An Auto-Encoder Matching Model for Learning Utterance-Level Semantic Dependency in Dialogue Generation

Liangchen Luo*, Jingjing Xu*, Junyang Lin, Qi Zeng, Xu Sun

MOE Key Lab of Computational Linguistics, School of EECS, Peking University (* indicates equal contribution)

{luolc,jingjingxu,linjunyang.pkuzengqi,xusun}@pku.edu.cn

INTRODUCTION

Challenge

Generating semantically coherent responses is still a major challenge in *dialogue generation*. Different from conventional text generation tasks, the mapping between inputs and responses in conversations is more complicated, which highly demands the understanding of utterance-level semantic dependency.

Method

We propose an AUTO-ENCODER MATCHING (AEM) model to learn such dependency. The model contains two auto-encoders and one mapping module. The auto-encoders learn the semantic representations of inputs and responses, and the mapping module learns to connect the utterance-level representations.

CONTRIBUTIONS

- To promote coherence in dialogue generation, we propose a novel AUTO-ENCODER MATCHING model to learn the utterance-level dependency.
- In our proposed model, we explicitly separate utterance representation learning and dependency learning for a better expressive ability.
- Experimental results on automatic evaluation and human evaluation show that our model can generate much more coherent text compared to baseline models.

MODEL

Encoder

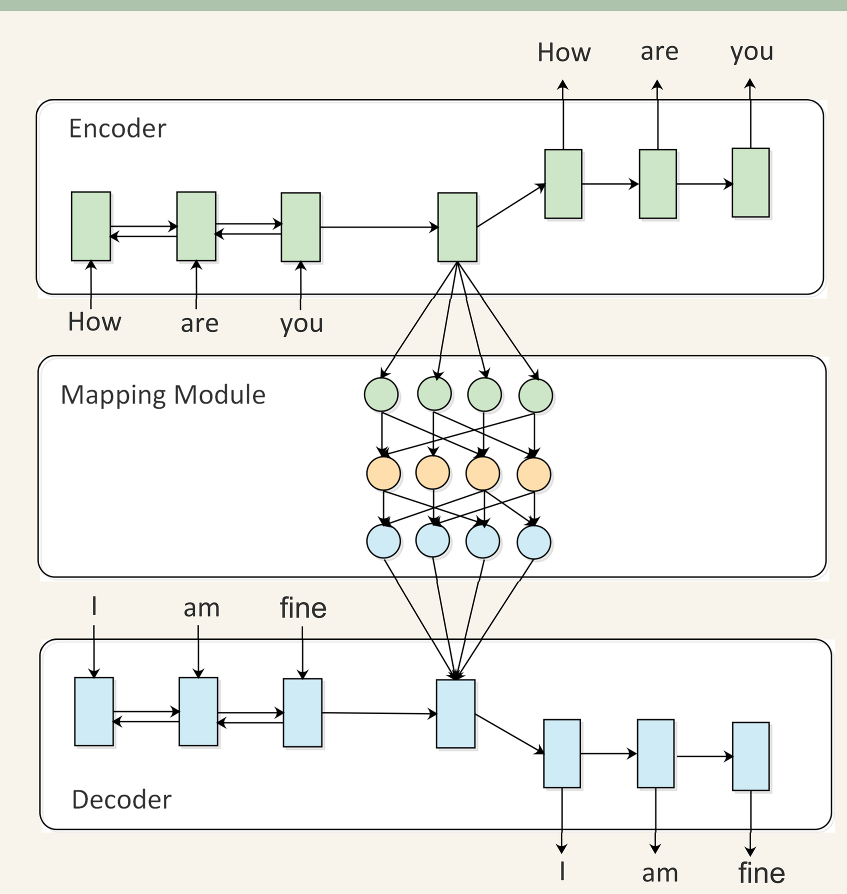
The encoder is an unsupervised auto-encoder based on LSTM. In training, the encoder receives the source text (dialog input), encodes it to an internal representation h, and then decodes h to a new sequence for the reconstruction of the input. We extract the hidden state h as the semantic representation.

Decoder

Similar to the encoder, the decoder is also a LSTM-based auto-encoder. We use s to indicate the utterance-level semantic representation.

Mapping Module

A simple feedforward network is used to transform the source semantic representation h to a new representation t. The mapping module is trained by minimizing the L2 loss between t and s.



EXPERIMENT

We conduct experiments on DailyDialog dataset (Li et al., 2017).

BLEU Scores

Models	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Seq2Seq	12.43	4.57	2.69	1.84
AEM	13.55	4.89	3.04	2.16
Seq2Seq+Attention	13.63	4.99	3.05	2.13
AEM+Attention	**14.17**	**5.69**	**3.78**	**2.84**

Table 1: BLEU scores for the AEM and Seq2Seq model.

Diversity of Generated Text

Considerable improvement of text diversity by AEM, reflected by the number of distinct 1-grams, 2-grams and 3-grams.

Models	Dist-1	Dist-2	Dist-3
Seq2Seq	0.8K	2.7K	5.5K
AEM	3.1K	14.8K	31.2K
Seq2Seq+Attention	2.5K	13.6K	34.6K
AEM+Attention	**3.3K**	**23.2K**	**53.9K**

Table 2: Diversity scores for the AEM and Seq2Seq model.

Human Evaluation

Models	Fluency	Coherence	G-Score
Seq2Seq	6.97	3.51	4.95
AEM	**8.11**	4.18	5.82
Seq2Seq+Attention	5.11	3.30	4.10
AEM+Attention	7.92	**4.97**	**6.27**

Table 3: Human evaluation results. Fluency represents whether each sentence is in correct grammar. Coherence evaluates whether the response is relevant to the input. The score ranges from 1 to 10 (1⇒bad, 10⇒good).

Examples

It is easy to see that the AEM model obviously outperforms the Seq2Seq model. It shows that the utterance-level dependency learnt by the AEM model is useful for generating responses with more diversity.

Input 1:	*My name is David. What is my name?*
Seq2Seq:	*Ours, maria.*
AEM:	*Pardon me? It is unreasonable!*
Input 2:	*I'm going to kill you!*
Seq2Seq:	*I know. I'm a good man.*
AEM:	*What?!?!!!! No!!*
Input 3:	*I hate you.*
Seq2Seq:	*I'm not.*
AEM:	*You do? You're a sweet girl.*
Input 4:	*How are you?*
Seq2Seq:	*Fine.*
AEM:	*Fine. I'm good.*
Input 5:	*Can you say 'Hello'?*
Seq2Seq:	*Yeah, I guess.*
AEM:	*Sure, I guess.*

Table 4: Examples generated by the AEM and Seq2Seq model.

CONCLUSION

- We propose an AUTO-ENCODER MATCHING model to learn the utterance-level semantic dependency, a critical dependency relation for generating coherent and fluent responses.
- The model contains two auto-encoders that learn the utterance representations in an unsupervised way, and a mapping module that builds the mapping between the input representation and response representation.
- Experimental results show that the proposed model significantly improves the quality of generated responses according to automatic evaluation and human evaluation, especially in coherence.

MISC

References

1. Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. In *Proc. of IJCNLP*.

Links

arXiv: arxiv.org/abs/1808.08795

code: github.com/lancopku/AMM

※ *Scan QR code to view the paper*→