ConceptNet infused DialoGPT for Underlying Commonsense Understanding and Reasoning in Dialogue Response Generation

Ye Liu[1], **Wolfgang Maier**[1], **Wolfgang Minker**[2] and **Stefan Ultes**[1]
[1]Mercedes-Benz AG, Sindelfingen, Germany
{ye.y.liu,wolfgang.mw.maier,stefan.ultes}@mercedes-benz.com
[2]Ulm University, Ulm, Germany
wolfgang.minker@uni-ulm.de

1 Introduction

Many pre-trained transformer-based (Vaswani et al., 2017) language models (LMs) have been widely applied in spoken dialogue systems (SDS) and shown promising performance. However, the probing experiments in Zhou et al. (2021b) demonstrated that pre-trained LMs (Zhang et al., 2020; Roller et al., 2021; Lewis et al., 2020) fail to capture commonsense (CS) knowledge hidden in dialogue utterances, even though they were already pre-trained with numerous datasets.

To improve the CS understanding and reasoning ability of a pre-trained model and to build a dialogue agent like shown in Figure 1, we have two main contributions in this work. We firstly inject external knowledge into a pre-trained conversational model to establish basic commonsense. Secondly, we leverage this integrated commonsense capability to improve open-domain dialogue response generation so that the dialogue agent is capable of understanding the CS knowledge hidden in dialogue history on top of inferring related other knowledge to further guide response generation.

2 Enabling commonsense capability

To enable the commonsense capability of the pre-trained conversational model DialoGPT, commonsense triplets of ConceptNet (Liu and Singh, 2004)–a large-scale knowledge graph–are infused through efficient Adapter tuning (Pfeiffer et al., 2021). By utilizing the AdapterHub (Pfeiffer et al., 2020), the Houlsby Adapter (Houlsby et al., 2019) is used, which includes two bottleneck adapters in each transformer layer: one after the multi-head attention sub-layer and another after the feed-forward sub-layer. To efficiently integrate this external knowledge into DialoGPT, only parameters of Adapter layers are updated and the parameters of DialoGPT are frozen during training.

To achieve this, we adapt the work from Perozzi et al. (2014); Lauscher et al. (2020) and induce a synthetic corpus from ConceptNet through a bias random walk (Grover and Leskovec, 2016). We convert relations in ConceptNet into natural language phrases (“IsA” to “is a”) and every relation is along with [] in collected data (Table 1) to distinguish relations from normal words/concepts. Finally, we created $359,421$ data points and split them into 80%/10%/10% train/valid/test set. During CS_Adapter tuning, we add one special token “<|commonsense|>” to the DialoGPT tokenizer and insert it at the beginning of every prompt input (Table 1). Given the auto-regressive property of DialoGPT, four prompt templates (Table 1) are proposed and randomly chosen as input. Given the knowledge prompt, the ConceptNet integrated DialoGPT+CS_Adapter can generate series of CS triplets (like the data in Table 1).

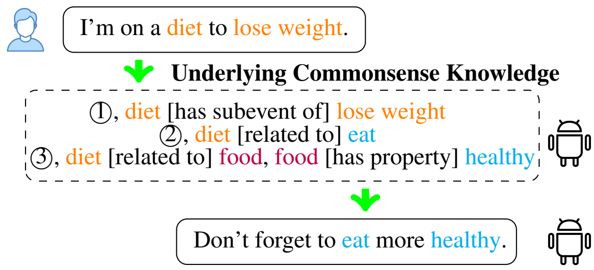

Figure 1: The ideal dialogue agent can understand the CS knowledge hidden in the dialogue history (①), meanwhile, infer the reasonable CS knowledge (② and ③) for further guiding an informative response generation. The key words/concepts are highlighted in orange for user utterance and blue for system response, respectively. The words highlighted in purple are middle concepts extracted in two-hop searching.

3 CS-based Response Model

To enable commonsense-based open-domain response generation, we utilize the Commonsense-Dialogues (Zhou et al., 2021a) dataset to continually train the DialoGPT+CS_Adapter presented in Section 2. This time, all parameters are updated.

Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue, August, 22-24, 2022, Dublin.

data	autobraking [related to] automatic, automatic [derived from] auto, auto [related to] automobile, ...
prompt templates	<|commonsense|> autobraking [related to] <|commonsense|> autobraking [related to] automatic, <|commonsense|> autobraking [related to] automatic, automatic [derived from] <|commonsense|> autobraking [related to] automatic, automatic [derived from] auto,

Table 1: One data example in synthetic corpus from ConceptNet and four prompt templates randomly as input.

model	perplexity↓	concepts Acc (%)	assertion Acc (%)
DialoGPT baseline	1.405	-	-
DialoGPT+CS_Adapter (ConceptNet Integration in 2)	-	56.88	47.29
DialoGPT+CS_Adapter (Response Model in 3)	1.365	62.43	45.27

Table 2: The automatic metrics of DialoGPT baseline, DialoGPT+CS_Adapter knowledge integration (Section 2) and DialoGPT+CS_Adapter response model (Section 3).

	annotator 1	annotator 2
yes *vs* no	87 *vs* 13	88 *vs* 12
positive agreement	93.71%	

Table 3: The human assessment results on generated CS triplets that do not officially exist in ConceptNet.

As shown in Figure 1, the goal is a dialogue agent that is capable of *understanding* CS knowledge hidden in the dialogue history (like 1 in Figure 1). Furthermore, the agent is also capable of *reasoning* other CS triplets for guiding the response generation. For this, we extract the knowledge triplets of keywords hidden in the dialogue history and the response (like 2 and 3 in Figure 1). For CS knowledge extraction, we firstly extract the key words in dialogue history and response reference. We adapt the work from Tang et al. (2019) and Zhong et al. (2021) that use TF-IDF and Part-Of-Speech (POS) features to select the keywords. Secondly, we extract the CS knowledge of these keywords from ConceptNet, i.e., one-hop and two-hop triplets with the keywords as root. Considering two-hop results includes triplets where the source and target concepts have no direct connection but share a common middle concept (3 in Figure 1).

During DialoGPT+CS_Adapter training, maximal 3 turns' dialogue context is as input for memory-efficiency, the extracted CS triplets, where the "<|commonsense|>" is inserted, and response are as label. Meanwhile, we add two new tokens: "[USER]" and "[SYSTEM]" to distinguish the user utterance from system response. Afterwards, the DialoGPT+CS_Adapter can generate both underlying CS knowledge and reasonable response.

4 Results and Discussion

To evaluate the CS knowledge integration in DialoGPT we use two automatic metrics. One is *concepts accuracy*, which represents the proportion of generated (head concept, tail concept) pairs that exist in ConceptNet without considering if the generated relation is officially correct. Another is *assertion accuracy*, which represents the proportion of generated (head concept, relation, tail concept) triplets that officially exist in ConceptNet. In order to further test our assumption–even if the generated commonsense triplets do not officially exist in ConceptNet, they still make sense for humans–we manually evaluate the generated CS knowledge. For this, two Master students with computational linguistic background were hired. And the human evaluation results shown in Table 3 support our assumption. The result comparison in Table 2 demonstrate that our final DialoGPT+CS_Adapter response model has comparative performance on knowledge generation compared to the plain DialoGPT+CS_Adapter after ConceptNet integration and lower perplexity (Serban et al., 2015) compared to the DialoGPT baseline.

In this work, we found several shortcomings that need to be discussed in our future work. One is that the relation distribution in ConceptNet is severely imbalanced which results in an over-generation of the "[related to]" relation. Another shortcoming is that our method of extracting CS triplets hidden in the dialogue is rule-based. It includes keywords extraction and knowledge extraction without considering the discourse information. A next step will be the application of neural network methods for knowledge extraction.

References

Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for nlp. In *International Conference on Machine Learning*, pages 2790–2799. PMLR.

Anne Lauscher, Olga Majewska, Leonardo FR Ribeiro, Iryna Gurevych, Nikolai Rozanov, and Goran Glavaš. 2020. Common sense or world knowledge? investigating adapter-based knowledge injection into pretrained transformers. In *Proceedings of Deep Learning Inside Out (DeeLIO): The First Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, pages 43–49.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880.

Hugo Liu and Push Singh. 2004. Conceptnet—a practical commonsense reasoning tool-kit. *BT technology journal*, 22(4):211–226.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710.

Jonas Pfeiffer, Aishwarya Kamath, Andreas Rücklé, Kyunghyun Cho, and Iryna Gurevych. 2021. Adapterfusion: Non-destructive task composition for transfer learning. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 487–503.

Jonas Pfeiffer, Andreas Rücklé, Clifton Poth, Aishwarya Kamath, Ivan Vulić, Sebastian Ruder, Kyunghyun Cho, and Iryna Gurevych. 2020. Adapterhub: A framework for adapting transformers. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 46–54.

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, et al. 2021. Recipes for building an open-domain chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325.

Iulian V Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. 2015. Hierarchical neural network generative models for movie dialogues. *arXiv preprint arXiv:1507.04808*, 7(8):434–441.

Jianheng Tang, Tiancheng Zhao, Chenyan Xiong, Xiaodan Liang, Eric Xing, and Zhiting Hu. 2019. Target-guided open-domain conversation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5624–5634.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and William B Dolan. 2020. Dialogpt: Large-scale generative pre-training for conversational response generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 270–278.

Peixiang Zhong, Yong Liu, Hao Wang, and Chunyan Miao. 2021. Keyword-guided neural conversational model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14568–14576.

Pei Zhou, Karthik Gopalakrishnan, Behnam Hedayatnia, Seokhwan Kim, Jay Pujara, Xiang Ren, Yang Liu, and Dilek Hakkani-Tur. 2021a. Commonsense-focused dialogues for response generation: An empirical study. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Singapore and Online. Association for Computational Linguistics.

Pei Zhou, Pegah Jandaghi, Hyundong Cho, Bill Yuchen Lin, Jay Pujara, and Xiang Ren. 2021b. Probing commonsense explanation in dialogue response generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4132–4146.