SG01 at the NTCIR-13 STC-2 Task

Haizhou Zhao, Yi Du, Hangyu Li, Qiao Qian, Hao Zhou, Minlie Huang, Jingfang Xu

Sogou Inc., Beijing, China | Tsinghua University, Beijing, China

NTCIR-13, Dec 5-8, 2017, Tokyo, Japan | contact: zhaohaizhou@sogou-inc.com

Introduction

We participate in NTCIR-13 Short Text Conversation (STC) **Chinese subtask**. In our system, we use the retrieval-based method and the generation-based method respectively. We have achieved **top performance** in both methods with 8 submissions.

Retrieval-based Method

In this part, we treat STC as an **IR problem**. We separate the process into stages, as it goes, we reduce the candidate set and introduce more complex features. In the end, we use learning to rank to get the final result list. Figure 1. describes the process of our retrieval-base method.

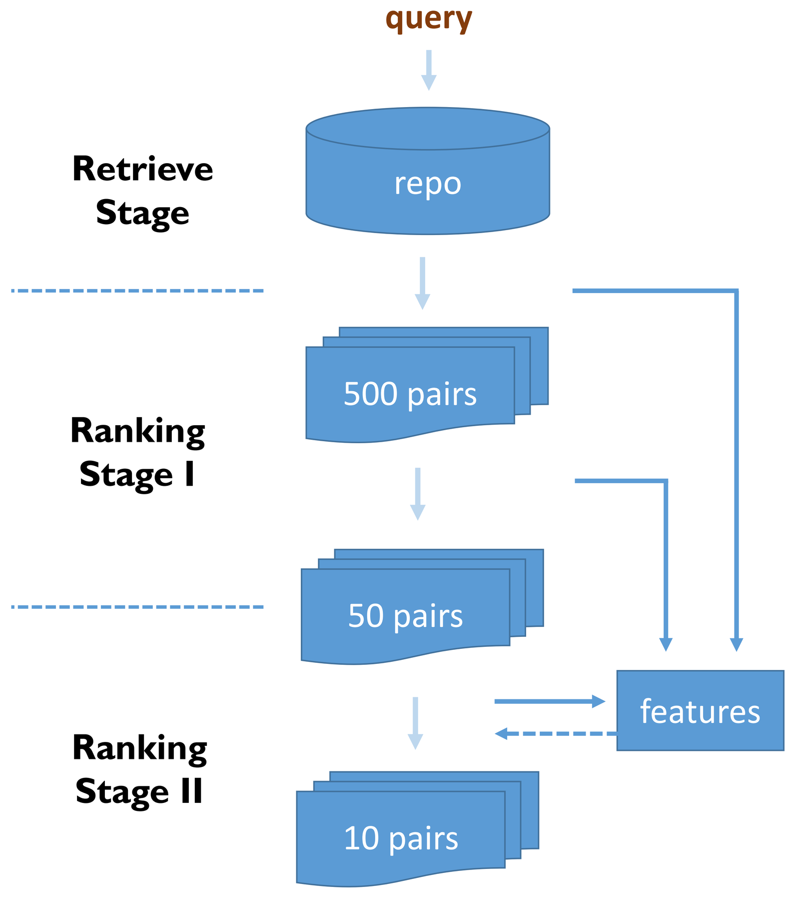

Figure 1. Diagram of Retrieval-based Method

Stage1: Retrieve Stage

At the beginning, we do data pre-processing to remove some low-quality post-comment pairs, then we put the repository into a light-weighted search engine, treating the post like a title and the comment like content.

For a given query, we retrieve **500** post-comment pairs from the repository for further comment selection. **Traditional features in IR** are used in this step, such as BM25, MRF for term dependency, Proximity, etc. These features will also be used in the final stage.

Stage2: Ranking Stage I

In this stage, we employ features designed for STC task:

- **cosine similarity** of TF-IDF Vector between:
- negative **Word Mover Distance** [M. J. Kusner 2015] between:
  - ✓ query ↔ post
  - ✓ query ↔ comment
  - ✓ query ↔ post + comment
- **Translation based language model** [Z. Ji 2014]

We treat each feature as a ranker, simply add the sequence number to get a final rank, we keep the **top 50** candidates.

Stage3: Ranking Stage II

We employ some **DNN features** to better capture rich structure in STC problem:

- $Score_{embd}$
- $Score_{BiLSTM+CNN}$ [R. Yan 2016]

↑ *Trained with a ranking-based objective, using given repository plus extra 12 million crawled post-comment pairs, noted as $Repo_{extn}$*

- $Score_{S2S-p2c}$
- $Score_{S2S-c2p}$

← *Defined in Generation-based Method*

At last, we use **LambdaMART** to perform learning to rank, all the features aforementioned will be used. The training data are 40 thous. labeled pairs. For each given query, we keep **top 10** pairs' comments as the final result.

Generation-based Method

In our generation-based method, we first generate various candidate comments, then perform ranking on them to get a preferable top 10 results. Figure 2. shows our generation-based method.

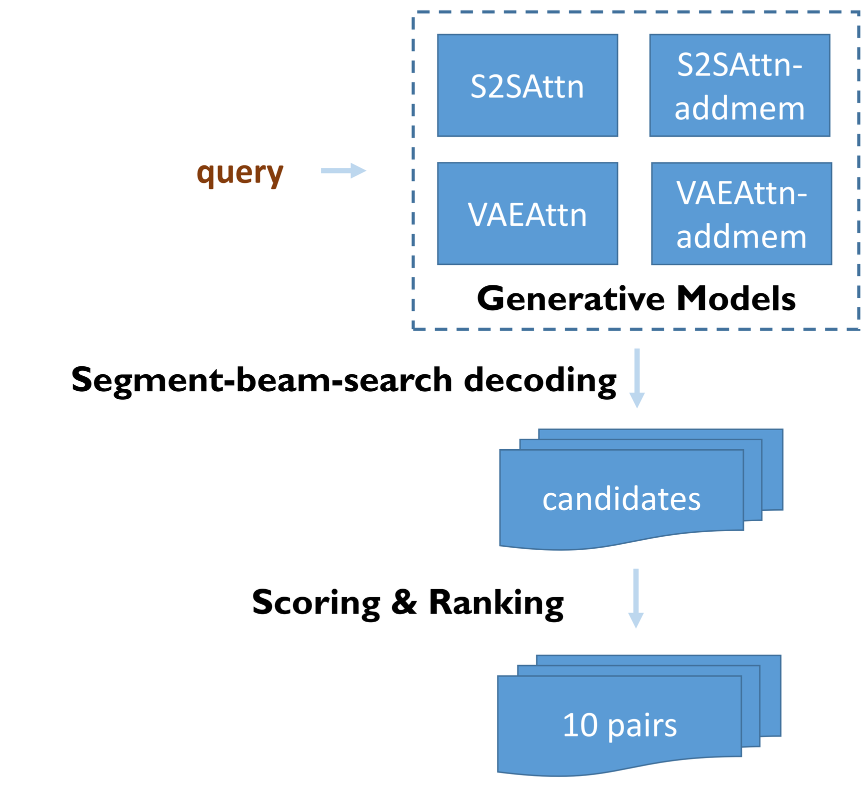

Figure 2. Diagram of Generation-based Method

Generative Models

We design 4 generative models to generate candidate comments, models are trained with $Repo_{extn}$, corpus is pre-processed by rules before training.

- ***S2SAttn***

  ↑ *seq2seq [I. Sutskever 2014] with attention mechanism*
- ***S2SAttn−addmem***

  ↑ *Add dynamic memory to the attention*
- ***VAEAttn***

  ↑ *Use Variational Auto-Encoder*
- ***VAEAttn−addmem***

Rank the Candidates

We define **likelihood** and **posterior** to rank the candidates. For a post X and a generated comment Y', we define $Score_{S2S-p2c}$ as a prediction of logarithmic $P(Y'|X)$, known as likelihood. We sum up likelihood scores from different models and implementations, noted as Li. As for posterior, we make the prediction $P(X|Y')$; so we have $Score_{S2S-c2p}$ and Po. We combine them in the following way to get the final ranking score:

$$score = \frac{\lambda * Li + (1-\lambda) * Po}{lp(Y')}$$

where $lp(Y') = \frac{(c+|Y'|)^{\alpha}}{(c+1)^{\alpha}}$ [Y. Wu 2016].

Before ranking, we also process the comments by rules to make them more fluent and to remove improper comments.

Submissions

Submission	L2R respect to	nG@1	P+	nERR@10
SG01-C-R1	nG@1	**0.5355**	0.6084	0.6579
SG01-C-R2	nERR@10	0.5168	0.5944	0.6461
SG01-C-R3	P+	0.5048	**0.6200**	**0.6663**

Table 1. Submissions of Retrieval-based Method

Submission	Fusion of candidates from	Scoring By	nG@1	P+	nERR @10
SG01-C-G5	$VAEAttn$, $VAEAttn-addmem$	Li	0.3820	0.5068	0.5596
SG01-C-G4	$S2SAttn$, $S2SAttn-addmem$	Li	0.4483	0.5545	0.6129
SG01-C-G3	$S2SAttn$, $S2SAttn-addmem$	Li & Po	0.5633	0.6567	0.6947
SG01-C-G2	$VAEAttn$, $VAEAttn-addmem$	Li & Po	0.5483	0.6335	0.6783
SG01-C-G1	*All 4 kinds of models*	Li & Po	**0.5867**	**0.6670**	**0.7095**

Table 2. Submissions of Generation-based Method

Case Study

We show some from our generation-based method submissions cases in Table 3. and Table 4. to reveal how improvements on baseline models benefit candidates generation and ranking.

Query 和家人一起喝喝茶，聊聊天，也是一种生活的乐趣 (Drink tea and chat with the family, what a joy of life)	
SG01-C-G3	我也是这样觉得 (I feel the same) 我也在看呢 (I'm watching too) 是啊，生活是一种享受 (Yes, life is joyful) 我也是。。。(Me too...) 是的，我也这么认为 (Yes, I also believe so) 我也是!!! (Me too!!!) 呵呵，是啊！ (Uh, yeah!) 是啊是啊！ (Yeah, yeah!) 是的，是的。 (Yes, yes.) 我也是这么想的 (I think so, too)
SG01-C-G4	是的，是的。 (Yes, yes.) 我也是。。。(Me too...) 我也是这么想的 (I think so, too) 我也是!!! (Me too!!!) 是啊，生活是一种享受 (Yes, life is joyful) 是啊是啊！ (Yeah, yeah!) 我也是这样觉得 (I feel the same) 是的，我也这么认为 (Yes, I also believe so) 呵呵，是啊！ (Uh, yeah!) 我也在看呢 (I'm watching too)

Table 3. Case Study 1

Query 杭州的亲们，我们已登机，等待起飞啦，暂别数日。 (My dear friends in Hangzhou, we are on board, waiting for take off, won't be seeing you for a while.)	
SG01-C-G1	辛苦了,注意安全! (You've had a long day, be safe!) 辛苦了。。。 (You've had a long day...) 也祝您节日快乐！ (Wish you a happy holiday, too!) 一定要注意安全啊！ (Must be safe!) 去哪啊? (Where are you going?) 一路平安,注意安全啊。。。 (Have a good trip, be safe...) 你要去哪里啊? (Where are you going?) 一路平安!!! (Have a good trip!!!) 祝您旅途愉快！ (Wish you a happy journey!) 我也在等飞机。。。 (I'm also waiting for boarding...)
SG01-C-G2	也祝您节日快乐！ (Wish you a happy holiday, too!) 一定要注意安全啊！ (Must be safe!) 祝您旅途愉快！ (Wish you a happy journey!) 杭州欢迎您！ (Welcome to Hangzhou!) 杭州欢迎你！ (Welcome to Hangzhou!) 回杭州了吗？ (Back to Hangzhou?) 什么时候来杭州啊？ (When coming to Hangzhou?) 来杭州了？ (Coming to Hangzhou?) 这么晚还不睡啊 (It's been late, still up?) 必须来支持！加油！ (Will support you! Good luck!)
SG01-C-G3	辛苦了,注意安全! (You've had a long day, be safe!) 去哪啊? (Where are you going?) 辛苦了。。。 (You've had a long day...) 你要去哪里啊? (Where are you going?) 一路平安,注意安全啊。。。 (Have a good trip, be safe...) 一路平安!!! (Have a good trip!!!) 我也在等飞机。。。 (I'm also waiting for boarding...) 好的，等你消息。 (Okay, wait for your message.) 谢谢亲们的支持！ (Thank you for your support!) 好的，谢谢！ (Okay, thanks!)

Table 4. Case Study 2

Analysis & Conclusions

On average, ***VAE*** does worse than traditional seq2seq, but it can bring in interesting candidates. The feature ***Po*** works, giving higher rank to more informative candidates. **Fusion of models** do better than single model, because the ranking will bring preferable candidates to top 10.

According to the evaluation results, the generation-based method does better, however, it still prunes to generate "safe" responses. Meanwhile, the retrieval-based method tends to get in-coherent comments. We also find that larger size of training data will help a lot.

References

- **Z. Ji**, Z. Lu, and H. Li. An information retrieval approach to short text conversation. CoRR, abs/1408.6988, **2014**.
- **M. J. Kusner**, Y. Sun, N. I. Kolkin, and K. Q. Weinberger. From word embeddings to document distances. In Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15, pages 957–966. JMLR.org, **2015**.
- **I. Sutskever**, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 27, pages 3104–3112. Curran Associates, Inc., **2014**.
- **Y. Wu**, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean. Google's neural machine translation system: Bridging the gap between human and machine translation. CoRR, abs/1609.08144, **2016**.
- **R. Yan**, Y. Song, X. Zhou, and H. Wu. "Shall I Be Your Chat Companion?": Towards an Online Human-Computer Conversation System. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM '16, pages 649–658, New York, NY, USA, **2016**. ACM.