EasyChair Preprint

№ 13434

An Open-Source ChatBot by Using ParlAI

Kiran Mai Narnavaram and Dan Lo

May 26, 2024

An Open-Source Chatbot by Using ParlAI

Kiran Mai Narnavaram
Kennesaw State University
Marietta, Georgia, USA
knarnava@students.kennesaw.edu

Dan Chia-Tien Lo
Kennesaw State University
Marietta, Georgia, USA
dlo2@kennesaw.edu

ABSTRACT

While existing work shows a great success, the accuracy, the response time, the interoperability, the extendibility, etc. remain to be improved. We proposed "ParlAIDialogTeacher" using the open-source ParlAI to improve the accuracy, the response time, the interoperability, and the extendibility. Training is based on Maximum Likelihood Estimation (MLE) approach for the generative model. This platform eases sharing, training, and evaluating dialog models with multiple datasets available. Our results indicate the proposed model outperforms x, y, z in terms of response time and accuracy. Future work include extending the model with other genitive models such as GAN.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; *Learning paradigms and algorithms*; *Semantic networks*; **Dialog systems**.

KEYWORDS

deep learning, chatbot, ParlAI - Blended Skill Talk

ACM Reference Format:
Kiran Mai Narnavaram and Dan Chia-Tien Lo. 2024. An Open-Source Chatbot by Using ParlAI. In *2024 ACM Southeast Conference (ACMSE 2024), April 18–20, 2024, Marietta, GA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3603287.3656160

1 INTRODUCTION AND RELATED WORK

Chatbots, advancing digital communication, face challenges in achieving human-like conversational abilities [1], with specialization enhancing efficiency but limiting versatility. Recent breakthroughs in neural approaches [2], particularly large pre-trained Transformer models, have shown promise in open-source conversational AI, as evidenced by successes in the ConvAI2 competition and subsequent improvements in dialogue system evaluations [10]. ParlAI and BlenderBot 2.0 exemplify the collaborative spirit and potential for chatbots to engage in more meaningful dialogues. Figure 1 shows that ParlAI and BlenderBot 2.0 exemplify the collaborative spirit and potential for chatbots to engage in meaningful dialogues. The development of Meena, a model with 2.6 billion parameters, marks a significant milestone in surpassing benchmarks and pushing the boundaries of chatbot capabilities [5].

ACMSE 2024, April 18–20, 2024, Marietta, GA, USA

ACM ISBN 979-8-4007-0237-2/24/04.
https://doi.org/10.1145/3603287.3656160

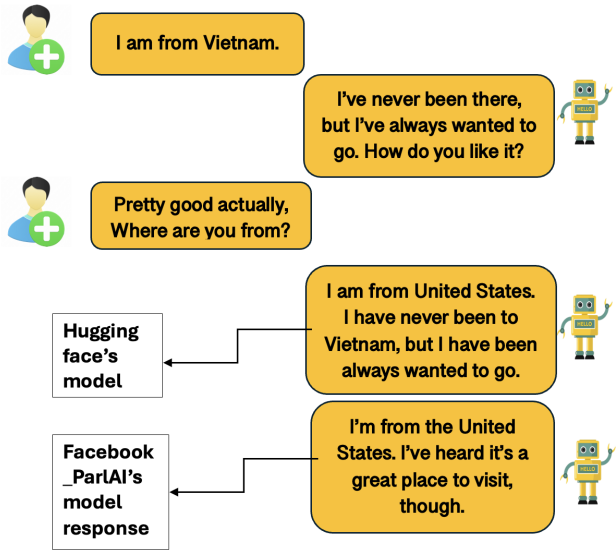

Figure 1: Chatting Example

2 METHODOLOGY

Our methodology revolutionizes chatbot development with a dynamic Seq2Seq transformer architecture in ParlAI for response generation, diverges from static databases. Employs Hugging Face Tokenizers for Byte-Level BPE, tests models with 90M to 9.4B parameters including a setup with four-layer encoder and 32-layer decoder [5]. Uses Maximum Likelihood Estimation (MLE) with an 8-layer RNN and GeLu-activated network [7], incorporates minimum length and predictive training for authentic dialogue, and enhances computational efficiency via remote server fine-tuning with ParlAI.

3 EXPERIMENTAL SETUP AND METRICS

Figure 2 illustrates our approach utilizing the Fairseq toolkit for pre-training [6], with models of 2.7B and 9.4B parameters optimized by the Adam optimizer. We incorporated Megatron-LM's model parallelism [9] for efficient node distribution, leveraging vertical transformer layer slicing for reduced cross-GPU communication. A sublinear memory footprint, the Adafactor [8] is used to facilitate larger batch size precision training.

4 PERPLEXITY

Automatic metric correlating with human judgment accelerates discourse model creation; perplexity, an uncertainty measure in Seq2Seq models. Our research demonstrates a positive correlation

Table 1: Fine Tuning Metrics

Model	timepoint1	timepoint2	timepoint3	timepoint4
Accuracy%	81.25	82.35	83.33	84.21
Loss	1.103	1.039	0.9869	0.9389
F1 %	81.25	83.82	84.72	85.52

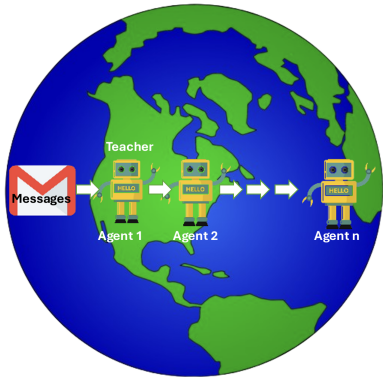

Figure 2: Chatting Model

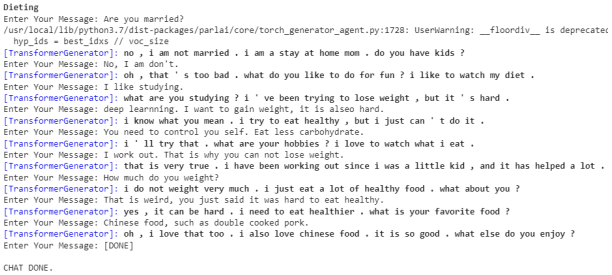

Figure 3: Controversial Results with ParlAI's Chatbot

with human evaluations, as depicted in Figure 2, highlighting that reduced perplexity correlates with increased model confidence in generating subsequent tokens.

5 PERFORMANCE EVALUATION

Evaluating chatbot performance using ACUTE-Eval [3], for bias-mitigated, pairwise dialogue evaluations; extending to "self-chat" for less human-dependent data generation, preserving human evaluation correlation [4].

6 BASELINE RESULTS

Interacting with the 90M blended skill talk model from the model zoo shows increasing accuracy in Table 1, decreasing loss, and rising F1 scores during fine-tuning, indicating promising results for the pre-trained polyencoder model fine-tuned on the Blended-Skill Talk dialogue task.

7 DESIRED GOALS AND FUTURE PROGNOSIS

The desired Goal of the project is to build higher level open-source bot which can deal with multiple tasks. The project aims to bring together three different architectures to provide the best response to get a possible question in terms of user-experience and basic knowledge of the world. This decision is illustrated by the interactions with our chatbot, as depicted in Figure 3, showcasing the performance improvements.

8 DISCUSSIONS AND LIMITATIONS

Enhanced chatbots using RNNs and a fully connected network with GeLu activation outperforms those with ReLu in diminishing both loss and perplexity. Training confined to a single epoch will limit thorough assessment through metrics such as BLEU and ROUGE, and constrain extensive evaluation. Subsequent research on multiple epochs remains to be studied with different architectures.

9 CONCLUSION

An open-source Blender bot aims for human-like interactions across multiple domains with a diverse dataset; explored chatbot implementation via specific neural networks; utilized ParlAI's comprehensive library for efficient building and testing; initial model marks project's beginning, with plans to enhance complexity and test alternative models; faced and learned from various challenges.

REFERENCES

[1] Daniel Adiwardana, Minh-Thang Luong, David R So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, et al. 2020. "Towards a Human-like Open-domain Chatbot". *arXiv preprint arXiv:2001.09977* (2020).

[2] Tianxing He, Jun Liu, Kyunghyun Cho, Myle Ott, Bing Liu, James Glass, and Fuchun Peng. 2019. "Analyzing the Forgetting Problem in The Pretrain-Finetuning of Dialogue Response Models". *arXiv preprint arXiv:1910.07117* (2019).

[3] Margaret Li, Jason Weston, and Stephen Roller. 2019. "Acute-eval: Improved Dialogue Evaluation with Optimized Questions and Multi-turn Comparisons"'. *arXiv preprint arXiv:1909.03087* (2019).

[4] Nitika Mathur, Timothy Baldwin, and Trevor Cohn. 2017. "Sequence Effects in Crowdsourced Annotations". In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2860–2865.

[5] Alexander H Miller, Will Feng, Adam Fisch, Jiasen Lu, Dhruv Batra, Antoine Bordes, Devi Parikh, and Jason Weston. 2017. "ParlAI: A Dialog Research Software Platform". *arXiv preprint arXiv:1705.06476* (2017).

[6] Myle Ott, Sergey Edunov, Alexei Baevski, and Angela Fan. 2019. "Fairseq: A Fast, Extensible Toolkit for Sequence Modeling". *arXiv preprint arXiv:1904.01038* (2019).

[7] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. "Language Models are Unsupervised Multitask Learners". *OpenAI blog* 1, 8 (2019), 9.

[8] Noam Shazeer and Mitchell Stern. 2018. "Adafactor: Adaptive Learning Rates with Sublinear Memory Cost". (2018), 4596–4604.

[9] Mohammad Shoeybi, Mostofa Patwary, Raul Puri, Patrick LeGresley, Jared Casper, and Bryan Catanzaro. 2019. "Megatron-lm: Training Multi-billion Parameter Language Models Using Model Parallelism". *arXiv preprint arXiv:1909.08053* (2019).

[10] Heung-Yeung Shum, Xiao-dong He, and Di Li. 2018. "From Eliza to XiaoIce: Challenges and Opportunities with Social Chatbots". *Frontiers of IT* 19, 1 (2018).