<<Neural Responding Machine for Short-Text Conversation>>

Tianyi Luo
2015-03-18

Author & Publication

- Lifeng Shang, Zhengdong Lu and Hang Li
  Noah's Ark Lab, Huawei Technologies Co. Ltd.
  Sha Tin, Hong Kong
- arXiv:1503.023640v2 [cs.CL] 27 April 2015
- ACL 2015 long paper

Abstract

We propose Neural Responding Machine (NRM), a neural network-based response generator for Short-Text Conversation. NRM takes the general encoder-decoder framework: it formalizes the generation of response as a decoding process based on the latent representation of the input text, while both encoding and decoding are realized with recurrent neural networks (RNN). The NRM is trained with a large amount of one-round conversation data collected from a microblogging service. Empirical study shows that NRM can generate grammatically correct and content-wise appropriate responses to over 75% of the input text, outperforming state-of-the-arts in the same setting, including retrieval-based and SMT-based models.

Short-Text Conversation

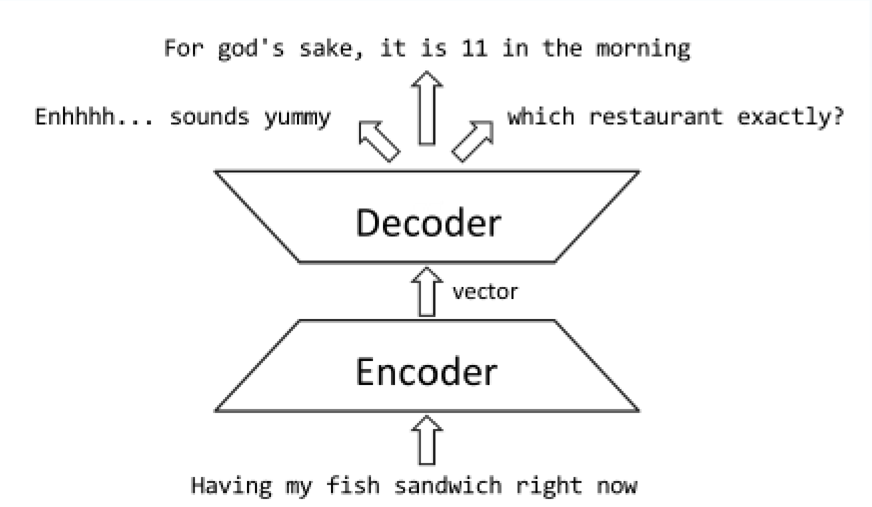
- What **problem** does this paper want to tackle?
  - Short-Text Conversation (STC).
  - Only considers one round of conversation.
  - Each round is formed by two short texts, with the **former being an input** (referred to as **post**) from a user and the **latter a response given by the computer**.
  - Utilize sina weibo dataset.

Short-Text Conversation

- What **contributions** does this paper make?
  - We propose to use an encoder-decoder-based neural network to generate a response in STC.
  - We have empirically verified that the proposed method, when trained with a reasonable amount of data, can yield performance better than traditional retrieval-based and translation-based methods.

Short-Text Conversation

- What **method** does this paper utilize?
  - Traditional methods: Retrieval-based and SMT-based models.
  - Employ a neural encoder-decoder for this task, named Neural Responding Machine (NRM).



Short-Text Conversation

- What **method** does this paper utilize?

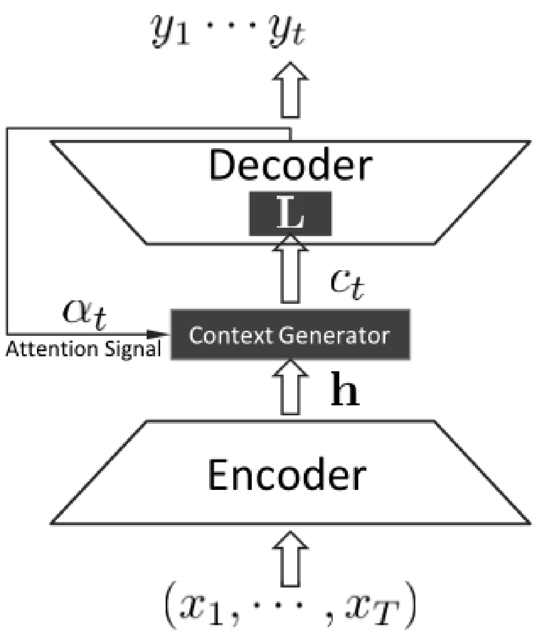


Figure 2: The general framework and dataflow of the encoder-decoder-based NRM.

Short-Text Conversation

- What **method** does this paper utilize?

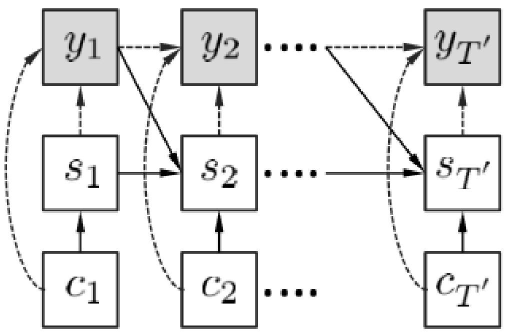


Figure 3: The graphical model of RNN decoder. The dashed lines denote the variables related to the function $g(\cdot)$, and the solid lines denote the variables related to the function $f(\cdot)$.

$$p(y_t|y_{t-1}, \cdots, y_1, \mathbf{x}) = g(y_{t-1}, s_t, c_t),$$

$$s_t = f(y_{t-1}, s_{t-1}, c_t),$$

f(·) can be logistic function, LSTM and **GRU(This paper utilizes GRU)**.

Short-Text Conversation

- What **method** does this paper utilize?

  We consider three types of encoding schemes, namely 1) the global scheme, 2) the local scheme, and the hybrid scheme which combines 1) and 2).

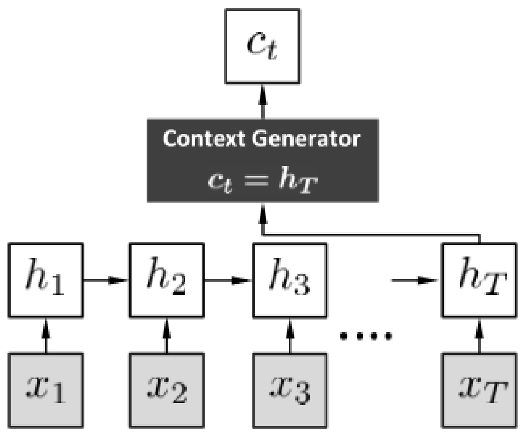


Figure 4: The graphical model of the encoder in NRM-glo, where the last hidden state is used as the context vector $c_t = h_T$.

Short-Text Conversation

- What **method** does this paper utilize?

  We consider three types of encoding schemes, namely 1) the global scheme, 2) the local scheme, and the hybrid scheme which combines 1) and 2).

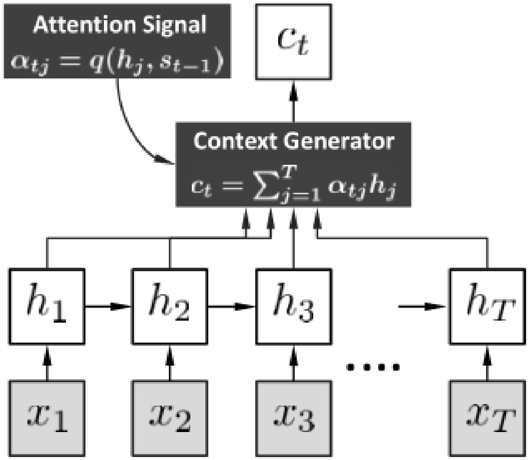


Figure 5: The graphical model of the encoder in NRM-loc, where the weighted sum of hidden sates is used as the context vector $c_t = \sum_{j=1}^{T} \alpha_{tj} h_j$.

Short-Text Conversation

- What **method** does this paper utilize?

We consider three types of encoding schemes, namely 1) the global scheme, 2) the local scheme, and the hybrid scheme

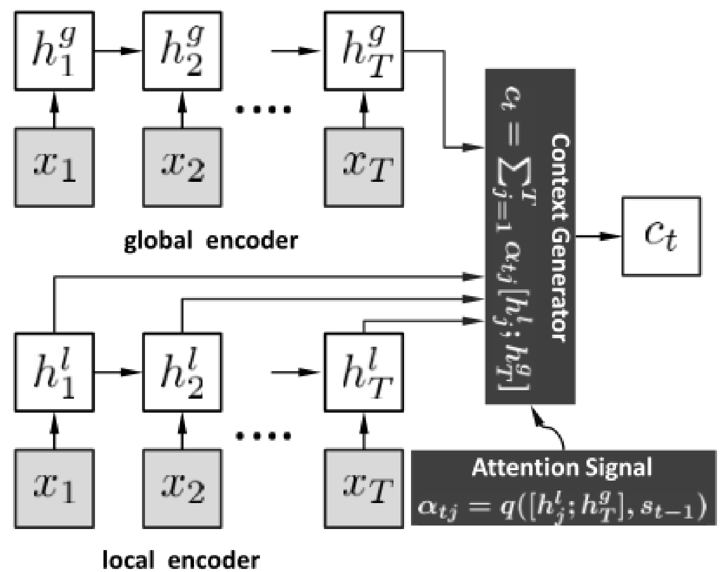


Figure 6: The graphical model for the encoder in NRM-hyb, while context generator function is $c_t = \sum_{j=1}^{T} \alpha_{tj}[h_j^l; h_T^g]$, here $[h_j^l; h_T^g]$ denotes the concatenation of vectors h_j^l and h_T^g

Short-Text Conversation

- What's the **experiments results**?
  - Data:

Training	#posts	219,905
	#responses	4,308,211
	#pairs	4,435,959
Test Data	#test posts	110
Labeled Dataset (retrieval-based)	#posts	225
	#responses	6,017
	#labeled pairs	6,017
Fine Tuning (SMT-based)	#posts	2,925
	#responses	3,000
	#pairs	3,000

Table 1: Some statistics of the dataset. **Labeled Dataset** and **Fine Tuning** are used by retrieval-based method for learning to rank and SMT-based method for fine tuning, respectively.

Short-Text Conversation

- What's the **experiments results**?
  - Labeled data:

Post	意大利禁区里老是八个人...太夸张了 There are always 8 players at the Italian restricted area. Unbelievable!	**Related Criterion**	**Labels**
Response1	我是意大利队的球迷，等待比赛开始。 I am a big fan of the Italy team, waiting for the football game to start	Logic Consistency	Unsuitable
Response2	意大利的食物太美味了 Italian food is absolutely delicious.	Semantic Relevance	Unsuitable
Response3	太夸张了吧！ Unbelievable!	Generality	Neutral
Response4	哈哈哈仍然是0：0 。还没看到进球。 Ha!Ha!Ha! it is still 0:0, no goal so far.	Scenario Dependence	Neutral
Response5	这正是意大利式防守足球。 This is exactly the Italian defending style football game.	——	Suitable

Figure 7: An example post and its five candidate responses with human annotation. The content of the post implies that the football match is already started, while the author of Response1 is still waiting for the match to start. Response2 talks about the food of Italy. Response3 is a widely used response, but it is suitable to this post. Response4 states that the current score is still 0:0, it is a suitable response only in this specific scenario.

Short-Text Conversation

- What's the **experiments results**?
  - The results of evaluated methods:

Models	Mean Score	Suitable (+2)	Neutral (+1)	Unsuitable (0)	Agreement
NRM-glo	0.969	34.0%	28.9%	37.1%	0.397
NRM-loc	1.065	36.0%	34.5%	29.5%	0.247
NRM-hyb	1.156	39.3%	37.1%	23.6%	0.309
Rtr.-based	0.984	29.8%	38.7%	31.5%	0.346
SMT-based	0.313	5.6 %	20.0%	74.4%	0.448

Table 2: The results of evaluated methods. Mean score is the average value of annotated scores over all annotations. (Rtr.-based means the retrieval-based method)

The agreement is evaluated by Fleiss' kappa (Fleiss, 1971), as a statistical measure of inter-rater consistency.

Short-Text Conversation

- What’s the **experiments results**?
  - Some cases of NRM:

ID	Post	NRM-glo	NRM-loc	NRM-hyb	Rtr.-based
1	每年元旦都来高烧。 High fever attacks me every New Year’s Day	多休息，增强免疫力！ Rest more to boost your immune system.	每年元旦都去医院看病 Go to see a doctor every New Year’s Day	祝早日康复，身体健康！ Get well soon and stay healthy!	高烧… High fever…
2	又老了一岁，感谢我们组，我很高兴 I gain one more year. Grateful to my group, so happy	生日快乐！永远支持你！ Happy birthday! Will stand by you forever!	生日快乐！永远年轻漂亮！ Happy birthday! Wish you always stay young and pretty!	老了老了，岁月不饶人啊！ Getting old now. Time has no mercy.	我只是老了一岁 I just gain only one more year.
3	吃素第一天，坚持住，崔朵拉。 First day of being a vegetarian. Hold on, **CuiDuoLa**	必须来支持！加油！ I come to offer my support. Keep it up!	坚持住，坚持下去。 Hold on, keep it up.	每天都是素食主义者 I have been being a vegetarian everyday	坚持住！ Hold on!
4	这种天气真应该带个帅锅出去踏青赏花 We should go out with some cute guys to enjoy a great outing in such a nice weather.	这是哪儿啊？真美！ Where is it? It is so beautiful!	阳光明媚，心情舒畅。 Such a nice sunny day! I am in a great mood.	这种天气真不是盖的。 It is indeed a very nice weather.	文山啊 出去踏青寻找灵感 哈哈 **WenShan**, let’s go out to get some inspiration. Ha! Ha!

Figure 8: Some responses generated by different models (originally in Chinese with their literal English translation), where the words in boldfaces are entity names.

Short-Text Conversation

- What's the **experiments results**?
  - Friedman test of different models:

Model A	Model B	Average rankings	p value
NRM-loc	NRM-glo	(1.463, 1.537)	2.01%
NRM-hyb	NRM-glo	(1.434, 1.566)	0.01%
NRM-hyb	NRM-loc	(1.465, 1.535)	3.09%
Rtr.-based	NRM-glo	(1.512, 1.488)	48.1%
Rtr.-based	NRM-loc	(1.533, 1.467)	6.20%
Rtr.-based	**NRM-hyb**	(1.552, 1.448)	0.32%
SMT	**NRM-hyb**	(1.785, 1.215)	0.00 %
SMT	**Rtr.-based**	(1.738, 1.262)	0.00 %

Table 3: p-values and average rankings of Friedman test for pairwise model comparison. (Rtr.-based means the retrieval-based method)

Text Understanding From Scratch

- Some ideas
  - 1. For tendency of deep learning, CNN, RNN and LSTM will use different problems like Short-Text Conversation, Text Understanding, Parser and etc.
  - 2. The methods of this paper is not so practical in the enterprise of industrial. The retrieval-based method is the mainly used method. But we can apply sentence vector to improve performance of learning to rank.

Thank You !