Information Engineering Express
International Institute of Applied Informatics
2024, Vol. 10, No. 1, IEE821

Improving the Consistency of Dialog Models Through Speaker Separation Learning

Sakuei Onishi [*] , Takamune Onishi [†] , Hiromitsu Shiina [‡]

Abstract

In recent years, dialog systems, a type of application in the field of natural language processing, have become more prevalent in our daily lives, such as through help desk services. In dialog response generation, responses generated for a specific context may differ from those for other contexts not only grammatically but also semantically in some cases. Thus, simply applying translation technologies would cause issues with the diversity of the generated responses. Previous studies, such as VHRED and GVT, used sampled latent variables for response generation to achieve response diversity. In this study, we propose a method (extended GVTSC) for classifying dialogs before reflecting them in internal dialog processing, in addition to the characteristics of each speaker, to improve diversity while maintaining consistency.

Keywords: Dialogue System, User-RNN, Conditional Variational Autoencoder, Global Variational Transformer, Extended GVT.

1 Introduction

The field of natural language processing is dramatically changing with the development of deep learning models. The Transformer[1] model has been proposed as an alternative to RNN and LSTM[2][3][4], which are suited to sequence-type data processing and can generate sentences. Language processing analysis that takes context into account is now possible, as seen in BERT[5], one of Transformer ' s models. Meanwhile, GTP-2[6] is capable of generating word definitions. The current dialog systems that are used for practical purposes are not generation-based dialog systems[7] that may generate inaccurate or risk-averse responses[8][9], but rather rule-based dialog systems where accuracy is guaranteed by manually generated rules.

RNN-based HRED[10] and VHRED[11] models have been announced. However, because these models do not distinguish between utterances made by two parties in a dialog, the characteristics of each party may become confused, resulting in a loss of consistency in response generation. The GVT model[12], in which the Transformer model is applied to

[*] Graduate School of Informatics, Okayama University of Science, Okayama, Japan
[†] Systems Nakashima, Okayama, Japan
[‡] A Faculty of Informatics, Okayama University of Science, Okayama, Japan

[10] I. V. Serban, A. Sordoni, Y. Bengio, A. Courville, and J. Pineau, "Building end-to-end dialogue systems using generative hierarchical neural network models," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, ser. AAAI'16. AAAI Press, 2016, p. 3776–3783.

[11] I. Serban, A. Sordoni, R. Lowe, L. Charlin, J. Pineau, A. Courville, and Y. Bengio, "A hierarchical latent variable encoder-decoder model for generating dialogues," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/10983

[12] Z. Lin, G. I. Winata, P. Xu, Z. Liu, and P. Fung, "Variational transformers for diverse response generation," *arXiv preprint arXiv:2003.12738*, 2020.

[13] B. Sun, S. Feng, Y. Li, J. Liu, and K. Li, "Generating relevant and coherent dialogue responses using self-separated conditional variational AutoEncoders," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online: Association for Computational Linguistics, Aug. 2021, pp. 5624–5637. [Online]. Available: https://aclanthology.org/2021.acl-long.437

[14] J. Li, M. Galley, C. Brockett, J. Gao, and B. Dolan, "A diversity-promoting objective function for neural conversation models," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, California: Association for Computational Linguistics, Jun. 2016, pp. 110–119. [Online]. Available: https://aclanthology.org/N16-1014

[15] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "Bertscore: Evaluating text generation with bert," *In International Conference on Learning Representation*, 2019.

[16] M. Inaba, "A example based dialogue system using the open 2channel dialogue corpus," *Journal of Japanese Society for Artificial Intelligence*, vol. 87, pp. 129—-132, 2019.

[17] T. Zhao, R. Zhao, and M. Eskenazi, "Learning discourse-level diversity for neural dialog models using conditional variational autoencoders," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 654–664. [Online]. Available: https://aclanthology.org/P17-1061

[18] S. R. Bowman, L. Vilnis, O. Vinyals, A. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," in *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*. Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 10–21. [Online]. Available: https://aclanthology.org/K16-1002

[19] X. Zhou and W. Y. Wang, "MojiTalk: Generating emotional responses at scale," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 1128–1137. [Online]. Available: https://aclanthology.org/P18-1104