Sequence to backward and forward sequences: A content-introducing approach to generative short-text conversation

Lili Mou,[1] Yiping Song,[2] Rui Yan,[3] Ge Li,[1] Lu Zhang,[1] Zhi Jin[1]

[1]Institute of Software
[2]Institute of Network Computing and Information Systems
[3]Institute of Computer Science and Technology

School of EECS, Peking University, P. R. China

COLING, Osaka, Japan
December, 2016

Outline

1. Introduction: Human-Computer Conversation Systems
2. Approach: seq2BF Model
3. Experimental Results
4. Conclusion: Related Work

Outline

Human-Computer Conversation

Human-computer conversation has long attracted interest in both academia and industry.

- Task/Domain-oriented systems
- Open-domain conversation systems

Task/Domain-Oriented Dialog Systems

- Transportation domain: TRAIN-95 [Ferguson et al., 1996]
- Education: AutoTutor [Graesser et al., 2005]
- Restaurant booking [Wen et al., 2016]

Approaches:

- Planning
- Rule-based, Slot-filling, etc.

Open-Domain Conversation

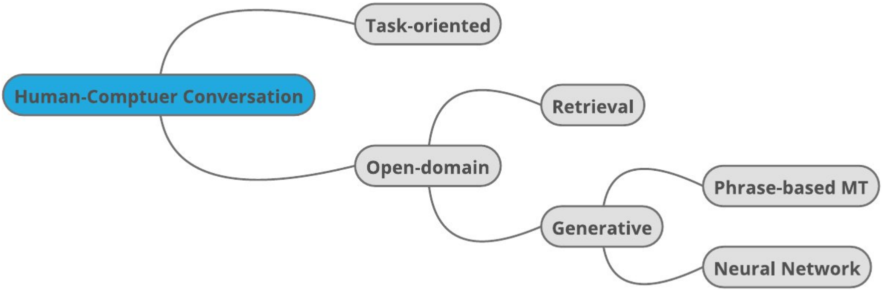
Why is chatbot-like conversation important?

- Tackles the problem of natural language understanding and generation
- Commercial needs

Approaches:

- Retrieval-based systems [Isbell et al., 2000, Wang et al., 2013]
- Generative systems
  - Phrase-based machine translation [Ritter et al., 2011]
  - Neural networks (seq2seq models) [Shang et al., 2015]

Where are we?



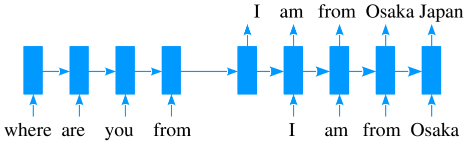
Open-domain, neural network-based, generative short-text conversation

Outline

Sequence-to-Sequence Model

Encode-decoder framework

- Encode a user-issued query as a vector
- Decode it as an utterance



Other applications:

- Machine translation
- Summarization
- etc

But the problem is. . .

Meaningless and universally relevant replies

- Example in a previous study (in English) [Li et al., 2016]
  - Q: How come you never say it?
  - Q: How much time do you have here?
  - R: “I don’t know”
- Our scenario (in Chinese)
  - 我也是 (Me too)

Why?

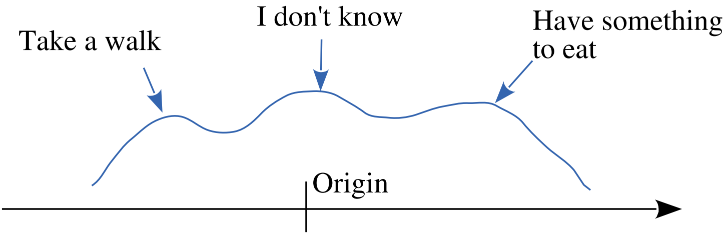
- The query does not convey sufficient information.
- A query may having multiple appropriate replies.
- Universally relevant utterances appear (slightly) more frequently than other replies

Conversation is different than translation.

Multi-Modality

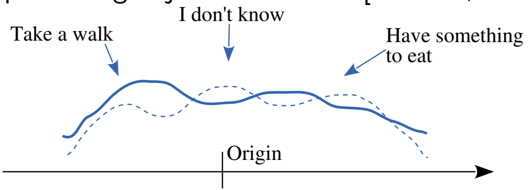
Query: What are you going to do?

Candidate replies:

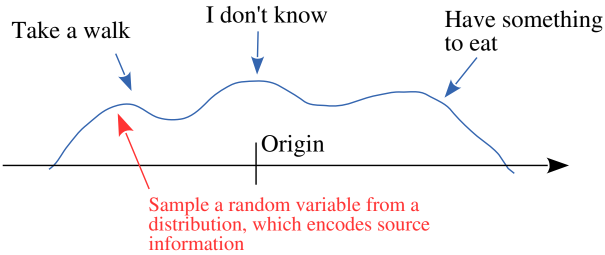


Solutions

- Diversity-promoting objective function [Li et al., 2016]



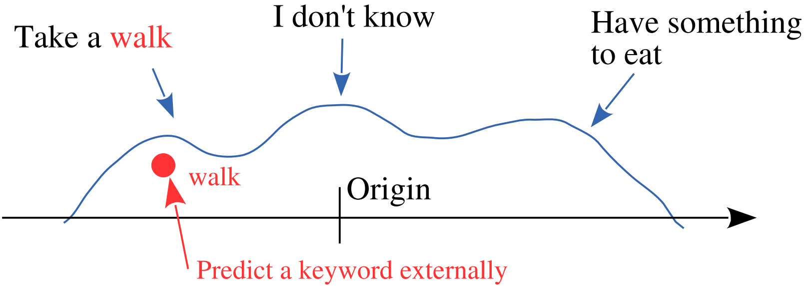
- Variational encoding the source [Serban et al., 2016]



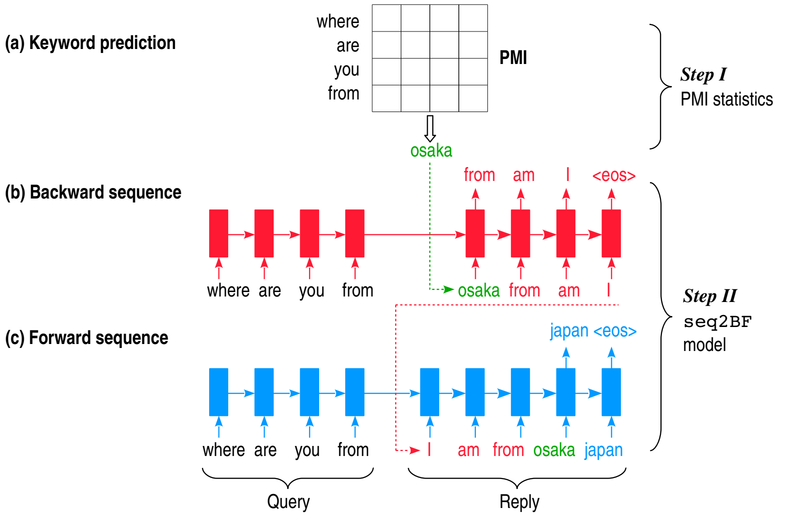
Our Intuition

- Some words in the utterance are highly correlated with the source.
  - Thank you
  - You're **welcome**
- Predict a keyword first, and generate a reply containing the keyword
- A "sequence to backward and forward sequences" model accomplishes this goal.

The View from Multi-Modality



Overview


(a) Keyword prediction
where
are
you
from
PMI
osaka
Step I
PMI statistics
(b) Backward sequence
from am I <eos>
where are you from
osaka from am I
(c) Forward sequence
japan <eos>
where are you from
I am from osaka japan
Query
Reply
Step II
seq2BF
model

Keyword Predictor

Computing the point-wise mutual information (PMI):

$$\mathrm{PMI}(w_q, w_r) = \log \frac{p(w_q, w_r)}{p(w_q)p(w_r)} = \log \frac{p(w_q|w_r)}{p(w_q)}$$

Prediction

$$w_r^* = \operatorname*{argmax}_{w_r} \mathrm{PMI}(w_{q_1} \cdots w_{q_n}, w_r)$$

where

$$\mathrm{PMI}(w_{q_1} \cdots w_{q_n}, w_r) = \log \frac{p(w_{q_1} \cdots w_{q_n}|w_r)}{p(w_{q_1} \cdots w_{q_n})}$$

$$\approx \log \frac{\prod_{i=1}^n p(w_{q_i}|w_r)}{\prod_{i=1}^n p(w_{q_i})} = \sum_{i=1}^n \log \frac{p(w_{q_i}|w_r)}{p(w_{q_i})} = \sum_{i=1}^n \mathrm{PMI}(w_{q_i}, w_r)$$

seq2BF Model

Traditional language models (sentence generators) start from the first word and generate following words in sequence.

$$\begin{aligned} p(r_1, \cdots, r_m | \boldsymbol{q}) &= p(r_1|\boldsymbol{q}) p(r_2|r_1, \boldsymbol{q}) \cdots p(r_m|r_1 \cdots r_{m-1}, \boldsymbol{q}) \\ &= \prod_{i=1}^{m} p(r_i|r_1 \cdots r_{i-1}, \boldsymbol{q}) \end{aligned}$$

The `seq2BF` model generates previous and future words conditioned on a given word.

$$p\left(\begin{array}{c} r_{k-1} \cdots r_1 \\ \hdashline r_{k+1} \cdots r_m \end{array} \middle| r_k, \boldsymbol{q} \right) = \prod_{i=1}^{k-1} p^{(\mathsf{bw})}(r_{k-i}|r_k, \boldsymbol{q}, \cdot) \prod_{i=1}^{m-k} p^{(\mathsf{fw})}(r_{k+i}|r_k, \boldsymbol{q}, \cdot)$$

Details

Asynchronously generating the two sequences

- First the backward half
- Then the forward half conditioned on the backward half sequence

Training:

- Randomly choose a word as the keyword, splitting an utterance into two subsentences

$\Rightarrow$ "half" language model

Outline

Datasets and Settings

- A Chinese conversation dataset containing up to $500k$ query-reply pairs
- RNNs are in character level, but keyword prediction is in phrase level.
- Embeddings and the hidden layer: 500-dimensional
- Rmsprop, generally default hyperparameters

Evaluation Metrics

Attempt#1: BLEU scores (0 BLEU-2 score by one annotator) (Abandoned)

Attempt#2: Human evaluation

- Pointwise annotation
- Pairwise annotation

Attemp#3: Intrinsic metric (entropy)

$$-\frac{1}{|R|}\sum_{w\in R}\log_2 p(w)$$

Results

Method	PointHuman	Length	Entropy
seq2seq	0.58	**5.61**	6.960
seq2BF$_-$	0.46	5.60	6.971
seq2BF$_+$	**0.67**	5.31	**9.139**
Groundtruth	–	9.19	8.832

	PairHuman		
Method	Wins	Ties	Loses
seq2seq	24.7	26.0	49.3
seq2BF$_+$	**49.3**	26.0	24.7

Examples

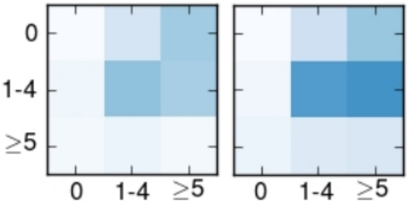
	Chinese	English (translated)
Query	李有男友公开过了	It's known that *Li*† has a boyfriend.
Grountruth	都已经分了 之前李的贴吧都在讨论了	Broken up. There's discussion in her *Tieba*.‡
seq2seq	我是男的	I am a male
seq2BF_{-}	我的男友	My boyfriend
seq2BF_{+}	有**绯闻**男友	Has a **rumor**ed boyfriend
Query	人大复试飘过	Passed second-round exam of Renming Univ.
Grountruth	这么牛，什么专业	Cool, what's your major
seq2seq	我也是！	Me too!
seq2BF_{-}	我也是！	Me too!
seq2BF_{+}	**分数**是什么	What is your **score**
Query	挺漂亮的 祝福祝福	So beautiful, congratulations
Grountruth	下手慢了哈。现在	You're late till now
seq2seq	我是你的头像	I'm in your photo
seq2BF_{-}	我是你的头像	I'm in your photo
seq2BF_{+}	第一张**图像**是谁	Who is in your first **photo**

Additional Analysis

Entropy

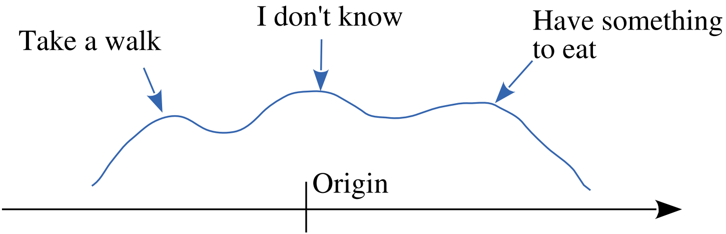
Model	seq2seq	seq2BF$_{-}$	seq2BF$_{+}$ keyword	seq2BF$_{+}$ remaining
Entropy	6.960	6.971	11.630	7.422

Length [Mou et al., 2015]



Outline

Conclusion



- Topic-augmenting [Xing et al., 2016]
- Combination of retrieval and generative dialog systems [Song et al., 2016]

Conclusion

Thank you for listening!

Q & A

References

Ferguson, G., Allen, J., and Miller, B. (1996).
TRAINS-95: Towards a mixed-initiative planning assistant.
In *Proceedings of Artificial Intelligence Planning Systems Conference*, pages 70–77.

Graesser, A. C., Chipman, P., Haynes, B. C., and Olney, A. (2005).
AutoTutor: An intelligent tutoring system with mixed-initiative dialogue.
IEEE Transactions on Education, 48(4):612–618.

Isbell, C. L., Kearns, M., Kormann, D., Singh, S., and Stone, P. (2000).
Cobot in LambdaMOO: A social statistics agent.
In *Proceedings of the 17th National Conference on Artificial Intelligence*, pages 36–41.

Li, J., Galley, M., Brockett, C., Gao, J., and Dolan, B. (2016).
A diversity-promoting objective function for neural conversation models.
In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 110–119.

Mou, L., Yan, R., Li, G., Zhang, L., and Jin, Z. (2015).
Backward and forward language modeling for constrained natural language generation.
arXiv preprint arXiv:1512.06612.

Ritter, A., Cherry, C., and Dolan, W. B. (2011).
Data-driven response generation in social media.
In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 583–593.

Serban, I. V., Sordoni, A., Lowe, R., Charlin, L., Pineau, J., Courville, A., and Bengio, Y. (2016).
A hierarchical latent variable encoder-decoder model for generating dialogues.
arXiv preprint arXiv:1605.06069.

Shang, L., Lu, Z., and Li, H. (2015).
Neural responding machine for short-text conversation.
In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 1577–1586.

Song, Y., Yan, R., Li, X., Zhao, D., and Zhang, M. (2016).
Two are better than one: An ensemble of retrieval-and generation-based dialog systems.
arXiv preprint arXiv:1610.07149.

Wang, H., Lu, Z., Li, H., and Chen, E. (2013).
A dataset for research on short-text conversations.
In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 935–945.

Wen, T.-H., Gasic, M., Mrksic, N., Rojas-Barahona, L. M., Su, P.-H., Ultes, S., Vandyke, D., and Young, S. (2016).
A network-based end-to-end trainable task-oriented dialogue system.
arXiv preprint arXiv:1604.04562.

Xing, C., Wu, W., Wu, Y., Liu, J., Huang, Y., Zhou, M., and Ma, W.-Y. (2016).
Topic augmented neural response generation with a joint attention mechanism.
arXiv preprint arXiv:1606.08340.