Generating Long Financial Report using Conditional Variational Autoencoders with Knowledge Distillation

Yunpeng Ren[1,2], Ziao Wang[1], Yiyuan Wang[1] and Xiaofeng Zhang[1]
[1]School of Computer Science, Harbin Institute of Technology Shenzhen, China
[2]Qianhai Financial Holdings Co., Ltd., Shenzhen, China
Shenzhen University Town, Nanshan District, Shenzhen 518055
zhangxiaofeng@hit.edu.cn

Abstract

Automatically generating financial report from a piece of news is quite a challenging task. Apparently, the difficulty of this task lies in the lack of sufficient background knowledge to effectively generate long financial report. To address this issue, this paper proposes the conditional variational autoencoders (CVAE) based approach which distills external knowledge from a corpus of news-report data, and experimental results show that it can achieve the SOTA performance.

Introduction

Text generation has long been investigated in the domain of natural language processing with several sub problems, e.g., long text generation and summary generation. Among these sub problems, long text generation from short text is quite challenging especially in a domain-specific task, such as financial report generation. Particularly, the difficulty to generate financial reports given a piece of short news lies in the lack of sufficient information.

Recently, deep generative model based approaches (McCarthy et al. 2020; Wang and Wan 2019) and generative adversarial network (GAN) based approaches (Guo et al. 2017) were proposed for long text generation. For instance, the conditional variational autoencoder (Yang et al. 2018) with a hybrid decoder could learn topics to generate Chinese poems. And (Shen et al. 2019) designed a hierarchy of stochastic layers between the encoder and decoder component to learn a VAE model for generating long coherent text. However, there exist two challenges which invalidate most existing approaches. First, the length of the input news data is rather short. Second, the generation of financial reports by human specialists often involves their intellectual efforts, such as inferring and reasoning abilities. This paper is thus motivated to address these research difficulties. Particularly, we distill external knowledge from a large corpus of training data to provide sufficient information to guide the generation of financial report, and the overall framework of the proposed approach with knowledge distillation (CVAE-KD) is plotted in Figure 1.

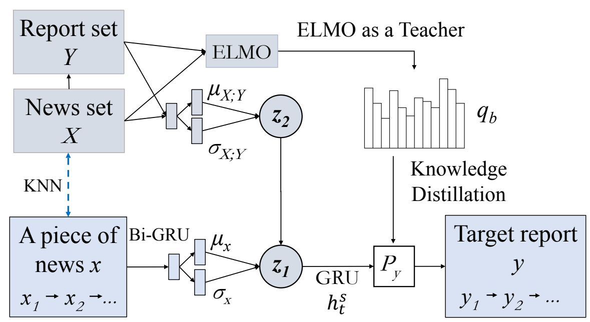

Figure 1: Framework of the proposed approach.

The Proposed Approach

In the proposed approach, the input news X is first embedded using lookup table. Then, Bi-GRU (Zhou et al. 2016) is adopted to encode each input x, and the latent variable distribution is learnt. For the background knowledge learning, a set of similar news are extracted as well as the corresponding reports to estimate a higher level latent variable distribution. At last, a GRU component is employed to decode the output report. The proposed CVAE-KD has three components, i.e., encoder component, background knowledge extraction component, decoder component.

The encoder component It employs a Bi-GRU component to embed the input x. The hidden state h_t^e of Bi-GRU is treated as the output of this component, h^e is calculated as

$$h_t^{xe} = (1 - c_t)h_{t-1}^{ex} + c_t h_t^{x'}; \quad h^e = [\overrightarrow{h_t^{xe}}, \overleftarrow{h_t^{xe}}],$$

where h_{t-1}^{ex} is the previous activation, $h_t^{x'}$ is the candidate activation, h_t^{xe} is the current activation and h^e is the concatenation of activation from both direction. We assume that the latent variables z_1 of CVAE follows a gaussian distribution. After learning the learnt latent variable distribution, we sample latent variable z_1 as

$$z_1^i = \mu_x^i + \sigma_x^i \epsilon, \epsilon \sim N(0, I)/N(\mu, \sigma).$$

Learning latent variable from background Knowledge We first extract the subset X_s and Y_s, as the background knowledge, for each x from the historical data. Another gaussian distribution is assumed for this background knowledge and the latent variable z_2 is then sampled from the learnt gaussian distribution, calculated as

$$\mu_{X_{em};Y_{em}} = f_{\mu_x}([X_{em}; Y_{em}]); \quad \sigma_{X;Y} = f_{\sigma_x}([X_{em}; Y_{em}])$$

$$z_2^i = \mu_{X;Y}^i + \sigma_{X;Y}^i \epsilon_{X;Y}, \epsilon_{X;Y} \sim N(0, I)/N(\mu, \sigma)$$

where X_{em} and Y_{em} feature representations of each input pair of data (news-report).

Knowledge distillation To distill knowledge from external knowledge base, X_s and Y_s are embedded using a pretrained model ELMO. To align the output of this component , a two-layers MLP is adopted. Y_{em} is considered as the teacher to supervise the generated financial report y. Accordingly, the student is the output of decoder component during model training process. The employed ELMO is to predict probability of next token to be generated, given a sequence of generated tokens in y. The knowledge distillation is to optimize below equation

$$L_{kd}(\theta) = -\sum_{w \in V} [P_\phi(y_t = \omega|x, y) log P_\theta(y_t = \omega|x, y_{1:t-1})],$$

where $P_\phi(y_t)$ is the soft target.

Model Loss We simply choose a GRU as the decoder component. Its output is fed into a MLP to generate the probability of a wordThe overall model loss of the proposed CVAE-KD contains two terms, i.e., CVAE loss and knowledge distillation loss. The CVAE loss contains two terms, and the first term is to calculate the reconstruction loss and the second term is to force the latent variable z_1 to approximate the latent variable z_2, written as

$$\begin{aligned} L_{CVAE}(x, y; \theta, \phi) = &-D_{KL}[q_\varphi(z_2|X, Y)||p_\theta(z_1|x)] \\ &+ E_{q_\phi(z_1|x, z_2)}[log p_\theta(y|z_1, x)] \end{aligned}$$

where $p_\theta(z_1|x)$, $q_\phi(z_1|x, z_2)$ and $q_\varphi(z_2|X, Y)$ are respectively calculated as

$$q_\phi(z_1|x, z_2) = N(z_1; \mu_x, \mu_{z_2}, \sigma_x, \sigma_{z_2})$$

$$p_\theta(z_1|x) = N(z_1; \mu_x, \sigma_x).$$

$$q_\varphi(z_2|X, Y) = N(z_2; \mu_{X,Y}, \sigma_{X,Y}).$$

At last, the overall model loss is written as

$$L_{total} = \alpha L_{CVAE} + (1 - \alpha) L_{kd}.$$

Experiments

The experimental results on the real-world news-report dataset are reported in Table 1 and 2. It is shown that the proposed CVAE-KD achieves the best results w.r.t. BLEU and ROUGE criteria. The corresponding BLEU-1, BLEU-2, BLEU-3 and BLEU-4 scores of CVAE-KD are 28.0%, 44.4%, 27.6%, 14.8% higher than the second best model, respectively. Similar observations could be found from Table 2. This verifies the effectiveness of the proposed approach.

Methods	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Seq2seq	32.69	7.65	4.85	2.75
Seq2seq+Attn	33.64	13.85	9.89	6.92
Pointer-Generator	36.45	9.51	5.75	2.45
CVAE	33.5	14.07	10.04	6.97
CVAE-KD	**46.67**	**20.32**	**12.81**	**8.00**

Table 1: The BLEU results of all compared methods.

Methods	ROUGE-1	ROUGE-2	ROUGE-L
Seq2seq	8.46	1.30	3.59
Seq2seq+Attn	15.66	3.02	3.89
Pointer-Generator	12.08	1.40	3.44
CVAE	16.69	**3.32**	4.65
CVAE-KD	**18.27**	2.64	**6.95**

Table 2: The ROUGE results of all compared methods.

Conclusion

This paper proposes the conditional variational autoencoders based approach with knowledge distillation (CVAE-KD) to automatically generate long financial reports from a piece of short news. Empirical results on a public news-report dataset demonstrate that the proposed approach achieves the state-of-the-art performance.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant No.61872108, and the Shenzhen Science and Technology Program under Grant No.JCYJ20170811153507788.

References

Guo, J.; Lu, S.; Cai, H.; Zhang, W.; Yu, Y.; and Wang, J. 2017. Long Text Generation via Adversarial Training with Leaked Information. In *AAAI*, 5141–5148.

McCarthy, A. D.; Li, X.; Gu, J.; and Dong, N. 2020. Addressing Posterior Collapse with Mutual Information for Improved Variational Neural Machine Translation. In *ACL*, 8512–8525.

Shen, D.; Celikyilmaz, A.; Zhang, Y.; Chen, L.; Wang, X.; Gao, J.; and Carin, L. 2019. Towards Generating Long and Coherent Text with Multi-Level Latent Variable Models. In *ACL*, 2079–2089.

Wang, T.; and Wan, X. 2019. T-CVAE: transformer-based conditioned variational autoencoder for story completion. In *IJCAI*, 5233–5239.

Yang, X.; Lin, X.; Suo, S.; and Li, M. 2018. Generating Thematic Chinese Poetry using Conditional Variational Autoencoders with Hybrid Decoders. In *IJCAI*, 4539–4545.

Zhou, P.; Shi, W.; Tian, J.; Qi, Z.; Li, B.; Hao, H.; and Xu, B. 2016. Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification. In *ACL*, volume 2, 207–212.