



Deconvolution-Based Global Decoding for Neural Machine Translation

Junyang Lin^{1,2}

Xu Sun²

Xuancheng Ren²

Shuming Ma²

Qi Su¹

School of Foreign Languages, Peking University
MOE Key Laboratory of Computational Linguistics, Peking University

{linjunyang, xusun, renxc, shumingma, sukia}@pku.edu.cn

Abstract

- A new NMT model with deconvolution-based decoder;
- Generate translation based on the structural prediction of the target-side context;
- Improved results on benchmark datasets;
- Translation is robust to inputs of diverse lengths and with less repetition.

Linguistic Insight

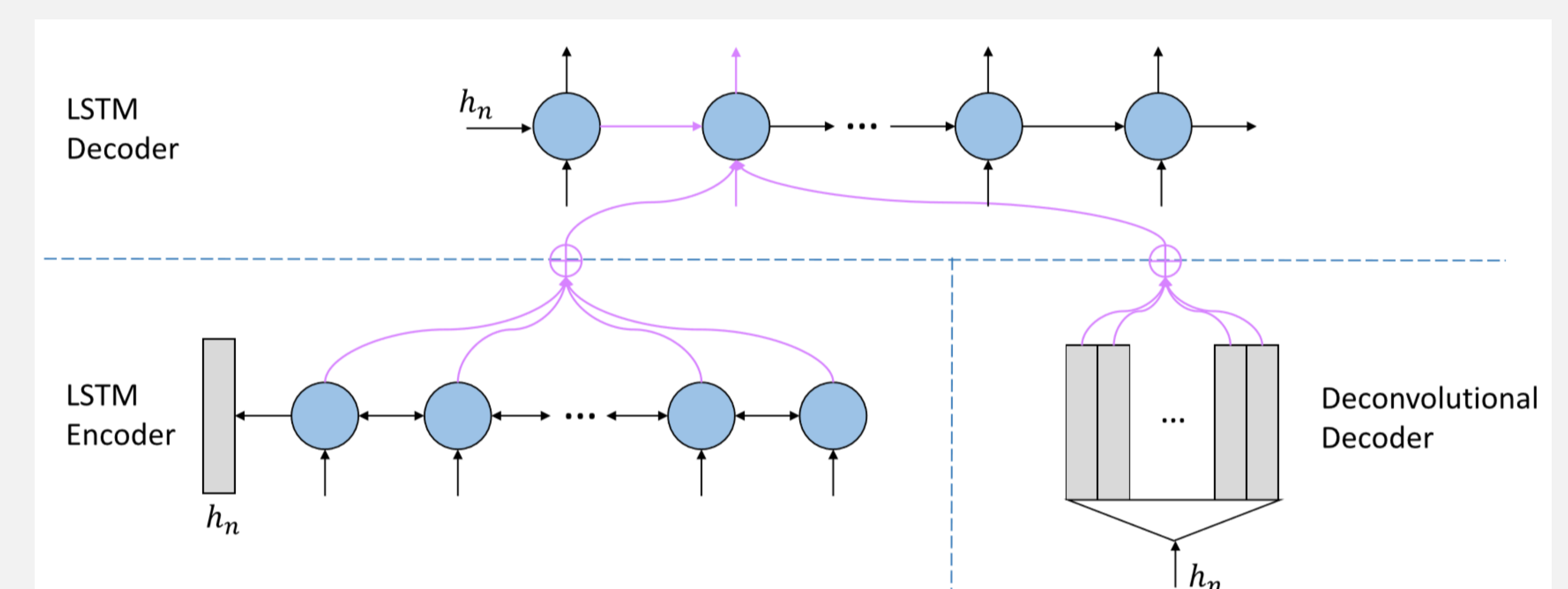
- Language generation involves complex syntactic analysis and semantic integration;
- Translation can be guided by the global lexical-semantic and discourse information;
- The process of translation is in need of the global information from the target-side context.

Sequence-to-Sequence as Baseline

- Encoder: Bidirectional LSTM
- Decoder: LSTM for sequential decoding. Training is with teacher forcing.
- Attention mechanism: global attention for the relevant source-side information

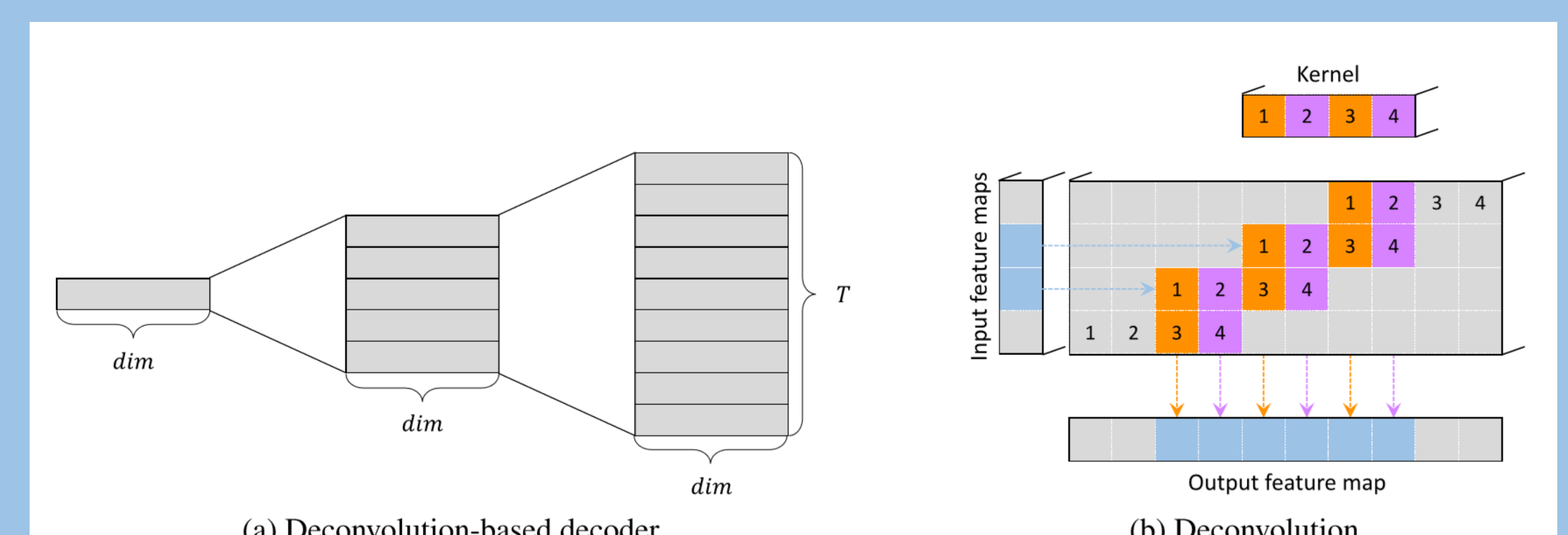
Model Architecture

- Two decoders: the deconvolution-based decoder for global target-side context and the RNN-based decoder for final translation.



Deconvolution-Based Decoder

- Based on deconvolution, able to expand sentence embedding to large matrix, trained to resemble the target word embedding matrix.



Mechanism

- Dual attention for the source-side and target-side contexts;
- Training with multiple losses: to minimize the cross-entropy loss of the generation and the target sequence as well as the smooth L1 loss of the generated matrix by deconvolution-based decoder and the target word embedding matrix (truncation and padding are required for the fixed size).

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \log P(y_t^{(i)} | \tilde{y}_{<t}^{(i)}, x^{(i)}, \theta) + \sum_{m=1}^M \text{smooth}_{L1}(E_m - \tilde{E}_m) + \sum_{t=1}^T \log P(y_t^{(i)} | x^{(i)}, \theta')$$

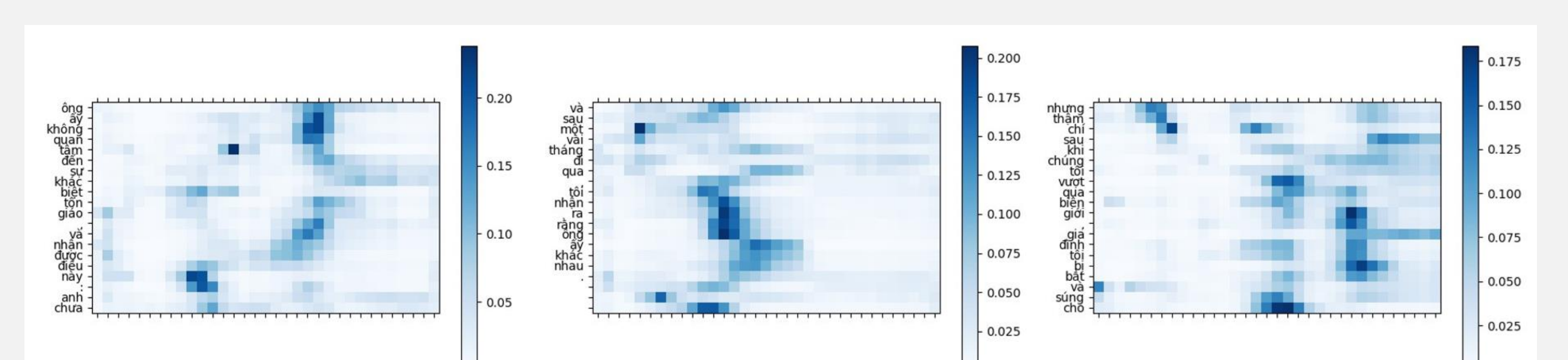
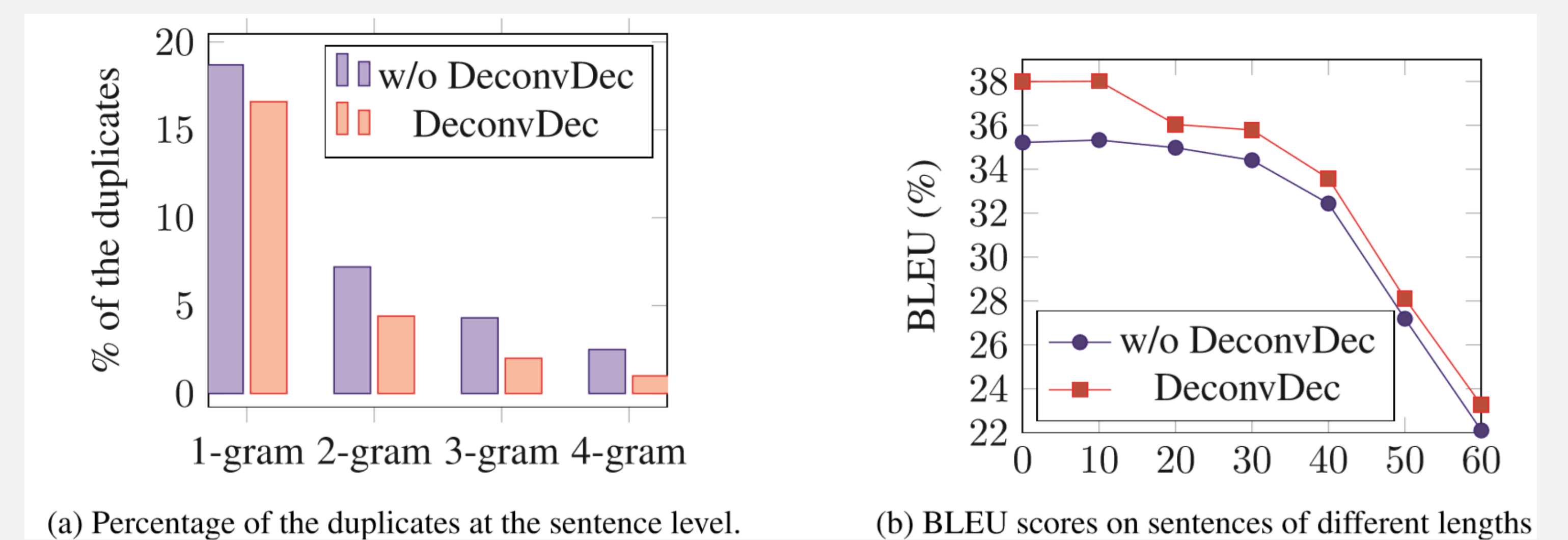
Experiments

- Chinese-English translation (LDC corpora);
- English-Vietnamese translation (IWSLT 2015)

Model	MT-03	MT-04	MT-05	MT-06	Ave.
Moses	32.43	34.14	31.47	30.81	32.21
RNNSearch	33.08	35.32	31.42	31.61	32.86
Lattice	34.32	36.50	32.40	32.77	34.00
Coverage	34.49	38.34	34.91	34.25	35.49
InterAtten	35.09	37.73	35.53	34.32	35.67
MemDec	36.16	39.81	35.91	35.98	36.97
Seq2Seq+Attention	35.32	37.25	33.52	33.54	34.91
+DeconvDec	38.04	39.75	36.77	36.32	37.73

Model	BLEU
RNNSearch-1	23.30
RNNSearch-2	26.10
LabelEmb	26.80
NPMT	27.69
Seq2Seq+Attention	26.93
+DeconvDec	28.47

Qualitative Analyses



Text: 叛军暗杀两位菲国国会议员后, 菲律宾总统雅罗育在二零零一年中期停止与共产党谈判。

Gold: after the rebels assassinated two philippine legislators , philippine president arroyo ceased negotiations with the communist party in mid 2001 .

Seq2Seq: philippine president gloria arroyo stopped the two philippine parliament members in the mid - autumn festival .

DeconvDec: philippine president gloria arroyo stopped holding talks with the communist party after the rebels assassinated two philippine parliament members .

Conclusion

- A new model with the global decoding mechanism;
- Coherent and accurate translation with fewer irrelevant contents;
- Will figure out its generalized patterns for the construction of the target-side contexts