Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 5821–5828
Marseille, 20-25 June 2022

SPORTSINTERVIEW
A Large-Scale Sports Interview Benchmark for Entity-centric Dialogues

Hanfei Sun, Ziyuan Cao, Diyi Yang
Georgia Institute of Technology
Atlanta, GA
hsun315@gatech.edu, zcao300@gatech.edu, diyi.yang@cc.gatech.edu

Abstract

We propose a novel knowledge grounded dialogue (interview) dataset SPORTSINTERVIEW set in the domain of sports interview. Our dataset contains two types of external knowledge sources as knowledge grounding, and is rich in content, containing about 150K interview sessions and 34K distinct interviewees. Compared to existing knowledge grounded dialogue datasets, our interview dataset is larger in size, comprises natural dialogues revolving around real-world sports matches, and have more than one dimension of external knowledge linking. We performed several experiments on SPORTSINTERVIEW and found that models such as BART fine-tuned on our dataset are able to learn lots of relevant domain knowledge and generate meaningful sentences (questions or responses). However, their performance is still far from humans (by comparing to gold sentences in the dataset) and hence encourages future research utilizing SPORTSINTERVIEW.

Keywords: Entity-centric sports interview, conversation dataset, text generation, deep neural network

1. Introduction

Recent years, large-scale corpora containing human-human interaction made the modeling of human dialogues possible (Li et al., 2017). To think about it, as humans, most of our daily dialogues are grounded on external knowledge we are aware of. Hence, it's natural for previous works to focuses on incorporating external knowledge in the task of response generation (Dinan et al., 2018; Zhou et al., 2018). However, most of these datasets only contain external knowledge from one knowledge source (one dimension), which limits the diversity of knowledge sources and may contain unwanted bias. On the other hand, we humans can easily acquire knowledge from multiple sources, like from Internet search or from others' words in a conversation.

QUESTION: Just a high ankle sprain just lingered and lingered? SAMAJE PERINE: Yeah.
QUESTION: Is that one of those things that you deal with a lot? I know in high school you carried the ball, too. Are those just things that happen? SAMAJE PERINE: Yeah, when you get the ball that many times, you know it's bound to happen, it's just a matter of time. You just have to be prepared for it and know how to play through it, or when it hurts too bad, know when to tell the coach that you've had enough and that you have to sit out a couple games.

Table 1: Sample question and answers from our dataset

Putting this concern for a while, there are actually many different forms of dialogue, ranging from chatting on the Internet to group discussions. In particular, interview dialogues are a specific kind of natural dialogue consisting of interviewers and interviewees chatting structurally about certain topics (usually focusing on recently occurring events). It differs from open-domain chit-chat in that interviewers and interviewees have different responSportsInterviewilities during the conversation. Specifically, the interviewer typically brings up a topic first, which the interviewee will expand on and talk about his/her perspectives about it. Then the interviewer can choose to follow up on that topic or change the topic altogether to guide the flow of the conversation. Interview dialogues are great candidates for knowledge-grounded dialogue modeling tasks as the discussions usually resolve around some certain topic and often go much deeper, requiring sufficient domain knowledge. There have been recent efforts to collect a large-scale interview dataset, notably from (Majumder et al., 2020), where an interview dataset of National Public Radio (NPR) transcripts are collected (NPR Interview). However, there are drawbacks of this dataset in terms of knowledge grounding, as there is no profile information about each speaker. Also, on average, each speaker does not speak much (the average number of sentences spoken per speaker is about 41), which limits the amount of information in this dataset.

Besides news interviews, there are other types of interviews, and one such type is sports interview. Sports interview has its own characteristics: it usually takes place after a sports match when the interviewer asks coaches and players about their feelings about the match and their opinions about their teammates or opponents. The wide variety of personalities and backgrounds of interviewees can be well embodied in

DATASET	DOMAIN	# DIALOGUES	# UTTERANCES	# WORDS	# USERS	GROUNDING
SPORTSINTERVIEW (ours)	Sports, Media Dialogue	150 K	14.5 M	231.3 M	34 K	Wiki
SPORTSINTERVIEW-ESPN (ours)	Sports, Media Dialogue	99 K	7.1 M	105.4 M	10 K	Wiki, ESPN
Interview	Media Dialogue	106K	3.2 M	126.7 M	184 K	NPR news
Wizard of Wikipedia	Chit Chat	22 K	N/A	N/A	N/A	Wiki
PERSONA-CHAT	Chit Chat	10 K	0.16 M	N/A	1155	persona
CMU_DoG	Chit Chat, Movie	4 K	0.13 M	N/A	N/A	Wiki

Table 2: Comparison of statistics of existing dialogue datasets.

this type of interview. We thus collect our dataset from ASAP Sports[1], a website archiving sports interviews and press conference transcripts of numerous types of sports. In addition to interviews, to provide grounding facts and personalized profiles for our dataset, we additionally scraped Wikipedia articles (articles about players/coaches) and ESPN[2] news articles related to these interviews. As a comparison, in our dataset, each speaker speaks 427 sentences on average, which is far more than that compared to NPR Interview. It's worth to mention that in our dataset, we have two dimensions of external knowledge: Wikipedia articles and ESPN news reports, hence providing a much richer external knowledge linking compared to other datasets.

To sum up, our contribution are: (1) We propose a large-scale sports interview dataset: SPORTSINTERVIEW, featuring interviews of players and coaches of multiple types of sports, along with grounding documents (Wikipedia articles). We also created a subset called SPORTSINTERVIEW-ESPN which are grounded additionally by ESPN news articles. (2) We applied several generation models on SPORTSINTERVIEW-ESPN and demonstrated that while current generation models perform well, they still hallucinate on key facts and struggle with entity-centric generation.

2. Related Work

Among the dialogue datasets that involve knowledge grounding, two noticeable ones are the Wizard of Wikipedia (Dinan et al., 2019) and CMU_DoG zhou-etal-2018-dataset. In each dialogue of Wizard of Wikipedia, a "wizard" answers questions asked by an eager "apprentice" based on selected sentences in Wikipedia articles related to the subject. Therefore each dialogue is grounded on a Wikipedia article. In each dialogue of CMU_DoG, two persons discuss the content of a movie based on the Wikipedia article of the movie. Another dialogue dataset involving knowledge grounding is (Majumder et al., 2020). It is different than two previously mentioned datasets in the way that the external knowledge is linked after the collection of dialogues.

Some dialogue datasets constructed for training personalized generation models involve knowledge grounding in the form of user profiles. For example, the PERSONA-CHAT dataset (Zhang et al., 2018) was constructed by instructing crowd-source workers on Amazon Mechanical Turk to engage in conversations with each other. Each pair of Turkers were instructed to condition their dialogue on a given profile, and the goal of the conversation is to get to know each other. This resulted in a dataset consisting of crowd-sourced conversation with the speakers' persona information.

3. Dataset

Data Collection Our dataset comprises interview transcripts and complementary Wikipedia and ESPN articles. Within our dataset, the interviewees are either sports players and coaches. The Wikipedia articles are about interviewees, matches, or sports leagues (for example, the Wikipedia pages of Bo Pelini, Youngstown State vs James Medison, and NCAA Division I). The ESPN news are from the news archives of ESPN news (for example, the news archive of college football). We gather all interview transcripts from ASAP Sports, and these come from 18 different sports categories, ranging from football to auto racing. Since there are only news available for several kinds of sports, we chose to create a subset of SPORTSINTERVIEW for these kinds of sports and call this subset SPORTSINTERVIEW-ESPN. This diverse set of interview transcripts, accompanied by Wikipedia articles and ESPN news, serves as a valuable resource for building response generation models.

Processing and Data Format After the raw interview transcripts are scraped, we performed preprocessing to construct a dataset from them. We filter out all non-English interviews using `langdetect`[3]. Then, we performed sentence-level tokenization using `pySBD`[4] to separate questions from responses. After preprocessing, every interview has the following structure: a title, a date, linked Wikipedia articles, linked ESPN news if there is any, participants (interviewees), moderator's speaking at the start if there is any, and the question-response part of the interview. For the question-response part, the anonymous interviewer asks questions, and the interviewee responds to those. Also, there are two types of interviews in the dataset, one kind is interviews happening after a match, where the interviewer usually asks players and coaches about questions related to the match, and the other kind is

[1] http://www.asapsports.com

[2] https://www.espn.com/

[3] https://pypi.org/project/langdetect/

[4] https://github.com/nipunsadvilkar/pySBD

called conference, where more general and non-match-specific questions tend to be asked. Note that in our dataset, responses are usually much longer than questions. In summary, there are 150,204 distinct interviews, 1,707,837 questions asked (with 1,890,652 distinctive responses) and 34,180 distinctive interviewees (sports players and coaches) in SPORTSINTERVIEW.

Wikipedia Pages We further scraped Wikipedia pages using `Wikipedia-API`[5] as additional grounding contexts for our interview dataset. We mainly collected three kinds of Wikipedia pages: interviewees, matches, or sports leagues. By analyzing SPORTSINTERVIEW, we found that the interviews are from 3439 distinct events including after-game interview and conference. Among them, 1548 out of the 3439 events are about a specific section of games and 1540 out of the 1548 games have corresponding Wikipedia articles. The other 1891 events are not about a specific section of games. These include 711 media conferences, 96 media opens, 53 announcements, etc. In summary, out of all 150,204 interviews in SPORTSINTERVIEW, 149,363 are linked to at least one Wikipedia page (99.44%).

SPORTSINTERVIEW-ESPN ESPN news also contain detailed information about the games they are reporting. Hence, we scraped these news and linked them to some of our interviews. These include five types of sports: football, basketball, baseball, golf and hockey. We selected these as a subset and called it SPORTSINTERVIEW-ESPN. Additionally, we filtered out all interviews which are not linked to any Wikipedia article, and all responses from interviewees who are not linked to any Wikipedia article in SPORTSINTERVIEW-ESPN. Thus, SPORTSINTERVIEW-ESPN can be regarded as a more knowledge grounded subset of SPORTSINTERVIEW. There are 99,812 distinct interviews in SPORTSINTERVIEW-ESPN and about 33% of interviews are linked to at least one ESPN news.

Comparison with Other Datasets Table 2 contains the comparison of our dataset with several other personalized dialogue datasets. Our dataset is on the larger side in terms of number of utterances. It is also the only large-scale dataset within both sports and media dialogue categories. For SPORTSINTERVIEW-ESPN, it's the only dataset backed up with two different kinds of knowledge sources (Wikipedia and ESPN news). Also, to the best of our knowledge, our dataset is the most comprehensive dialogue dataset in the sports domain. Hence, it can be used as a valuable resource for conducting research on sports interviews by its own virtue.

4. Methods and Results

To investigate the entity-related knowledge embedded in our dataset and the potential usage of our dataset, we designed two tasks and tested the performance of several models on both of the tasks.

Interview question generation We formulate the first experiment as interview question generation. The purpose of this experiment is to examine if we can train a model that can ask interview questions to a given player or coach if the model has some background information of the interviewee and the current game. Depending on the specific interview, the background information can include the Wikipedia article of the game (G), the Wikipedia article of the section (S), the ESPN news (E), the Wikipedia article of the interviewee (I), and the previous response in the interview (P). This task is formulated as

$$\arg\max_{Q} \mathbb{P}(Q \mid G, S, E, I, P)$$

One natural choice of baseline models for this task is to use pretrained language model. We use GPT-2 (Radford et al., 2019) in our experiments. Since the background information is usually very long, we first summarized the available background information using Huggingface's summarization pipeline powered by a pretrained distilled BART[6](Lewis et al., 2020). For each player or coach that the model will ask question to, we design the following prompt, "*As a reporter, you ask X*", where X is the name of the player or coach. We provide the summarized background information and the prompt to GPT-2 and collect the truncated output of the model as the interview question. Baseline results are summarized in Table 3.

To investigate how our dataset can be used to train a question generation model, we fine-tuned BART for this task. As in the case of GPT-2, for each interviewee in each interview, we summarized the corresponding background information. We fed the summarized background into the encoder of the BART and minimized the negative log-likelihood of the interview question. We tried different combinations of background. The results are shown in Table 3

Interview response generation In addition to generating interview questions, we also experimented on generating appropriate responses to interview questions. In this case, we employed Seq2Seq, Speaker Model, DialoGPT (Zhang et al., 2020), as well as BART as our models. We experimented on all these models and for BART in particular, we tried to feed different inputs to it. All models are trained/fine-tuned on the training set[7] of SPORTSINTERVIEW-ESPN and model performances are evaluated on the test set. Our evaluation metrics include BLEU score (Post, 2018), perplexity, BERTScore (Zhang et al., 2019) and distinct-n (Li et al., 2015). Our results are summarized in Table 4. We observe that state-of-the-art models

[5] `https://pypi.org/project/Wikipedia-API/`

[6] `https://huggingface.co/docs/transformers/master/en/main_classes/pipelines#transformers.SummarizationPipeline`

[7] Random split with ratio 0.98:0.1:0.1

Model	Perplexity	BLEU	BERTScore	Distinct-1	Distinct-2	Avg. Length
GPT-2 (not fine-tuned)	45.22	0.3	0.804	0.37	0.69	27
BART	**17.93**	**0.7**	**0.848**	**0.42**	**0.76**	15

Table 3: Question generation results on test set of SportsInterview-ESPN

Model	**Perplexity**	**BLEU**	**BERTScore**	**Distinct-1**	**Distinct-2**	**Avg. Length**
Seq2Seq	29.68	1.0	0.82	0.07	0.14	65
Speaker	28.22	1.0	0.82	0.11	0.21	59
DialoGPT	16.54	0.5	0.82	0.45	0.62	21
BART	**14.36**	**1.2**	0.85	0.34	0.62	33
BART with previous QA	15.42	0.5	0.85	**0.51**	**0.68**	23
BART with Wiki	16.27	0.5	0.85	0.44	**0.68**	25

Table 4: Response generation results on test set of SportsInterview-ESPN

can generally achieve a low perplexity score after fine-tuning and by introducing additional knowledge like Wiki, the generated response are more diverse (distinct-n scores are higher). Some generated results are included in Appendix 11.

Interview transcripts are inherently knowledge-rich We believe that even our interview transcripts themselves contain numerous information which models can learned from. To test this, we fine-tuned BART on the interview transcripts only, without Wikipedia and ESPN news, and queried this fine-tuned model. One question we asked the model is "What sports does this player play?", and we asked this question to the model for a total of 17k times (each time we ask about a different interviewee, and we only ask the model about interviewees who have at least reasonable occurrences in the dataset). As a result, the model answers approximately 26% of all questions asked correctly. We also asked the same question to the same BART model, but without fine-tuning it on our dataset and this time, the BART model could hardly guess who these interviewees are. This proves that our interview transcripts alone contain numerous knowledge in the sports domain, not to mention that we have linked Wikipedia articles (and ESPN news in the case of SportsInterview-ESPN).

5. Conclusion

In this work, we present large-scale sports interview datasets SportsInterview and SportsInterview-ESPN for knowledge-based generation and dialogue modeling. SportsInterview and its variant are large in size, and comprise of interviews coming from a broad range of sports with external knowledge linking. By training and testing several models on SportsInterview-ESPN, we showed that it is a challenging dataset and can be a valuable resource in conducting further research on knowledge-based generation models. Potential future directions include improving knowledge selection procedures and modeling sports interview dynamics.

6. Ethical Considerations

The corpus being used in this work includes interview transcripts from ASAP Sports, Wikipedia articles, and ESPN news. Hence all these information are copyrighted by their affiliating corporations or organizations. In terms of privacy, all of the data sources present in our corpus are already public, yet we still make sure that our datasets SportsInterview and SportsInterview-ESPN will be shared for academic use only. We also open source all the code used for scraping and processing the aforementioned data sources.

7. References

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.*

Dinan, E., Roller, S., Shuster, K., Fan, A., Auli, M., and Weston, J. (2019). Wizard of Wikipedia: Knowledge-powered conversational agents. In *Proceedings of the International Conference on Learning Representations (ICLR).*

Holtzman, A., Buys, J., Du, L., Forbes, M., and Choi, Y. (2019). The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751.*

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. (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, Online, July. Association for Computational Linguistics.

Li, J., Galley, M., Brockett, C., Gao, J., and Dolan, B. (2015). A diversity-promoting objective function for neural conversation models. *arXiv preprint arXiv:1510.03055.*

Li, Y., Su, H., Shen, X., Li, W., Cao, Z., and Niu, S. (2017). DailyDialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of the Eighth International Joint Conference on Natural Language*

Processing (Volume 1: Long Papers), pages 986–995, Taipei, Taiwan, November. Asian Federation of Natural Language Processing.

Majumder, B. P., Li, S., Ni, J., and McAuley, J. (2020). Interview: Large-scale modeling of media dialog with discourse patterns and knowledge grounding. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8129–8141.

Post, M. (2018). A call for clarity in reporting bleu scores. *arXiv preprint arXiv:1804.08771*.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

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)*, pages 2204–2213, Melbourne, Australia, July. Association for Computational Linguistics.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y. (2019). Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.

Zhang, Y., Sun, S., Galley, M., Chen, Y.-C., Brockett, C., Gao, X., Gao, J., Liu, J., and Dolan, B. (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, Online, July. Association for Computational Linguistics.

Zhou, K., Prabhumoye, S., and Black, A. W. (2018). A dataset for document grounded conversations. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 708–713, Brussels, Belgium, October-November. Association for Computational Linguistics.

Appendix

8. Statistics of *SportsInterview*

8.1. Statistics

The per-sports-category statistics is shown in Table 5. We also made a histogram showing the distribution of number of questions asked in each interview (see Figure 1). We observe that the majority of interviews contain from 4 to 15 interviews. We also measured the number of utterances (an utterance is a single sentence) every interviewee speaks, as we want to identify how many interviewees have spoken a significant amount of words. The result is in Figure 2. We observe that about 10,000 interviewees have spoken more than 100 utterances in our dataset, hence indicating that these interviewees' responses can provide a rich amount of training data.

Sports Type	# Interviews	# Interviewees	Avg # Questions
Golf	66513	6719	9.54
Tennis	29830	2596	12.86
Basketball	19671	10155	11.24
Football	11444	4686	18.42
Baseball	9367	3988	11.11
Auto Racing	5892	2196	12.93
Hockey	4707	1832	10.13
Cricket	600	242	11.41
Track & Field	499	483	15.6
Equestrian	396	742	8.4
Soccer	359	180	8.39
Wrestling	305	190	8.32
Swimming	227	88	8.63
Volleyball	199	242	10.46
Lacrosse	102	249	11.89
CoSIDA	60	171	0.68
Boxing	31	112	43.48
Extreme	2	8	11.5

Table 5: Statistics of different sport types

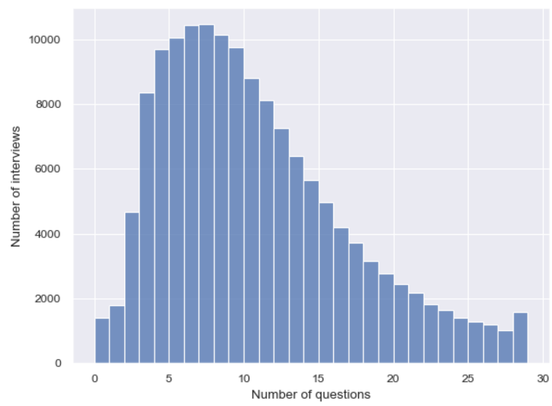

Figure 1: Distribution of question counts

8.2. Response Length

In a typical interview setting, the length of responses are generally much longer than that of the questions. Our dataset is not an exception to this rule. Hence, to better understand our corpus, we plot the distribution of length of responses, as shown in Figure 3. We observe that most responses are of length shorter than 150, thus making the training of generation models fea-SportsInterviewle (but still in the domain of long-form generation). Also, the average length of responses in SportsInterview is 80.53 words.

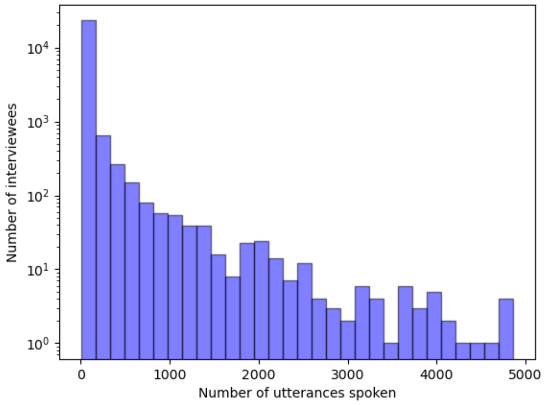

Figure 2: Distribution of utterance counts

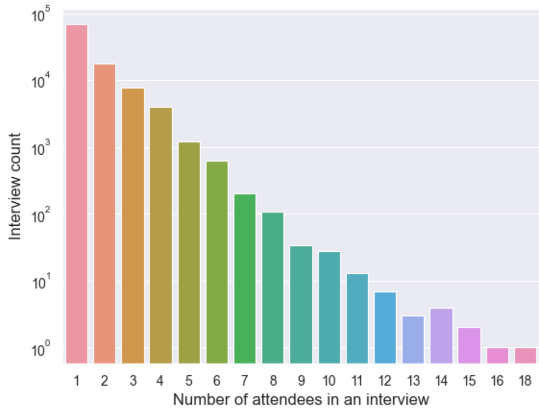

Figure 4: Distribution of number of attendees for interviews

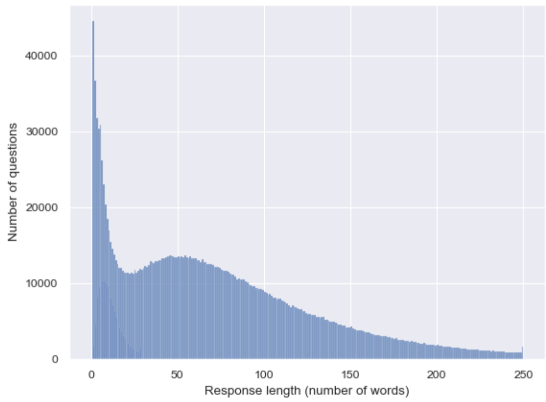

Figure 3: Distribution of response length

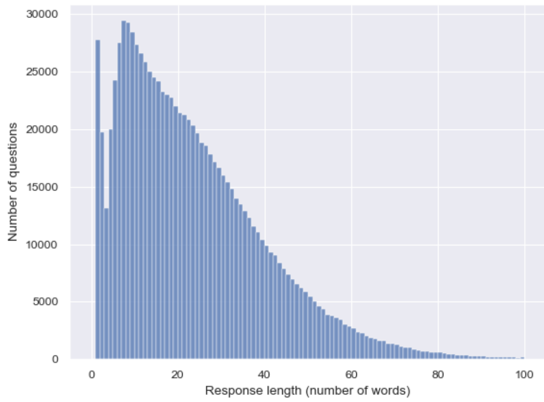

Figure 5: Distribution of input length of BART

9. Statistics about some experiments

We plotted some distributions regarding the model input length of some models, namely: BART (Figure 5) and BART with previous question and response (Figure 6).

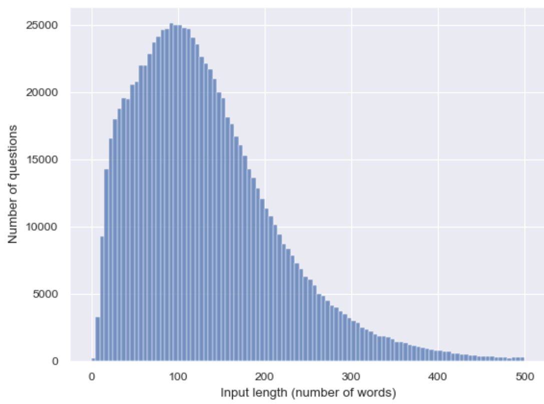

Figure 6: Distribution of input length of BART with previous question and response

10. Full experiment results

See Table 6 for the detailed version of our experiment results.

11. Sample generated responses

Here are some generated responses from our models using sampling based decoding (see Table 7, Table 8 and Table 9). In Table 9, the predictions of BART and BAR W/BACKGROUND mentioned *birdie*, a terminology in golf, when describing the performance of the player Henrik. We suspect that the model gets the cue from the name Henrik, a name of a golf player. This is because the model remembers the name Henrik from the training set and knows who he is (by querying

MODEL	PERPLEXITY	BLEU	BERTSCORE	DISTINCT-1	DISTINCT-2	AVG. LENGTH
Seq2Seq (5 ep)	29.68	1.0	0.82	0.07	0.14	65
Speaker (5 ep)	28.22	1.0	0.82	0.11	0.21	59
DialoGPT (1 ep, 0.9 top-p)	16.54	0.5	0.82	0.45	0.62	21
BART						
1 ep	15.10	1.9	0.85	0.21	0.51	45
5 ep	14.36	2.1	0.86	0.22	0.51	47
5 ep, 10 bm	14.36	0.5	0.85	0.39	0.68	24
5 ep, 0.9 top-p	14.36	1.2	0.85	0.34	0.62	33
BART with background						
5 ep	15.75	1.7	0.86	0.33	0.54	41
5 ep, 10 bm	15.69	0.2	0.85	0.53	0.72	18
5 ep, 0.9 top-p	15.75	0.7	0.86	0.45	0.64	27
BART with previous QA						
5 ep	15.42	1.4	0.86	0.37	0.57	36
5 ep, 10 bm	15.42	0.1	0.85	0.58	0.73	15
5 ep, 0.9 top-p	15.42	0.5	0.85	0.51	0.68	23
BART with Wiki						
1 ep	16.27	1.4	0.85	0.28	0.55	38
1 ep, 10 bm	16.27	0.0	0.85	0.54	0.76	14
1 ep, 0.9 top-p	16.27	0.5	0.85	0.44	0.68	25

Table 6: Performance of different models. If unspecified, greedy decoding is used. ep: epoch, bm: number of beams, top-p: nucleus sampling probability (Holtzman et al., 2019)

the raw pre-trained BART model who Henrik is, the model simply says it does not know).

When comparing generated responses across different models, we observe that all models generate responses using fluent English and conveys opinions meaningfully. But inspecting more carefully, we see that BART seems to be able to provide the most concise and clear meaning. For example, in Table 7, BART is the only model that gives a decisive answer (by choosing Moseley) instead of giving ambivalent ones. In table 9, BART uses the word "but", which correctly conveys the transitive meaning that although Henrik is hard to beat, the player will still try his best to win the match. The exact same meaning is also conveyed by the gold response. This result corresponds with the fact that the perplexity of BART is the lowest among these models, as the generated responses of BART have the closest meaning with the gold responses.

QUESTION (BASEBALL): After a game like that can you just kind of update us on your pitching situation, who you could maybe start tomorrow and would Howard or Moseley be available tomorrow?
BART: Yeah, Moseley will start tomorrow.
BART W/ BACKGROUND: We'll see tomorrow. We'll see how it goes. I mean, we've got a couple guys that are going to be available, but we'll see what happens.
BART W/ PREV. QR:Yeah, I mean, we'll see. We'll see how it goes.
BART W/ WIKI: I don't know. I'm not going to get into that right now.
GOLD: Absolutely. That would be the answer there as far as those two would be available. It would be really tough to bring Dugger back I would say, but everybody else I'd say all hands on deck.

Table 7: Sample generated responses of baseball interviews.

QUESTION (FOOTBALL): How badly does your team want this?
BART: I think they want it very badly. You know, it's a great opportunity for us. I mean, we're going to go out there and play hard and try to win the game. I think that's the most important thing.
BART W/ BACKGROUND: I think they want it very badly. I think they're excited. They want to win a Stanley Cup. They know what it takes to get there. They're excited to be here.
BART W/ PREV. QR:I don't know if they want it too bad. I think they want to win every game they play. I don't think there's any question about it.
BART W/ WIKI: I don't know. I mean, you know, I don't think it matters. I think it's important for our team. It's important to us. We want to win. We don't want to lose. That's why we're here. We're here to win, and we're going to do everything we can to win the game.
GOLD: We want every game really bad. We look at every game like a Super Bowl. That's how we got ready for this game. When we were in Eugene two weeks ago, our mentality for each practice was as if we were getting ready for a game in three or four days. I think that's what made our practices the way they have been. We've been having some pretty good practices. Really excited about this game.

Table 8: Sample generated responses of football interviews.

QUESTION (GOLF): And in context of the race, Henrik is proving a fair old frontrunner, isn't he. You saw him up close today.
BART: Yeah, he's a great player. I think he's got a lot of game, and it's going to be tough to catch him. But I think you've just got to go out there and play your own game and try to make as many *birdies* as you can and see what happens.
BART W/ BACKGROUND: Yeah, he's a great player. He's been playing well for a while now, and he's going to be tough to catch. I'm just going to go out and play my own game and try to make as many *birdies* as I can and see what happens.
BART W/ PREV. QR:Yeah, he's playing great. He's a great player, and he's got a lot of game. So yeah, it's going to be a tough match. I'm just going to try to play as good as I can and see what happens.
BART W/ WIKI: Yeah, he's been playing really well. He's a really good player. I think he's going to be in contention for a long time.
GOLD: He seems to be getting off to strong starts in all his tournaments recently. He's got his name on the board. He's making it hard for the rest of us. But I knew that could be the case. I didn't that's why I wanted it to be in my own hands this week, if I won a golf tournament which is obviously what we try to do every week, and then it would be in my hands and that's still the case. My goal remains the same for the rest of the week.

Table 9: Sample generated responses of golf interviews.