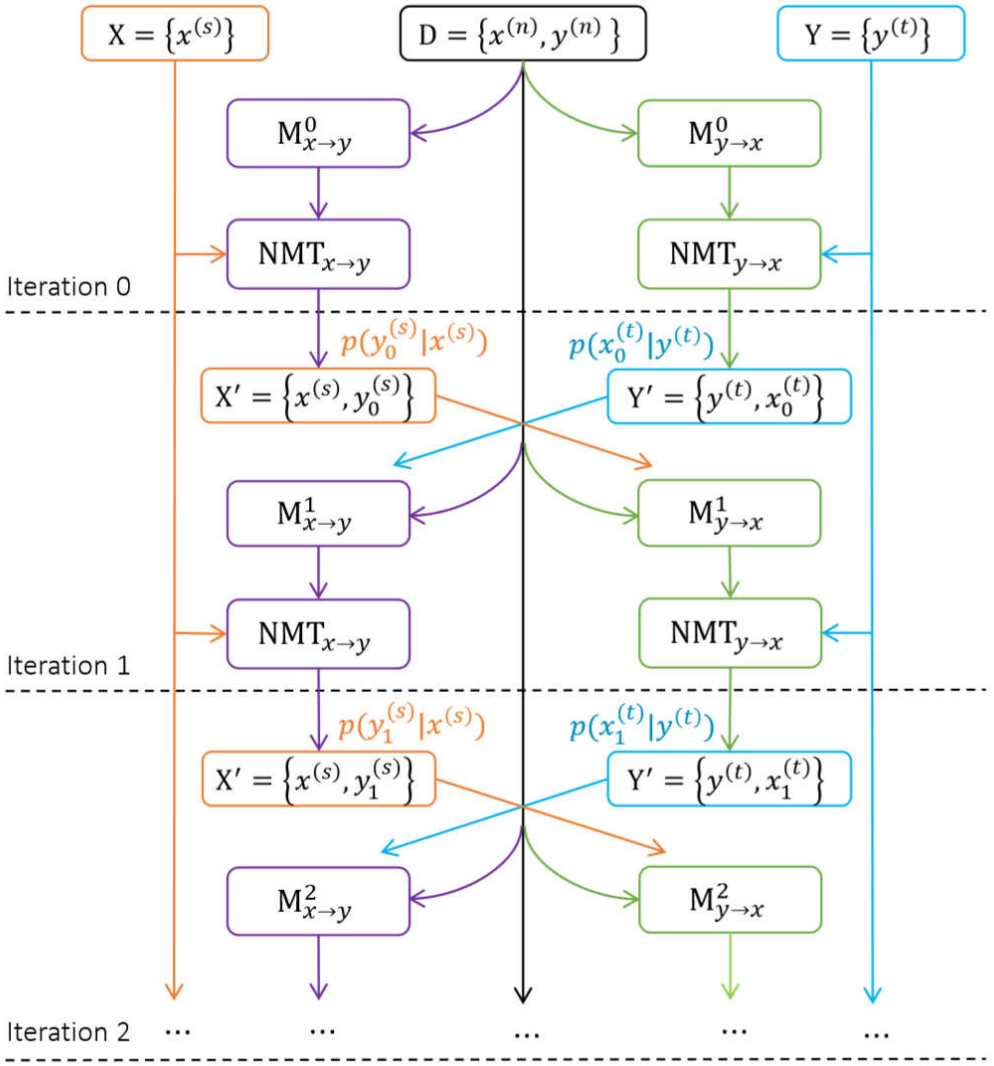
- 三个关键点
  - 词向量初始化
  - 语言建模：降噪自编码器
  - 迭代式的反向翻译
- **提升11个BLEU值**
  - 不同语言共享BPE表
  - 共享编码器，解码器参数
  - NMT和SMT的集成

Phrase-Based & Neural Unsupervised Machine Translation. Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, Marc'Aurelio Ranzato, *EMNLP2018*. Best Paper Award!

半监督NMT

- **单向NMT训练**
  - Back-Translation Sampling by Targeting Difficult Words in Neural Machine Translation. Fadaee and Monz., EMNLP2018.
- **双向NMT训练**
  - Joint Training for Neural Machine Translation Models with Monolingual Data., Zhang et al. AAAI2018.
  - Dual Transfer Learning for Neural Machine Translation with Marginal Distribution Regularization. Wang et al., AAAI2018.

半监督NMT



Joint Training for Neural Machine Translation Models with Monolingual Data. Zhirui Zhang, Shujie Liu, Mu Li, Ming Zhou, Enhong Chen, *AAAI2018.*

基于枢轴的NMT

- Zero-Resource Neural Machine Translation with Multi-Agent Communication Game. Chen et al., AAAI2018.
- Triangular Architecture for Rare Language Translation. Ren et al., ACL2018.

领域自适应

- 语料角度
  - Sentence Weighting for Neural Machine Translation Domain Adaptation. Zhang et al., COLING2018.
  - Sentence Selection and Weighting for Neural Machine Translation Domain Adaptation. Wang et al., TASLP2018.
- 模型
  - Multi-Domain Neural Machine Translation with Word-Level Domain Context Discrimination. Zeng et al., EMNNLP2018.

领域自适应

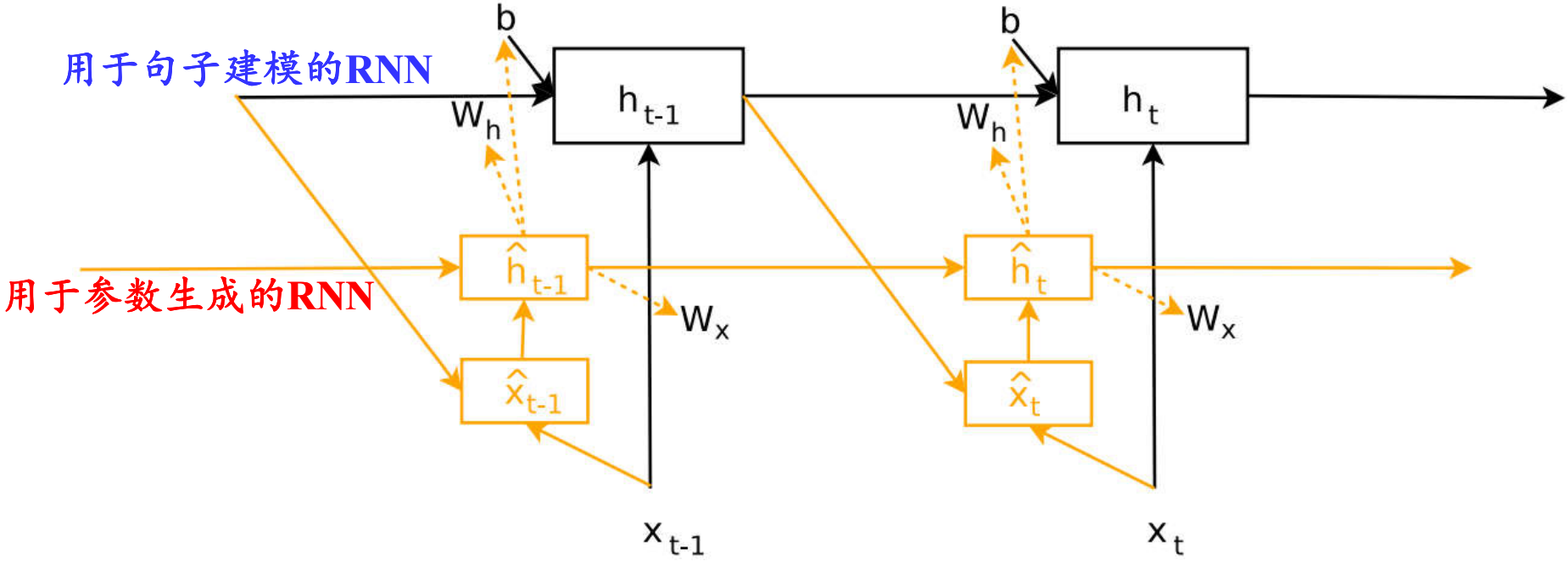
参数生成

- Hypernetworks. Ha et al., ICLR2017.
- Contextual Parameter Generation for Universal Neural Machine Translation. Platanios et al., EMNLP2018.

综述

- A Survey of Domain Adaptation for Neural Machine Translation. Chu and Wang., COLING2018.

领域自适应



1. 深层次信息共享，减少参数
2. 动态调整每步参数

Hypernetworks. David Ha, Andrew Dai, Quoc V. Le, *ICLR2017*.

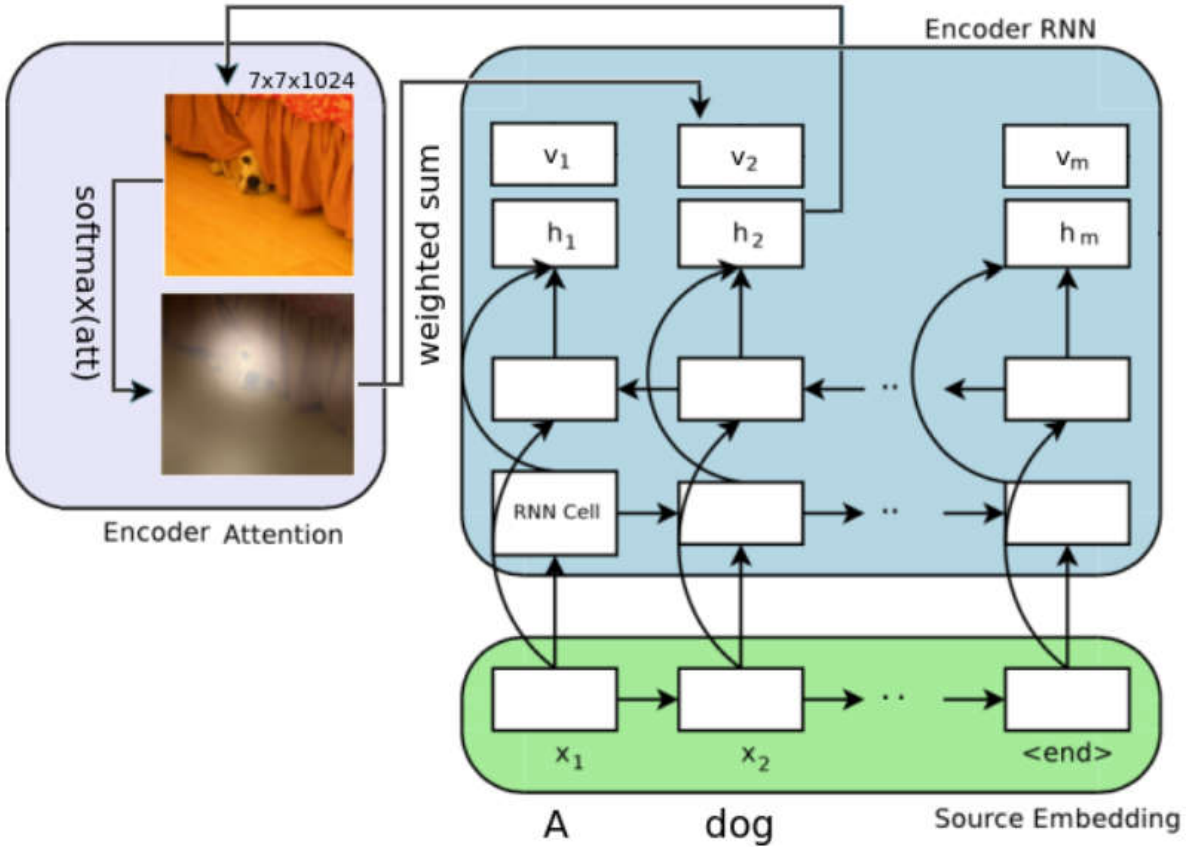
多模态NMT

- Modulating and Attending the Source Image During Encoding improves Multimodal Translation. Delbrouck and Dupont. NIPS2017.
- A Visual Attention Grounding Neural Model for Multimodal Machine Translation. Zhou et al., EMNLP2018.

多模态NMT

1. 利用文本的语义信息优化CNN图片特征提取

2. 源语言Annotation对CNN图片特征进行Attention，获得的表示作为源语言Annotation的补充



Modulating and Attending the Source Image during Encoding Improves Multimodal Translation. Jean-Benoit Delbrouck and Stéphane Dupont. *NIPS2017.*

多语言，多任务

- 多语言
  - Universal Neural Machine Translation for Extremely Low Resource Lan guages. Gu et al., NAACL2018.
  - Multilingual Neural Machine Translation with Task-Specific Attention. Blackwood et al., COLING2018.
  - Meta-Learning for Low-Resource Neural Machine Translation. Gu et al., EMNLP2018.
  - Three Strategies to Improve One-to-Many Multilingual Translation. Wang et al., EMNLP2018.

多语言，多任务

- 多任务
  - Scheduled Multi-Task Learning: From Syntax to Translation. Kiperwasser et al., TACL2018.
  - Neural Machine Translation for Bilingually Scarce Scenarios: A Deep Mul titask Learning Approach. ZareMoodi et al., NAACL2018.
  - Multi-Task Neural Models for Translating Between Styles Within and Across Languages. Niu et al., COLING2018.

3.模型

- 模型架构
  - **RNMT**
  - **Transformer**
  - **ConvS2S**
- 模型对比和分析
- 模型训练

RNMT

- 编码器
- 注意力机制
- 解码器
- 其它改进
- 未来信息建模
- 跨句子信息建模
- SMT知识
- 句法知识
- 其他外部知识

编码器改进

- A Hierarchy-to-Sequence Attentional Neural Machine Translation Model. Su et al., TASLP2018.
- Multi-Channel Encoder for Neural Machine Translation. Xiong et al., AAAI2018.
- Dense Information Flow for Neural Machine Translation. Shen et al., NAACL2018.
- Refining Source Representations with Relation Networks for Neural Machine Translation. Zhang et al., COLING2018.

注意力机制改进

- Syntax-Directed Attention for Neural Machine Translation. Chen et al., AAAI2018.
- Neural Machine Translation with Key-Value Memory-Augmented Attention. Meng et al., IJCAI2018.
- Target Foresight based Attention for Neural Machine Translation. Li et al., NAACL2018.
- Neural Machine Translation with Deep Attention. Zhang et al., TPAMI2018.

解码器改进

- 目标上下文
  - Neural Machine Translation with Decoding History Enhanced Attention. Wang et al., COLING2018.
  - Self-Attentive Residual Decoder for Neural Machine Translation. Werlen et al., NAACL2018.
- 译文生成
  - Towards Neural Phrase-based Machine Translation. Huang et al., ICLR2018.
- 译文搜索
  - Fast Lexically Constrained Decoding with Dynamic Beam Allocation for Neural Machine Translation. Post et al., NAACL2018.
  - Speeding Up Neural Machine Translation Decoding by Cube Pruning. Zhang et al., EMNLP2018.

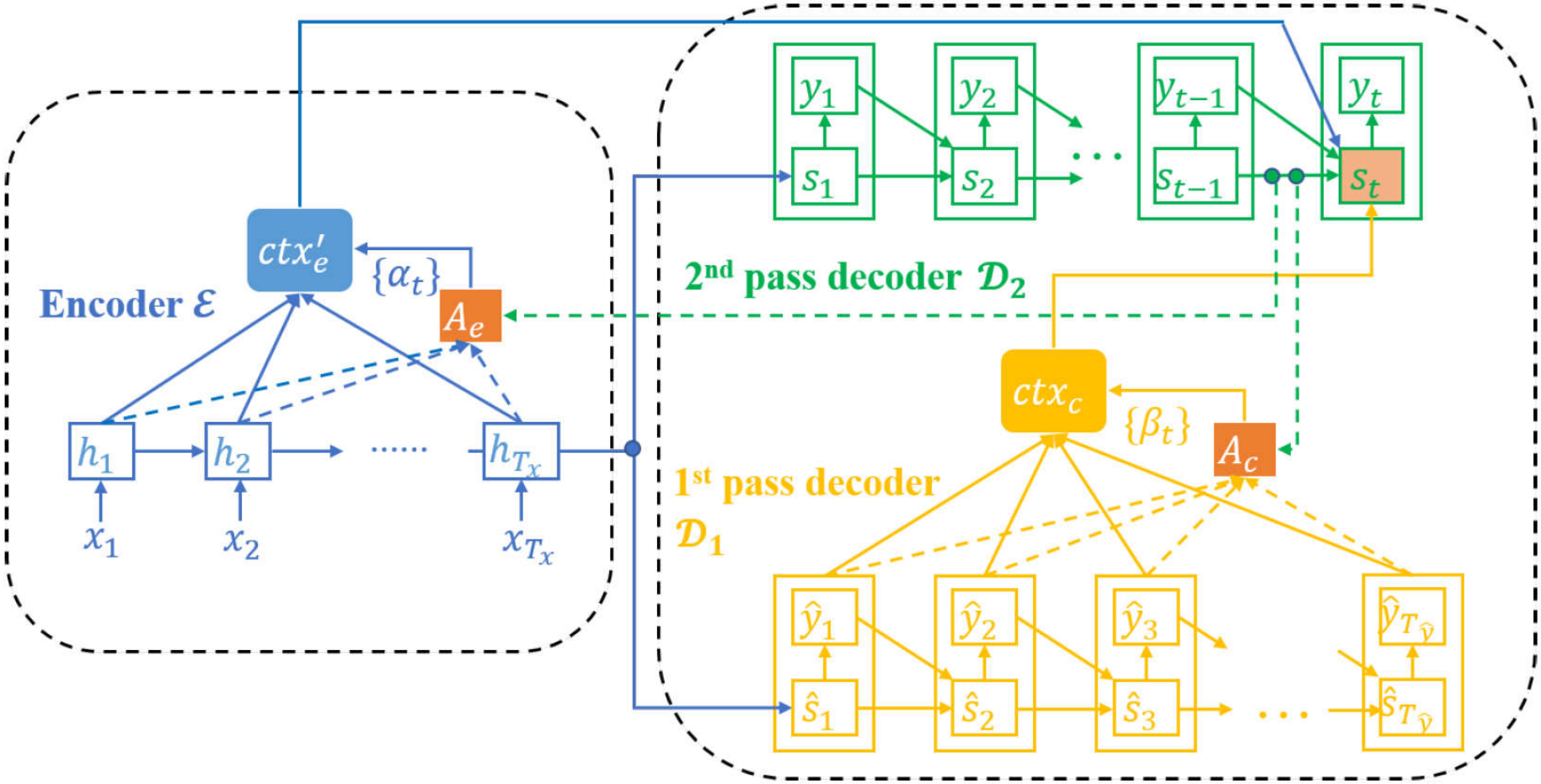
其它改进

- Translating Pro-Drop Languages with Reconstruction Models. Wang et al., AAAI2018.
- Attention Focusing for Neural Machine Translation by Bridging Source and Target Embeddings. Kuang et al., ACL2018.
- Improving Lexical Choice in Neural Machine Translation. Nguyen et al., NAACL2018.
- Handling Homographs in Neural Machine Translation. Liu et al., NAACL2018.
- Adaptive Weighting for Neural Machine Translation. Li et al., COLING2018.
- Simplifying Neural Machine Translation with Addition-Subtraction Twin-Gated Recurrent Networks. Zhang et al., EMNLP2018.

未来信息建模

- 多次解码
  - Deliberation Networks: Sequence Generation beyond One-pass Decoding. Xia et al., NIPS2017.
  - Asynchronous Bidirectional Decoding for Neural Machine Translation. Zhang et al., AAAI2018.
  - Adaptive Multi-pass Decoder for Neural Machine Translation with Reinforcement Learning. Geng et al., EMNLP2018.

未来信息建模

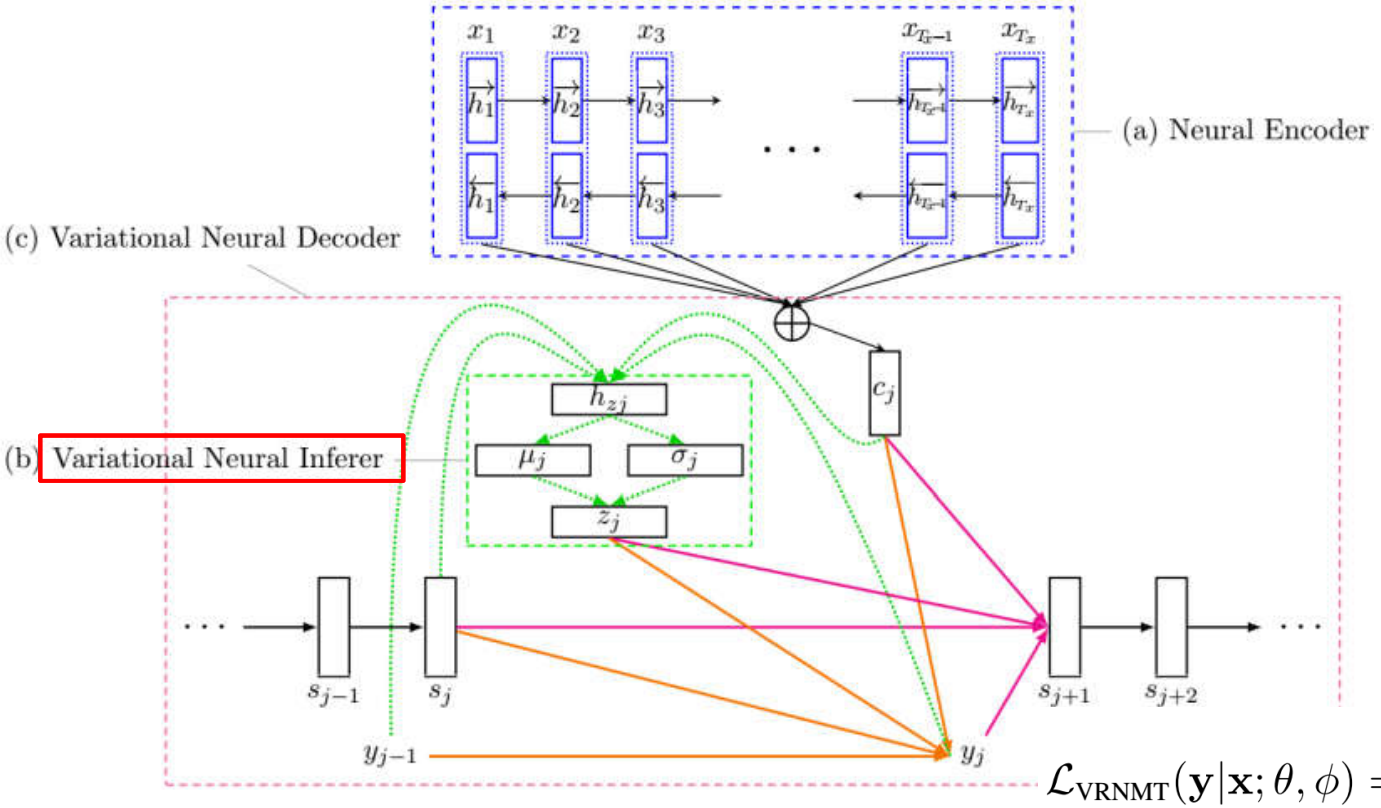


Deliberation Networks: Sequence Generation beyond One-pass Decoding. Yingce Xia, Fei Tian, Lijun Wu, Jianxin Lin, Tao Qin, Nenghai Yu, Tie-Yan Liu, *NIPS2017*.

未来信息建模

- 生成全局语义信息
  - Variational Recurrent Neural Machine Translation. Su et al., AAAI2018.
  - A Stochastic Decoder for Neural Machine Translation. Schulz et al., ACL2018.
  - Deconvolution-Based Global Decoding for Neural Machine Translation. Li et al., COLING2018.
- 其它
  - Modeling Past and Future for Neural Machine Translation. Zheng et al., TACL2018.

未来信息建模



$$\mathcal{L}_{\text{VRNMT}}(\mathbf{y}|\mathbf{x};\theta,\phi) = \sum_{j=1}^{T_y} \mathcal{L}_{\text{VRNMT}}(y_j|\mathbf{x},y_{<j};\theta,\phi)$$

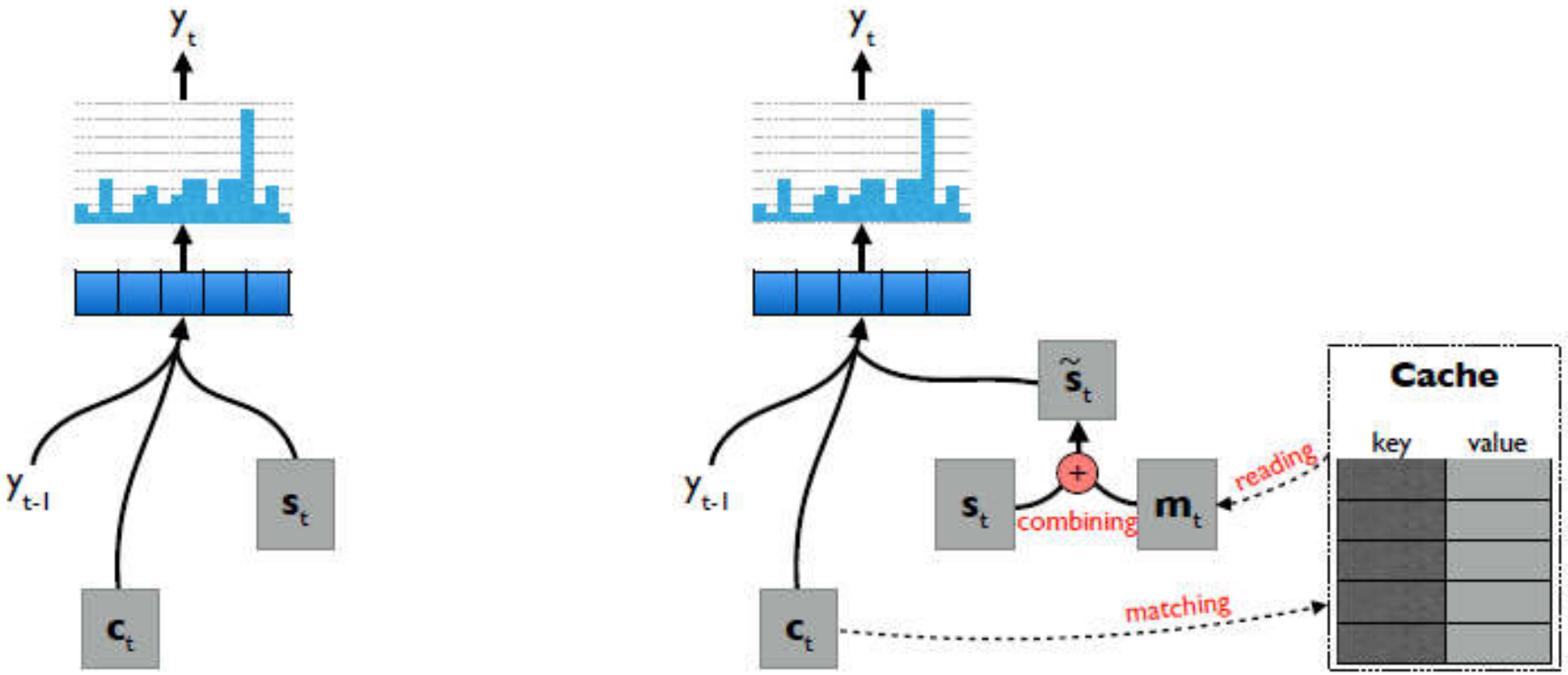
$$= \sum_{j=1}^{T_y}\{ -KL(q_\phi(z_j|\mathbf{x},y_{\le j})\,|\,p_\theta(z_j|\mathbf{x},y_{<j})) + \mathbb{E}_{q_\phi(z_j|\mathbf{x},y_{\le j})}[\log p_\theta(y_j|\mathbf{x},y_{<j},z_j)]\},$$

Variational Recurrent Neural Machine Translation. Jinsong Su, Shan Wu, Deyi Xiong, Yaojie Lu, Xianpei Han, Biao Zhang, *AAAI2018*.

跨句子信息建模

- Document Context Neural Machine Translation with Memory Networks. Maruf and Haffari. ACL2018.
- Learning to Remember Translation History with a Continuous Cache. Tu et al., TACL2018.
- Modeling Coherence for Neural Machine Translation with Dynamic and Topic Caches. Kuang et al., COLING2018.
- Fusing Recency into Neural Machine Translation with an Inter-Sentence Gate Model. Kuang et al., COLING18.

跨句子信息建模



Learning to Remember Translation History with a Continuous Cache. Zhaopeng Tu, Yang Liu, Shuming Shi and Tong Zhang, *TACL2018*.

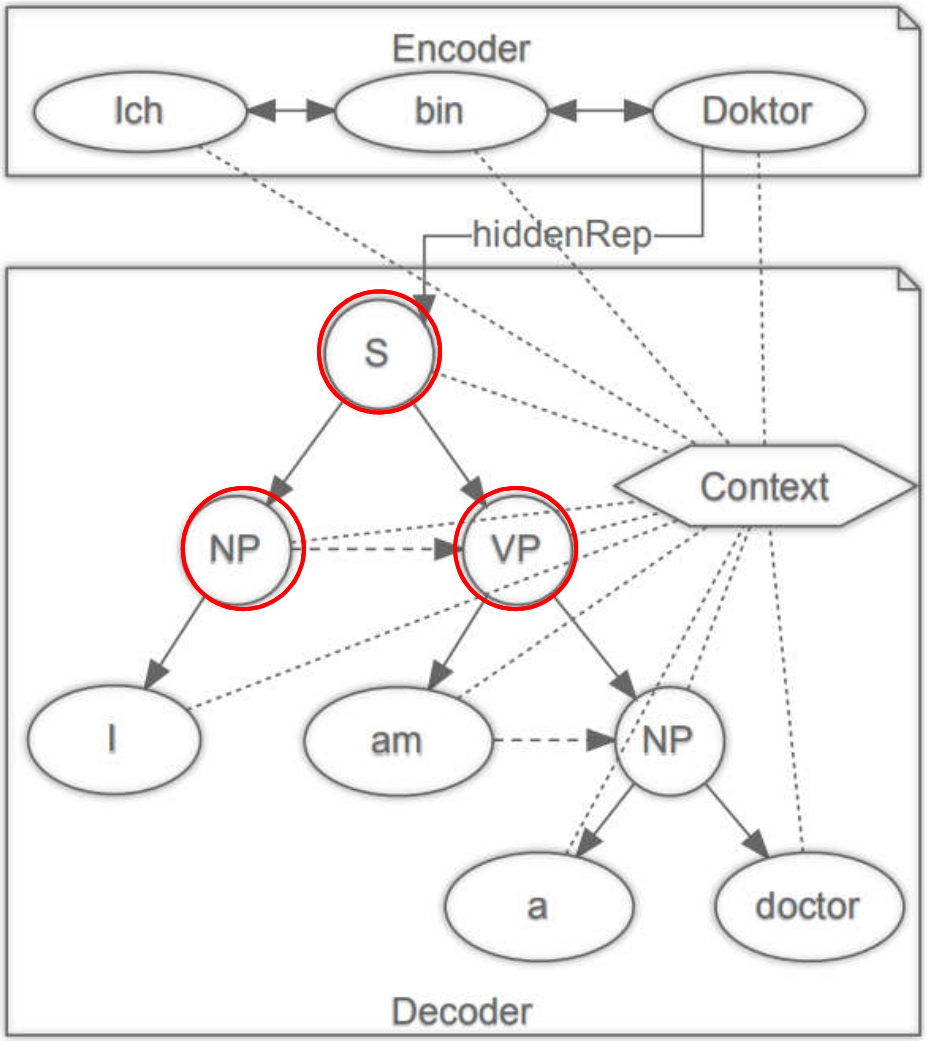
SMT知识

- Incorporating Statistical Machine Translation Word Knowledge Into Neural Machine Translation. Wang et al., TASLP2018.
- Phrase Table as Recommendation Memory for Neural Machine Translation. Zhao et al., IJCAI2018.
- Addressing Troublesome Words in Neural Machine Translation. Yang et al., EMNLP2018.

句法知识

- 源端（句法森林）
  - Forest-Based Neural Machine Translation. Ma et al., ACL2018.
  - Incorporating Syntactic Uncertainty in Neural Machine Translation with a Forest-to-Sequence Model. Zaremoodi and Haffari. COLING2018.
- 目标端（句法树）
  - Top-down Tree Structured Decoding with Syntactic Connections for Neural Machine Translation and Parsing. Gū et al., EMNLP2018.
- 源端+目标端（依存树）
  - Dependency-to-Dependency Neural Machine Translation. Wu et al., TALSP2018.

句法知识



Top-down Tree Structured Decoding with Syntactic Connections for Neural Machine Translation and Parsing. Jetic Gū, Hassan S. Shavarani, Anoop Sarkar, *EMNLP2018.*

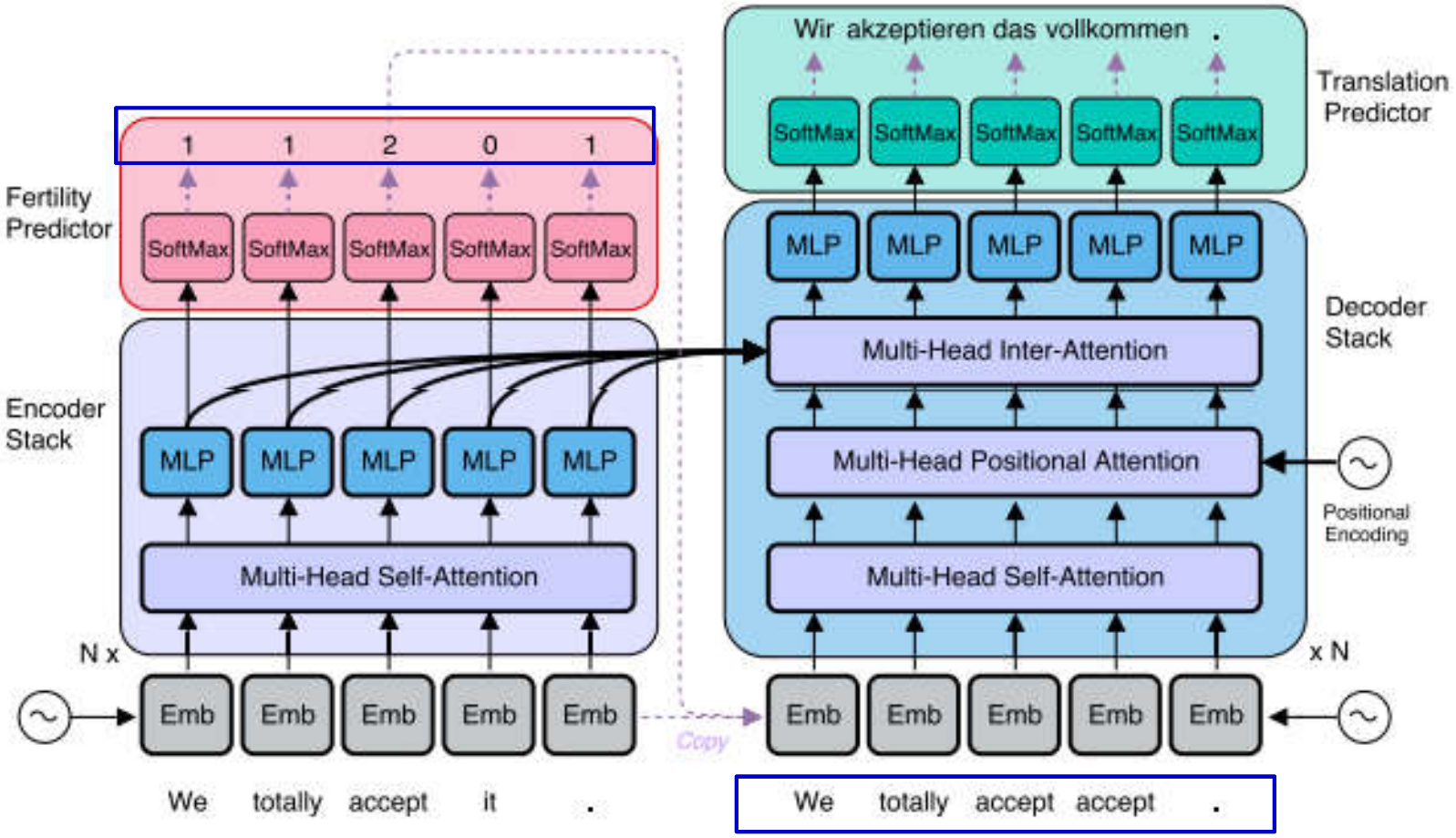
其它外部知识

- Linguistic Knowledge-Aware Neural Machine Translation. Li et al., TALSP2018.
- Search Engine Guided Neural Machine Translation. Gu et al., AAAI2018.
- Guiding Neural Machine Translation with Retrieved Translation Pieces. Zhang et al., NAACL2018.
- Neural Machine Translation Incorporating Named Entity. Ugawa et al., COLING2018.

Transformer

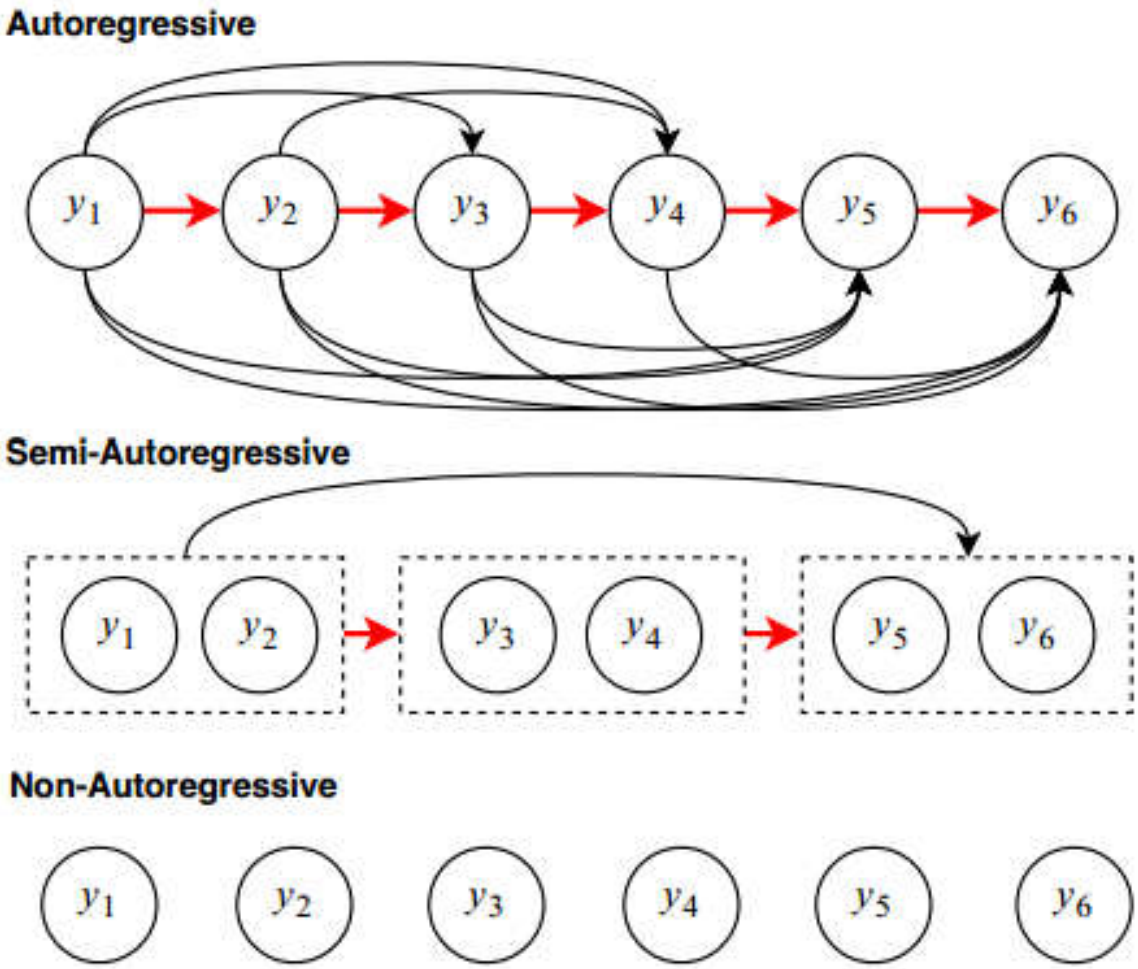
- 译文产生
  - Non-Autoregressive Neural Machine Translation. Gu et al. ICLR2018.
  - Semi-Autoregressive Neural Machine Translation. Wang et al. EMNLP2018.
- 加入上下文
  - Context-Aware Neural Machine Translation Learns Anaphora Resolution. Voita et al. ACL2018.
  - Improving the Transformer Translation Model with Document-Level Context. Zhang et al. EMNLP2018.

Transformer



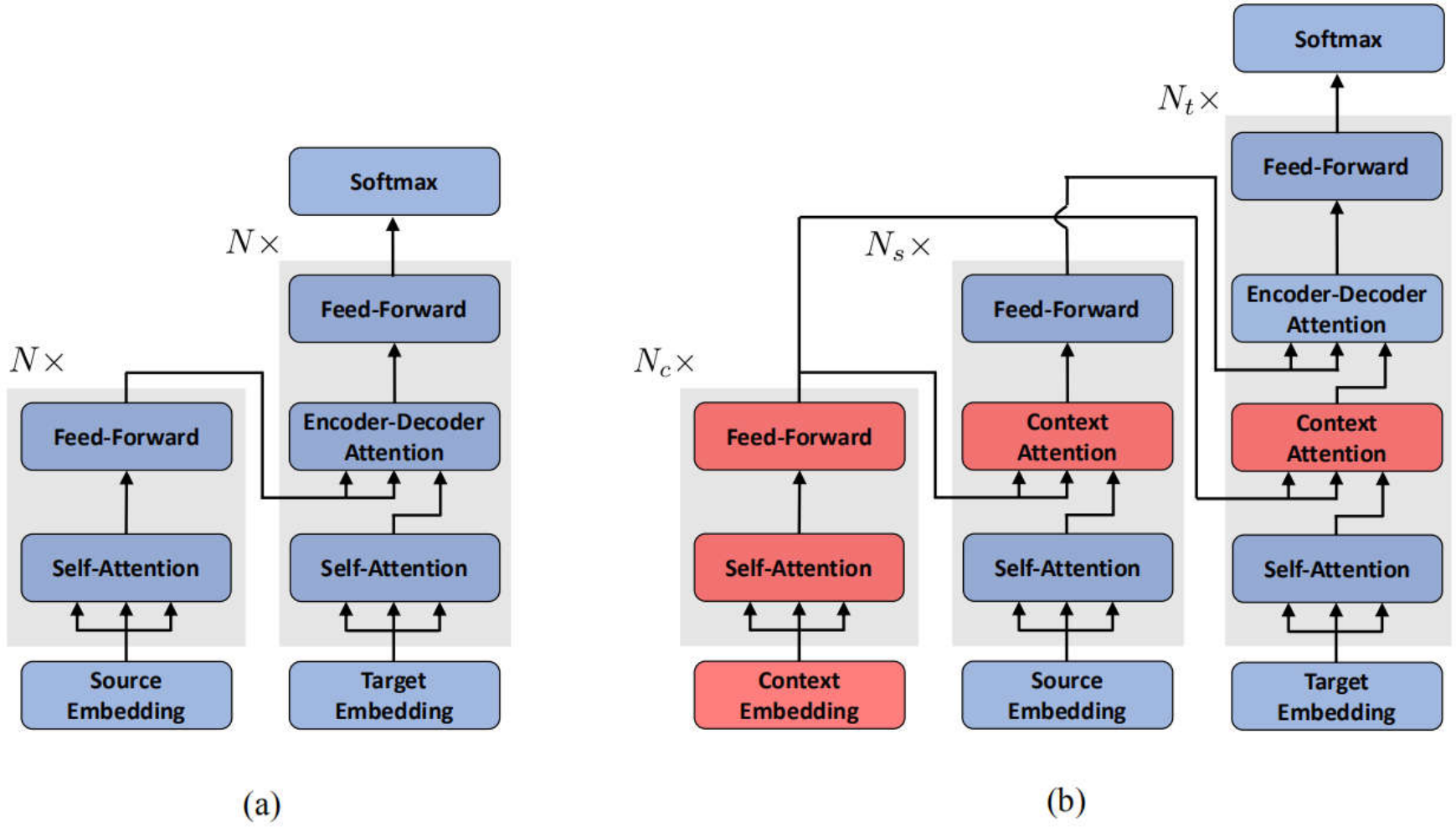
Non-Autoregressive Neural Machine Translation. Jiatao Gu, James Bradbury, Caiming Xiong, Victor O.K. Li, Richard Socher, *ICLR2018*.

Transformer



Semi-Autoregressive Neural Machine Translation. Chunqi Wang, Ji Zhang, and Haiqing Chen, *EMNLP2018.*

Transformer



Improving the Transformer Translation Model with Document-Level Context. Jiacheng Zhang, Huanbo Luan, Maosong Sun, Feifei Zhai, Jingfang Xu, Min Zhang and Yang Liu. *EMNLP2018*.

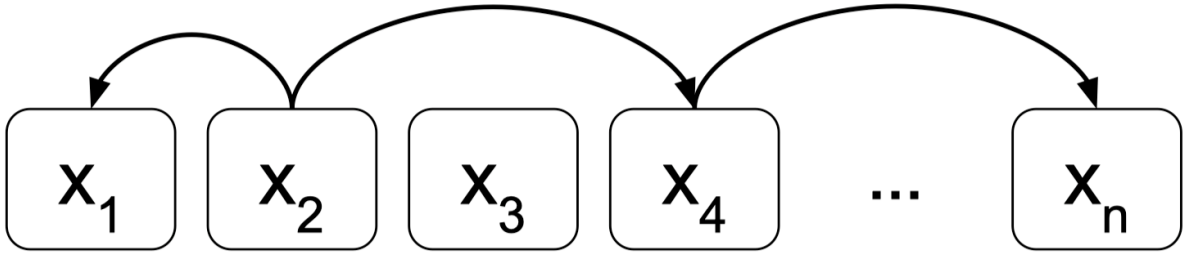
Transformer

- **Attention改进**
  - Self-Attention with Relative Position Representations. Shaw et al., NAACL2018.
  - Bi-directional Block Self-attention for Fast and Memory-efficient Sequence Modeling. Shen et al., ICLR2018.
  - DiSAN: Directional Self-attention Network for RNN/CNN Language Understanding. Shen et al., AAAI2018.
  - Accelerating Neural Transformer via an Average Attention Network. Zhang et al., ACL2018.
  - Modeling Localness for Self-Attention Networks. Yang et al., EMNLP2018.
- **多层融合**
  - Multi-layer Representation Fusion for Neural Machine Translation. Wang et al., COLING2018.
  - Exploiting Deep Representations for Neural Machine Translation. Dou et al., EMNLP2018.

Transformer

$a^V_{2,1}=w^V_{-1}$ $a^V_{2,4}=w^V_2$ $a^V_{4,n}=w^V_k$

$a^K_{2,1}=w^K_{-1}$ $a^K_{2,4}=w^K_2$ $a^K_{4,n}=w^K_k$

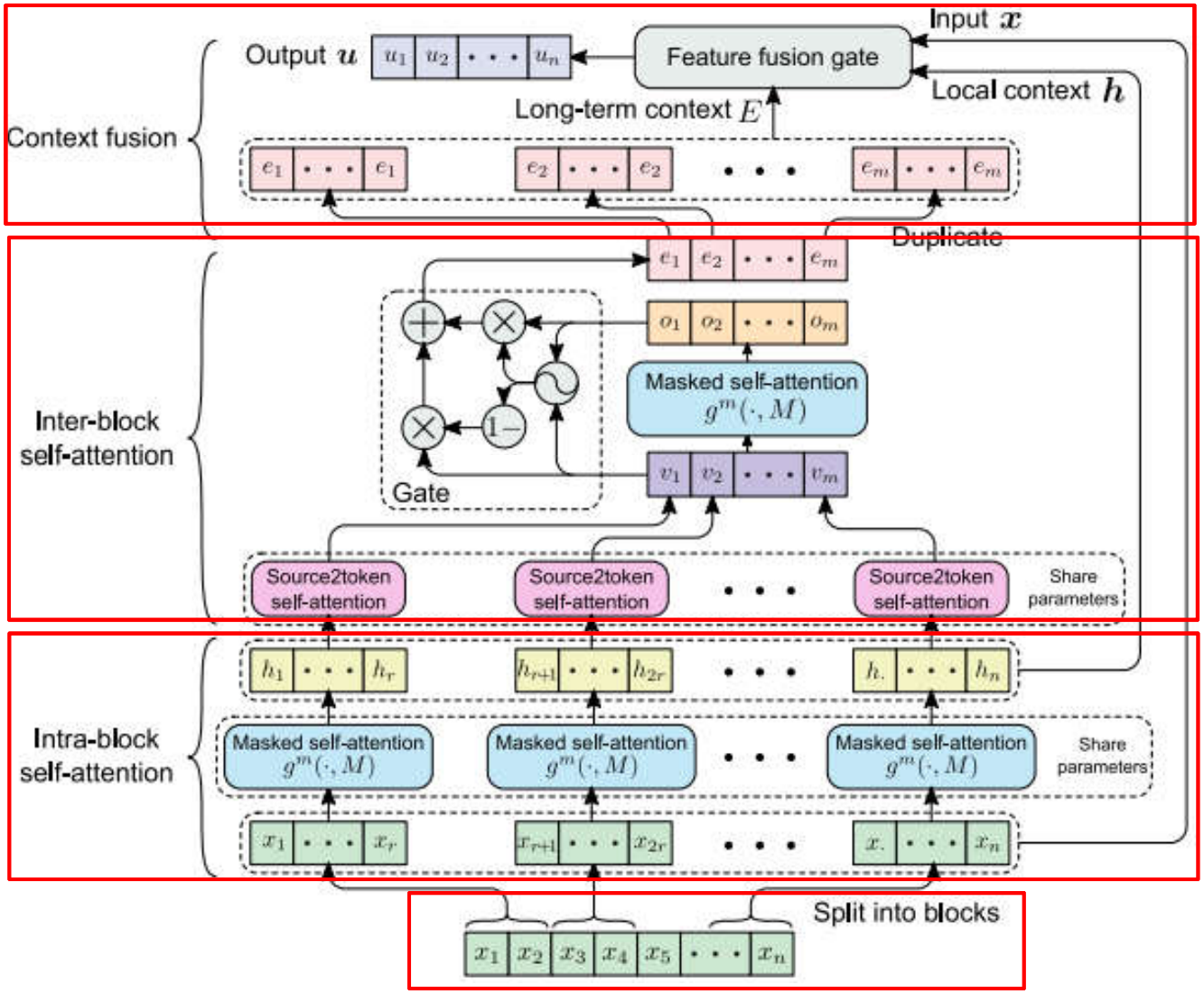


$$z_i = \sum_{j=1}^{n} \alpha_{ij}(x_j W^V) \longrightarrow z_i = \sum_{j=1}^{n} \alpha_{ij}(x_j W^V + a_{ij}^V)$$

$$e_{ij} = \frac{(x_i W^Q)(x_j W^K)^T}{\sqrt{d_z}} \longrightarrow e_{ij} = \frac{x_i W^Q (x_j W^K + a_{ij}^K)^T}{\sqrt{d_z}}$$

Self-Attention with Relative Position Representations. Peter Shaw, Jakob Uszkoreit and Ashish Vaswani, *NAACL2018.*

Transformer



Bi-directional Block Self-attention for Fast and Memory-efficient Sequence Modeling. Tao Shen, Tianyi Zhou, Guodong Long, Jing Jiang, and Chengqi Zhang., *ICLR2018*.

ConvS2S

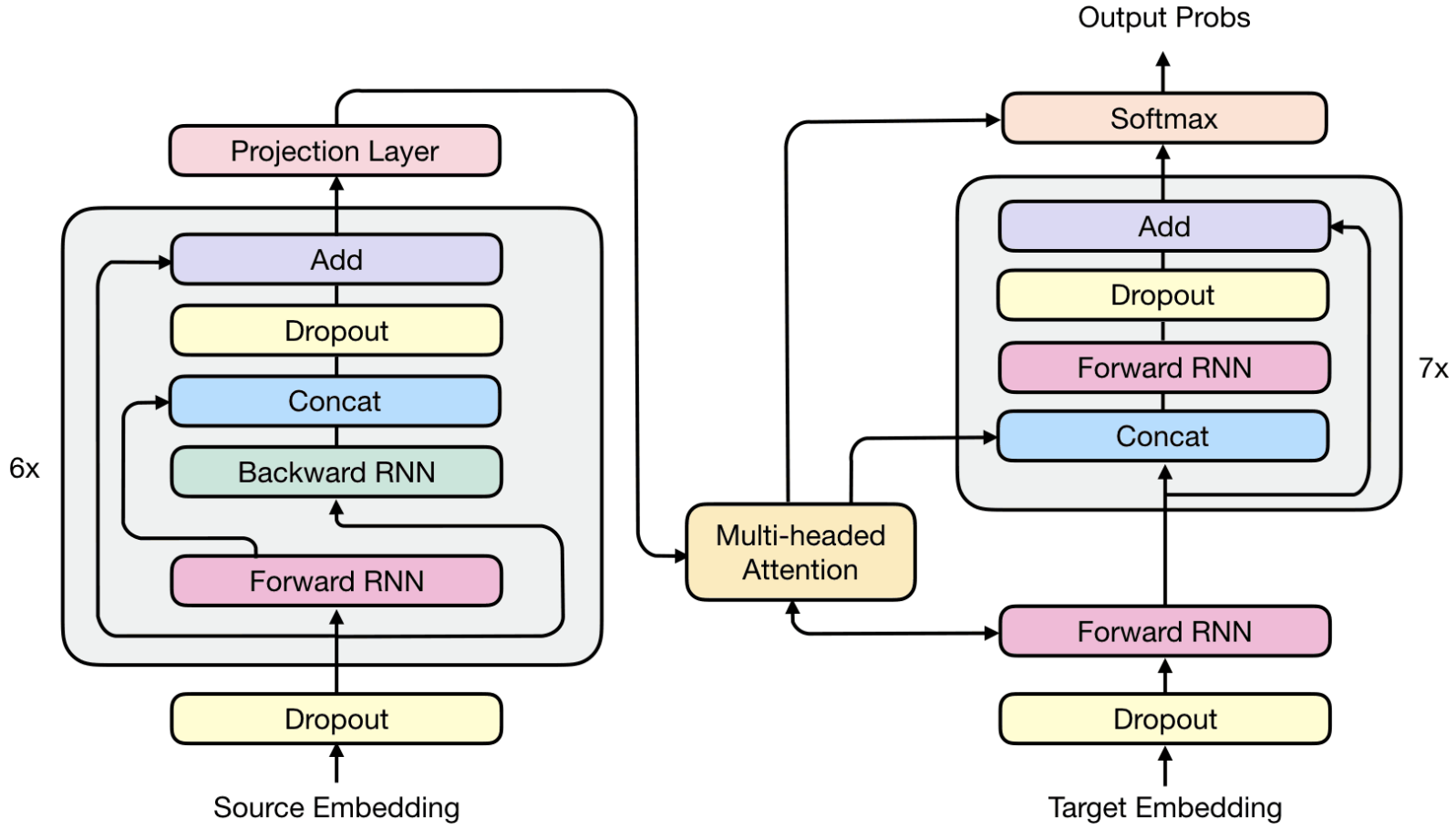
- A Convolutional Encoder Model for Neural Machine Translation. Gehring et al., ACL2017.
- Convolutional Sequence to Sequence Learning. Gehring et al., ICML2017.
- Depthwise Separable Convolutions for Neural Machine Translation. Kaiser et al., ICLR2018.

模型对比

- The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation. Chen et al., ACL2018.
- How Much Attention Do You Need? A Granular Analysis of Neural Machine Translation Architecture. Tobias Domhan, ACL2018.
- Neural Machine Translation with Decoding History Enhanced Attention. Wang et al., COLING2018.
- A Comparison of Transformer and Recurrent Neural Networks on Multilingual Neural Machine Translation. Lakew et al., COLING2018.

模型对比

- **RNMT+**



The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation. Chen et al., *ACL2018*.

模型对比

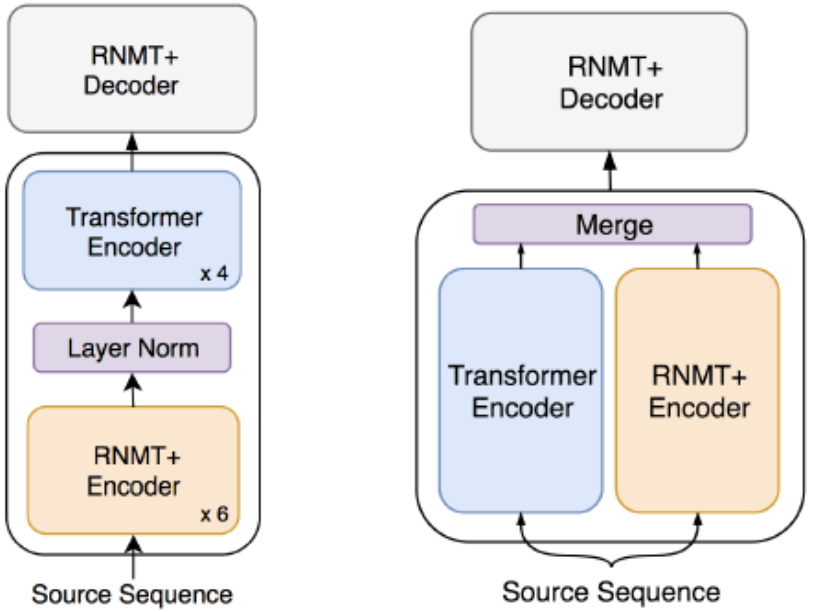
结论

- Label Smoothing对RNMT和Transformer都有效
- Multi-head Attention对RNMT和Transformer都有效
- Layer Normalization使得RNMT和Transformer的模型训练更加稳定
- 增大Batch Size对RNMT和Transformer都有效

The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation. Chen et al., *ACL2018*.

模型对比

- 实验



(a) Cascaded Encoder (b) Multi-Column Encoder

Model	En→Fr BLEU	En→De BLEU
Trans. Big	40.73 ± 0.19	27.94 ± 0.18
RNMT+	41.00 ± 0.05	28.49 ± 0.05
Cascaded	**41.67 ± 0.11**	28.62 ± 0.06
MultiCol	41.66 ± 0.11	**28.84 ± 0.06**

The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation. Chen et al., *ACL2018*.

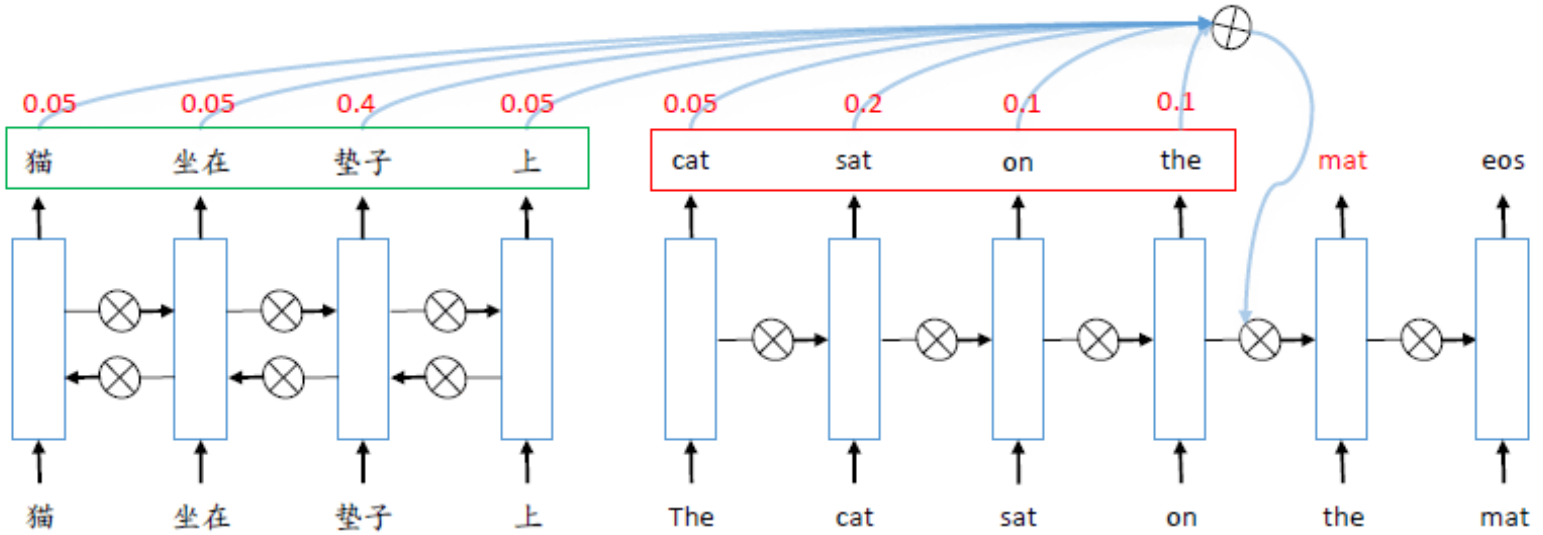
模型对比

结论

- RNMT，ConvS2S可以改到和Transformer性能相近
- Encoder最后一层语义表示最重要
- Multi-head Attention和Residual Feed-Forward层很重要
- 源端的Self-Attention比目标端的Self-Attention更重要

How Much Attention Do You Need? A Granular Analysis of Neural Machine Translation Architecture. Tobias Domhan., *ACL2018*.

模型对比



SYSTEM	MT03	MT04	MT05	MT06	AVE.
Existing systems					
Moses	31.61	33.48	30.75	30.85	31.67
DEEPLAU(Wang et al., 2017)	39.35	41.15	38.07	37.29	38.97
FRNN+PRNN (Zheng et al., 2017)	37.90	40.37	36.75	34.55	37.39
COVERAGE+Context Gate(Tu et al., 2016a)	-	-	34.13	34.83	-
Transformer (Vaswani et al., 2017)	45.16	46.81	44.62	43.53	45.03
Our deep NMT systems					
Source Attn.	45.08	46.90	44.32	43.13	44.85
DHEA + Hybird Comb.	46.58	47.20	45.45	43.47	45.66
DHEA + Sum Comb.	46.38	47.15	45.30	43.60	45.50
DHEA + Gate Comb.	46.60	47.73	45.35	43.97	**45.90**

SYSTEM	Architecture	EN-Fr BLEU	EN-DE BLEU
Existing systems			
Buck et al. (2014)	Winning WMT14 system	35.7	20.7
Wu et al. (2016)	GNMT + Ensemble	40.4	26.3
Gehring et al. (2017)	ConvS2S	40.51	25.16
Vaswani et al. (2017)	Transformer (small)	38.1	27.3
Vaswani et al. (2017)	Transformer (large)	41.0	28.4
Our deep NMT systems			
this work	Source Attn (small)	39.3	27.5
this work	DHEA + Gate Comb. (small)	40.4	27.9
this work	DHEA + Gate Comb. (large)	**41.6**	**28.7**

Neural Machine Translation with Decoding History Enhanced Attention. Mingxuan Wang, Jun Xie, Zhixing Tan, Jinsong Su, Deyi Xiong, Chao Bian, *COLING2018.*

模型分析

- Analyzing Uncertainty in Neural Machine Translation. Ott et al., ICML2018.
- Synthetic and Natural Noise Both Break Neural Machine Translation. Belinkov et al., ICLR2018.
- Beyond Error Propagation in Neural Machine Translation: Characteristics of Language Also Matter. Wu et al., EMNLP2018.
- Why Self-Attention? A Targeted Evaluation of Neural Machine Translation Architectures. Tang et al., EMNLP2018.

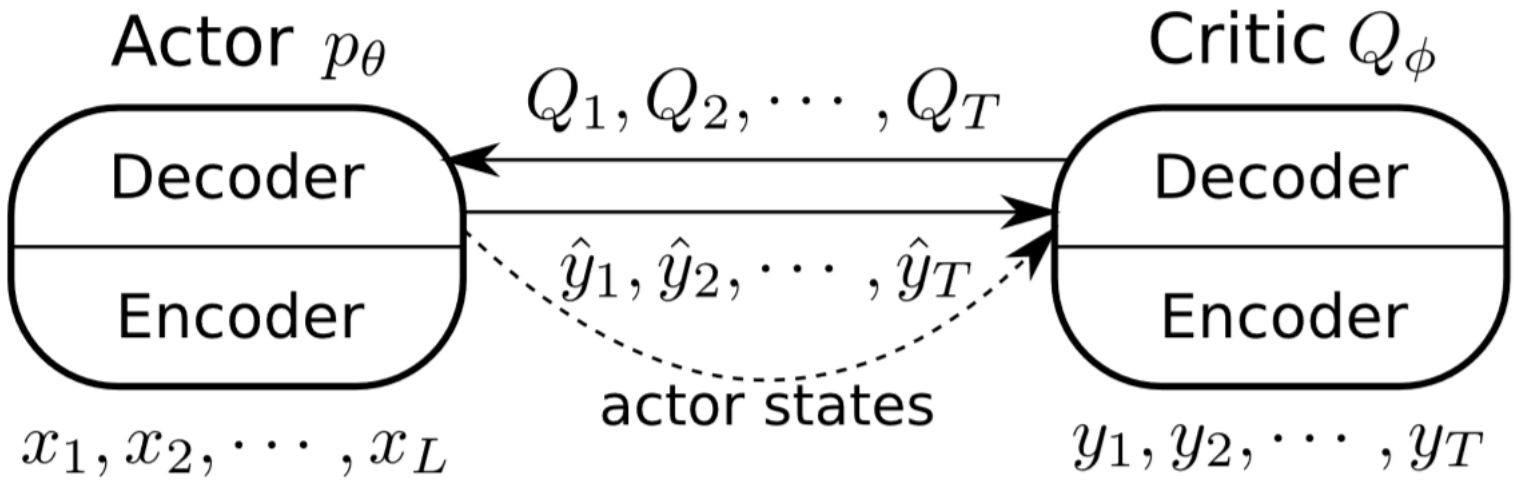
模型训练

- 强化学习
- 对抗

强化学习

- An Actor-critic Algorithm for Sequence Prediction. Bahdanau et al., ICLR2017.
- Decoding with Value Networks for Neural Machine Translation. He et al., NIPS2017.
- A Study of Reinforcement Learning for Neural Machine Translation. Wu et al., EMNLP2018.

强化学习

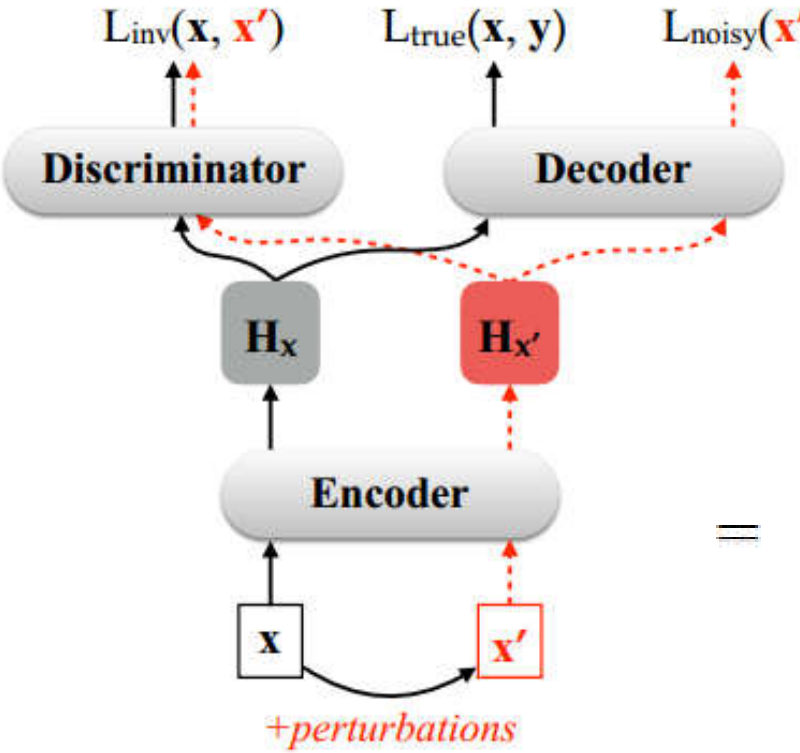


An Actor-critic Algorithm for Sequence Prediction. Dzmitry Bahdanau, Philemon Brakel, Kelvin Xu, Anirudh Goyal, Ryan Lowe, Joelle Pineau, Aaron Courville, Yoshua Bengio, *ICLR2017*.

对抗

- Towards Robust Neural Machine Translation. Cheng et al., ACL2018.
- Neural Machine Translation with Gumbel-Greedy Decoding. Gu et al., AAAI2018.
- Improving Neural Machine Translation with Conditional Sequence Generative Adversarial Nets. Yang et al., NAACL2018.
- On Adversarial Examples for Character-Level Neural Machine Translation. Ebrahimi et al., COLING2018.

对抗



$$
\begin{aligned}
\mathcal{J}(\boldsymbol{\theta}) = & \sum_{\langle \mathbf{x},\mathbf{y} \rangle \in \mathcal{S}} \Big(\mathcal{L}_{\text{true}}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}_{\text{enc}}, \boldsymbol{\theta}_{\text{dec}}) \\
& + \alpha \sum_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} \mathcal{L}_{\text{inv}}(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta}_{\text{enc}}, \boldsymbol{\theta}_{\text{dis}}) \\
& + \beta \sum_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} \mathcal{L}_{\text{noisy}}(\mathbf{x}', \mathbf{y}; \boldsymbol{\theta}_{\text{enc}}, \boldsymbol{\theta}_{\text{dec}}) \Big)
\end{aligned}
$$

Towards Robust Neural Machine Translation. Yong Cheng, Zhaopeng Tu, Fandong Meng, Junjie Zhai, Yang Liu, *ACL2018*.

未来值得关注的问题

- 资源受限的NMT
- 知识驱动的NMT
- RNMT+/Transformer如何做简化，降低训练复杂度
- RNMT+/Transformer语言学方面的可解释性
- 新翻译模型架构设计

谢谢！