The Adapter-Bot: All-In-One Controllable Conversational Model

Zhaojiang Lin*, Andrea Madotto*, Yejin Bang, Pascale Fung
Center for Artificial Intelligence Research (CAiRE)
The Hong Kong University of Science and Technology
{zlinao,amadotto,yjbang,pascale}@ust.hk

Abstract

In this paper, we present the Adapter-Bot, a generative chatbot that uses a fixed backbone conversational model such as DialGPT (Zhang et al. 2019) and triggers on-demand dialogue skills via different adapters (Houlsby et al. 2019). Each adapter can be trained independently, thus allowing a continual integration of skills without retraining the entire model. Depending on the skills, the model is able to process multiple knowledge types, such as text, tables, and graphs, in a seamless manner. The dialogue skills can be triggered automatically via a dialogue manager, or manually, thus allowing high-level control of the generated responses. At the current stage, we have implemented 12 response styles (e.g., positive, negative etc.), 6 goal-oriented skills (e.g. weather information, movie recommendation, etc.), and personalized and emphatic responses.

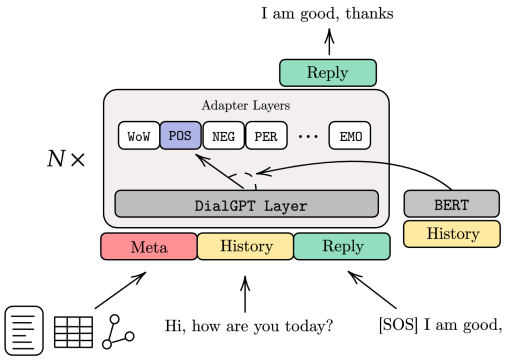

Figure 1: Adapter-Bot high-level architecture.

Introduction

Large pre-trained language models have greatly improved the state-of-the-art in many down-stream tasks. Similarly, transformer-based conversational models trained on large unlabeled human-to-human conversation (i.e. Reddit comments) (Zhang et al. 2019; Adiwardana et al. 2020; Roller et al. 2020) have shown excellent performance in modelling human responses. These models are capable of generating coherent and fluent responses.

Despite their capabilities, existing large conversational models are unable to on-demand control generated responses. For instance, once these conversational language models are fine-tuned on multiple conversational datasets, there is no mechanism (e.g., control codes or latent variables) for controlling which response 'skill' to use. Furthermore, these conversational models are unable to add dialogue 'skills' continuously without retraining all the model parameters. In advanced smart speakers such as Alexa, multiple skills can be added easily since the overall infrastructure is hard-coded, but in deep learning models, adding new conversational skills, without catastrophically forgetting all the previous ones, is challenging.

To overcome these challenges, we propose the Adapter-Bot, a dialogue model that uses a fixed DialGPT (Zhang et al. 2019) and triggers on-demand dialogue skills via different Adapters (Houlsby et al. 2019). Each adapter can be trained independently, thus allowing a continual integration of skills without retraining the entire conversational model. Moreover, we propose two interaction modes, automatic and manual. In the first, we train a Dialogue Manager (BERT (Devlin et al. 2019)) to select which adapter to use give a certain dialogue history. In the second, we let the user decide what style or skill to use for the response, to show high-level control over the chatbot.

System Description

Adapter Bot

The Adapter-Bot is composed of a fixed DialoGPT backbone, parameterized with Θ, and a set of residual adapters (Houlsby et al. 2019), parameterized as $\{\theta_1, \cdots, \theta_p\}$. Each of the adapter is indexed by its own skill-id (i.e., index t refers to adapter t with parameters θ_t). The skill-id can be either selected by the user or classified by a neural dialogue manager. As illustrated by Figure 1, given dialogue history X, and skill-id t, the Adapter-Bot retrieve external knowledge M, if need, then it generates the response:

$$Y = f_{\Theta,\theta_t}(X, M, t). \quad (1)$$

We train the Adapter-Bot with multiple datasets, such as Wizard-of-Wikipedia (Dinan et al. 2018), OpenDialKG (Moon et al. 2019), Stanford-Multi-Domain (Eric et al.

*Equal contributions.

2017), Empathetic Dialogue (Eric et al. 2017), Persona Chat (Zhang et al. 2018; Dinan et al. 2019), and the synthetic datasets generated by Plug-and-Play response style controlling method (Madotto et al. 2020). We denote the dialogue datasets as $\mathcal{D} = \{D^1, \cdots, D^p\}$, where each dataset D^t is made of dialogues with their corresponding meta-knowledge M aligned. Then, we optimize the adapter parameters in θ_t to minimize the negative log-likelihood over the dataset of dialogues D^t and its knowledge M. We evaluate our model using automatic evaluation by comparing it with existing state-of-the-art conversational models, all the results can be found in full version.

Dialogue Manager The dialogue manager is trained to select the right dialogue skill by predicting the index of the residual adapter. More formally, given the dialogue history X the dialogue manger predicts an index in $1, \cdots, p$. The dialogue adapter can be any classifier, and it is trained using the same set of dialogue datasets $\mathcal{D}$, but instead of using the response as supervision, we use the adapter index of the corresponding dialogue. For example, to select adapter t, we train the dialogue manager to predict the index t from the dialogue histories in D^t.

Knowledge Retrieval We apply different strategies to retrieve knowledge from different sources. To fetch the relevant information from Wikipedia, we use the TF-IDF retriever implemented by Chen et al. (2017), which computes the dot product of the TF-IDF weighted vector between the last user utterance and the Wikipedia articles. Then, the first paragraph of the highest score article is used as meta-information. To retrieve information from a knowledge graph, we first extract the entities from user utterances and match them with the entity node. Then we return the first-order neighbours. We store the extracted sub-graph as set of triples in the form (entity1, relation, entity2). To query the online API (e.g., weather API), we use heuristic rules to extract the slot values (e.g., location) from dialogue context.

User Interface

To make the system easily accessible, we establish a web-based demo, based on BotUI, for chatting with the Adapter-Bot. The demo supports both manual-mode and auto-mode. In addition to the above-mentioned dialogue skills, we also add multiple features:

- **Emotion Face Recognition** We deploy a javascript-based face emotion recognition model (face-api.js), for monitoring the involuntary reaction (Fan, Lam, and Li 2020) of the user while interacting with the Adapter-Bot. This model runs directly on the user browser; thus it does not require sending any images to the server. This is important for two reasons: privacy and real-time performance.
- **Text Emotion Recognition** We deploy a text-based emoji-classifier for text using deepMoji (Felbo et al. 2017). This is used to make the chat-bot more empathetic by showing the corresponding emoji turn by turn in the interface.

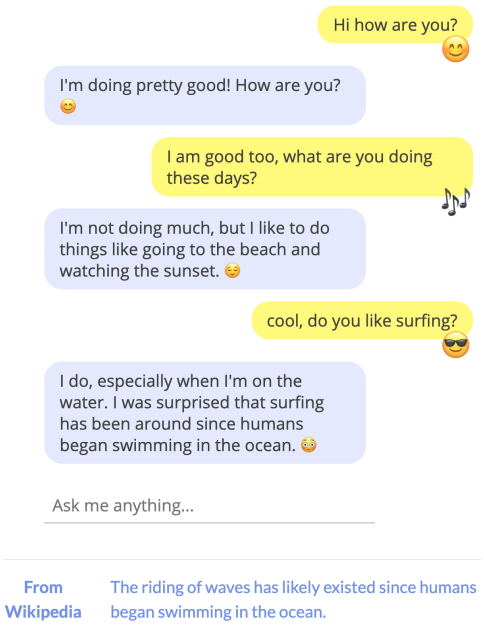

Figure 2: Adapter-Bot UI.

- **Toxic Classifier** We deploy a toxic classifier to detect possible offensive responses from the model. The classifier, BERT-base, is trained using the Toxic Comment Classification Dataset and deployed using IBM docker-container.
- **Covid-19** To show the flexibility of our model, we implement two further skills: Covid-19 QA (Su et al. 2020) and Covid-19 fact-checker (Lee et al. 2020). The first is accessed with an API-call that, given a question about Covid-19, returns the answer based on a large repository of scientific articles. The second is deploy with an adapter, but instead of training it to generate a response, it is used to score the falseness of a given claim. These two skills can be triggered in manual-mode only, and they show how the same backbone model can also be used to deploy non-dialogue skills.
- **Visualization** To show the grounding knowledge used by the model at each turn we deploy three visualization: graph, document (i.e., Wiki articles) and table (i.e., weather information). The visualization are developed using D3.js and an example of graph visualization is shown in Figure 2.

Conclusion

In this paper, we presented the Adapter-Bot, a dialogue model that is built with a fixed pre-trained conversational model and multiple trainable light-weight adapters. The model allows high-level control of different dialogue skills and continuous skills integration. We preliminarily showed 8 goal-oriented skills, 12 response styles, and personalized and emphatic responses. A web-based demo is established to make the system easily accessible.

References

Adiwardana, D.; Luong, M.-T.; So, D. R.; Hall, J.; Fiedel, N.; Thoppilan, R.; Yang, Z.; Kulshreshtha, A.; Nemade, G.; Lu, Y.; et al. 2020. Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977* .

Chen, D.; Fisch, A.; Weston, J.; and Bordes, A. 2017. Reading wikipedia to answer open-domain questions. *arXiv preprint arXiv:1704.00051* .

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186.

Dinan, E.; Logacheva, V.; Malykh, V.; Miller, A.; Shuster, K.; Urbanek, J.; Kiela, D.; Szlam, A.; Serban, I.; Lowe, R.; et al. 2019. The Second Conversational Intelligence Challenge (ConvAI2). *arXiv preprint arXiv:1902.00098* .

Dinan, E.; Roller, S.; Shuster, K.; Fan, A.; Auli, M.; and Weston, J. 2018. Wizard of wikipedia: Knowledge-powered conversational agents. *arXiv preprint arXiv:1811.01241* .

Eric, M.; Krishnan, L.; Charette, F.; and Manning, C. D. 2017. Key-Value Retrieval Networks for Task-Oriented Dialogue. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, 37–49.

Fan, Y.; Lam, J. C.; and Li, V. O. K. 2020. Facial Action Unit Intensity Estimation via Semantic Correspondence Learning with Dynamic Graph Convolution. In *AAAI*, 12701–12708.

Felbo, B.; Mislove, A.; Søgaard, A.; Rahwan, I.; and Lehmann, S. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524* .

Houlsby, N.; Giurgiu, A.; Jastrzebski, S.; Morrone, B.; De Laroussilhe, Q.; Gesmundo, A.; Attariyan, M.; and Gelly, S. 2019. Parameter-Efficient Transfer Learning for NLP. In *International Conference on Machine Learning*, 2790–2799.

Lee, N.; Bang, Y.; Madotto, A.; and Fung, P. 2020. Misinformation has High Perplexity. *arXiv preprint arXiv:2006.04666* .

Madotto, A.; Ishii, E.; Lin, Z.; Dathathri, S.; and Fung, P. 2020. Plug-and-Play Conversational Models. *arXiv preprint arXiv:2010.04344* .

Moon, S.; Shah, P.; Kumar, A.; and Subba, R. 2019. Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 845–854.

Roller, S.; Dinan, E.; Goyal, N.; Ju, D.; Williamson, M.; Liu, Y.; Xu, J.; Ott, M.; Shuster, K.; Smith, E. M.; et al. 2020. Recipes for building an open-domain chatbot. *arXiv preprint arXiv:2004.13637* .

Su, D.; Xu, Y.; Yu, T.; Siddique, F. B.; Barezi, E. J.; and Fung, P. 2020. CAiRE-COVID: A Question Answering and Multi-Document Summarization System for COVID-19 Research. *arXiv preprint arXiv:2005.03975* .

Zhang, S.; Dinan, E.; Urbanek, J.; Szlam, A.; Kiela, D.; and Weston, J. 2018. Personalizing Dialogue Agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2204–2213. Association for Computational Linguistics. URL http://aclweb.org/anthology/P18-1205.

Zhang, Y.; Sun, S.; Galley, M.; Chen, Y.-C.; Brockett, C.; Gao, X.; Gao, J.; Liu, J.; and Dolan, B. 2019. DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation. *arXiv preprint arXiv:1911.00536* .