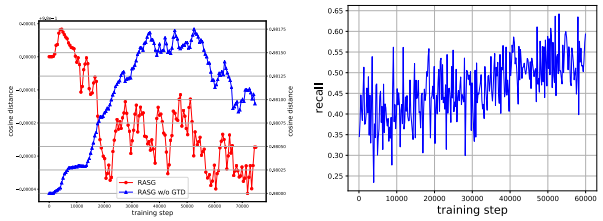

(a) (b)

Figure 2: (a) Cosine distance between decoding attention and reader attention. (b) Recall score of denoising module.

Table 5: Consistency and fluency comparison by human evaluation.

	Fluency		Consistency	
	mean	variance	mean	variance
S2S	2.17	0.24	1.98	0.28
CGU	2.20	0.26	2.08	0.29
RASG	2.65▲	0.21	2.48▲	0.26

a decrease of cosine distance and conversely, the RASG w/o GTD observes an increment of cosine distance. The fact that RASG can narrow the cosine distance proves that goal tracker and discriminator can lead the generator to follow the reader focused aspect.

Human evaluation

We ask three highly-educated Ph.D. students to rate 100 generated summaries of different models according to consistency and fluency. These annotators are all native speakers. The rating score ranges from 1 to 3 and 3 is the best. We take the average score of all summaries as the final score of each model, as shown in Table 5. It can be seen that RASG outperforms other baseline models in both sentence fluency and consistency by a large margin. We calculate the kappa statistics in terms of fluency and consistency, and the score is 0.33 and 0.29 respectively. To prove the significance of the above results, we also do the paired student t-test between our model and CGU model (row with shaded background), the p-value are 0.0017 and 0.0012 for fluency and consistency respectively.

Case analysis

Figure 3 shows a document and its corresponding summaries generated by different methods. We can observe that S2S does generate fluent summary. However, the generated aspect is contradictory to the focused aspect of reader or ground truth summary. Meanwhile, RASG overcomes this shortcoming by using goal vector and gap content given by goal tracker and supervisor at training stage, and produces the summary that is not only fluent but also consistent with main aspect of document.

document	徐麟表示，中央网信办将提供政策支持建立健全国有资本进入培育互联网企业，完善互联网企业国内上市等相关政策；通过在新闻网站核发新闻记者证，开展从业人员教育培训等措施，努力打造新媒体平台的国家队和主力军。详见长微博(Lin Xu said that the Central Network Office will provide policy support to establish a nationwide capital to enter the cultivation of internet companies, improve the policies of domestic listing of internet companies; through the press on the news website to issue a press card, carry out education and training for practitioners, and strive to create new the national team and the main force of the media platform. See Long Weibo for details.)
comments	国有资本必须介入(State-owned capital must be involved.)
	中央网信办不要把举措落实在文件上，要切实的执行！(The Central Network Office should not implement the measures on the documents and must implement them in a practical way !)
	我觉得网信办要好好治治这些害群之马了(I feel that the Central Network Office should cure these black sheep.)
reference	网信办副主任：建立健全国有资本进入培育互联网企业(Deputy director of the Central Network Office: Establishing state-owned capital to cultivate internet enterprises)
S2S	徐麟：努力打造新媒体平台的国家队和国家队(Lin Xu : National team and national team working hard to build a new media platform)
CGU	长微博：中央网信办将提供政策支持建立健全国有资本(Long Weibo: The Central Network Office will provide policy support to establish state-owned capital.)
RASG	中央网信办：努力打造新媒体平台的国家队(Central Network Office: Working hard to build a national team of new media platform.)

Figure 3: Examples of the generated summary by RASG and other models.

Conclusion

In this paper, we propose a new framework named *reader-aware summary generator* (RASG) which aims to generate summaries for document from social media incorporating the reader comments. In order to capture the reader focused aspect, we design a reader attention component with a denoising module to capture the alignment between comments and document. We employ a supervisor to measure the semantic gap between generated summary and reader focused aspect. A goal tracker uses the information of semantic gap and the feature extracted by the discriminator to produce a goal vector to guide the summary generator. In our experiments, we have demonstrated the effectiveness of RASG and have found significant improvements over state-of-the-art baselines in terms of ROUGE and human evaluations. Moreover, we have verified the effectiveness of each module in RASG for improving the summarization performance.

Acknowledgments

We would like to thank the anonymous reviewers for their constructive comments. We would also like to thank Zhujun Zhang, Sicong Jiang for their helps on this project. This work was supported by the National Key Research and Development Program of China (No. 2017YFC0804001), the National Science Foundation of China (NSFC No. 61876196, No. 61672058), Alibaba Innovative Research

(AIR) Fund. Rui Yan was sponsored by CCF-Tencent Open Research Fund and Microsoft Research Asia (MSRA) Collaborative Research Program.