ProphetChat: Enhancing Dialogue Generation with Simulation of Future Conversation

Liu et al., ACL 2022

2022.06.10
Yu-Hung Wu @ IIS, Academia Sinica

https://aclanthology.org/2022.acl-long.68/

Outline

- Previous Model
- Proposed Model
- Results & Studies
- Conclusion

Previous Model

Dialogue Generation in Open-domain Conversations

- Recent dialogue systems usually utilize the dialogue history to generate the response, i.e. estimate the probability of $P(response|history)$.
- There are some common problems in such systems:

1. Generating bland responses (e.g., yeah, ha ha) is the most critical problem.
2. It also poses a greater **one-to-many** problem than is typical in other text generation tasks

DialoGPT (Zhang et al., 2020)

- DialoGPT is formulated as an autoregressive language model, and uses the multi-layer transformer as model architecture.
- The model is trained on large-scale dialogue pairs/sessions extracted from Reddit discussion chains.
- DialoGPT successfully captures the joint distribution of $P(target|source)$ in conversational flow, which achieves SOTA results in 2020.

DialoGPT (Zhang et al., 2020).

- First, concatenate all dialog turns within a dialogue session into a long text $S = x_1, x_2, \ldots, x_m$ (m is the sequence length), ended by the `<end-of-text>` token.
- The conditional probability of $P(T|S)$ can be written as the product of a series of conditional probabilities:

$$p(T|S) = \prod_{n=m+1}^{N} p(x_n|x_1, \cdots, x_{n-1}) \quad (1)$$

where m is the length of the history sequence.

Maximum Mutual Information (MMI)

- MMI is a solution the problem that the response was sometimes too bland or uninformative.
- Given a dialogue in the training set:
  - A: How old are you?
  - B: I'm 22 years old.
  - A: Oh, that's terrible.
- To train the target model, we use the following as input sequence:

Maximum Mutual Information (MMI).

- Now, we train an additional "MMI model". MMI is the backward version of the original model, it's goal is to *predict source sentences from given responses*..
  - Original training objective: maximize $P(Target|Source)$, MMI training objective: $P(Source|Target)$.
- Thus, the input is the **reversed** sentence of the original model:
  - `[CLS]Oh, that's terrible.[SEP]I'm 22 years old.[SEP]How old are you?[SEP]`

How to Use the MMI Model?

1. Use the target model to generate a set of responses (use different sampling method to generate diverse responses, such as top-k), we denote it as S.
    - $S = [r_1, r_2, \ldots, r_n]$, where n is the number of candidate responses.
2. for each response in S, use the MMI model to calculate it's perplexity score. The input of the MMI model is backwarded.
3. Choose the one with least perplexity, this is the final response.

Proposed Model

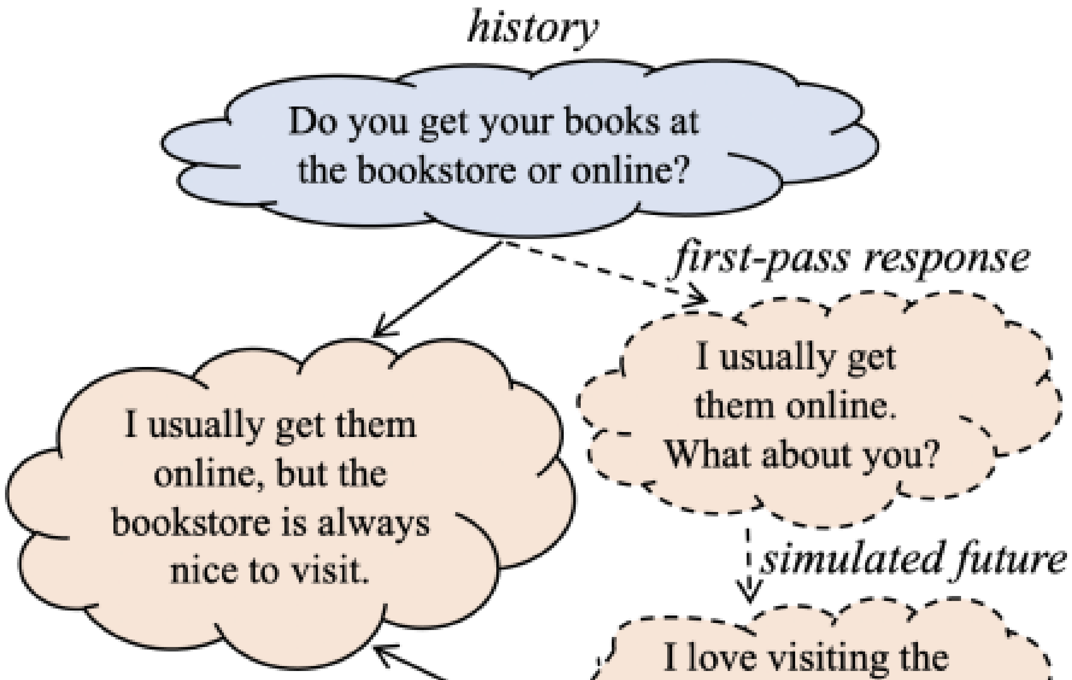
Motivation

- Since there exists a ***one-to-many*** relationship in dialogue generation, generating a desired response solely based on the historical information is not easy.
- If the chatbot can foresee in advance what the user would talk about after receiving its response, it could possibly provide a more informative response.
- Previous models usually consider $P(response|history)$, the proposed model considers $P(response|history, future)$.

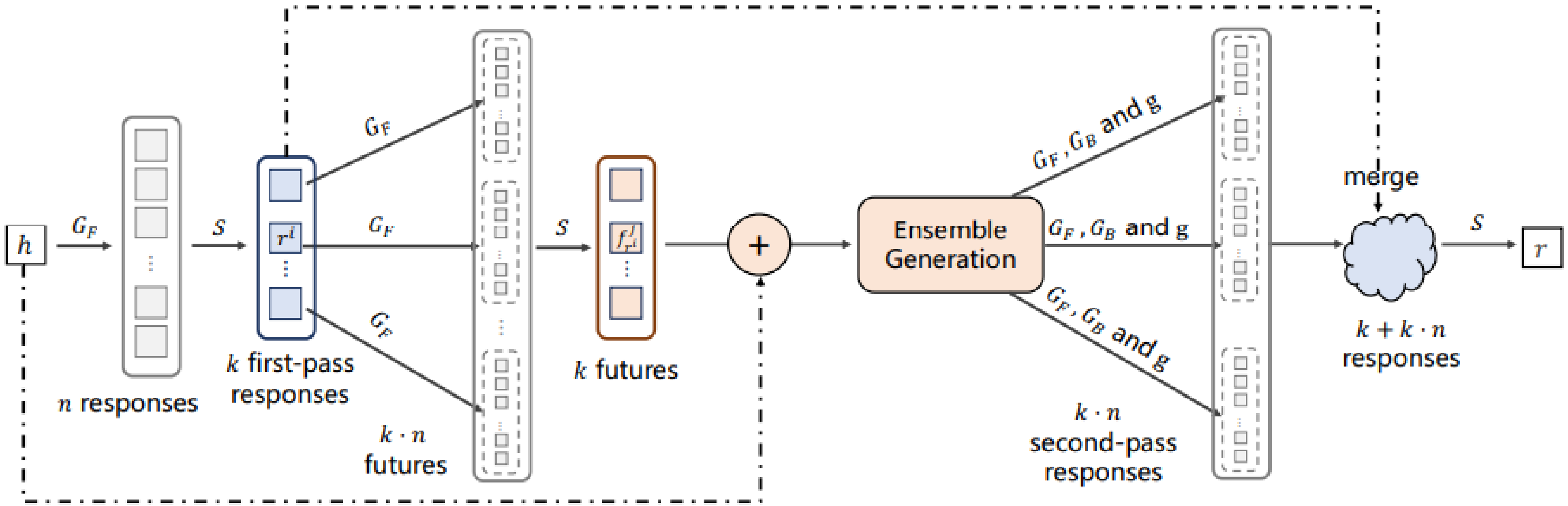
Motivation.

- Some PLMs such as BERT use bi-directional information instead of just considering the tokens on its left side, and have bring significant improvement.
- The model thus uses the "right side information" (i.e. dialogue future) to enhance the generated response.

Example


history
Do you get your books at the bookstore or online?
first-pass response
I usually get them online, but the bookstore is always nice to visit.
I usually get them online. What about you?
simulated future
I love visiting the

Overall Framework



Generating First-pass Responses

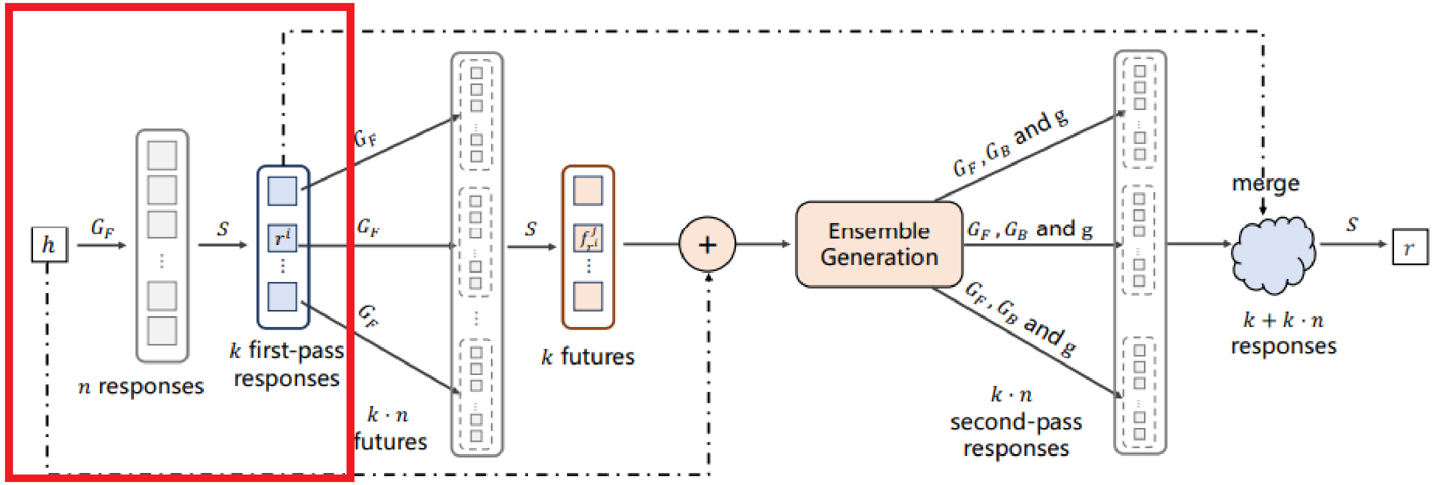


Figure 2: The overall framework of ProphetChat.

- Given a dialogue history h, first use G_F (a forward generator, you

Generating Futures

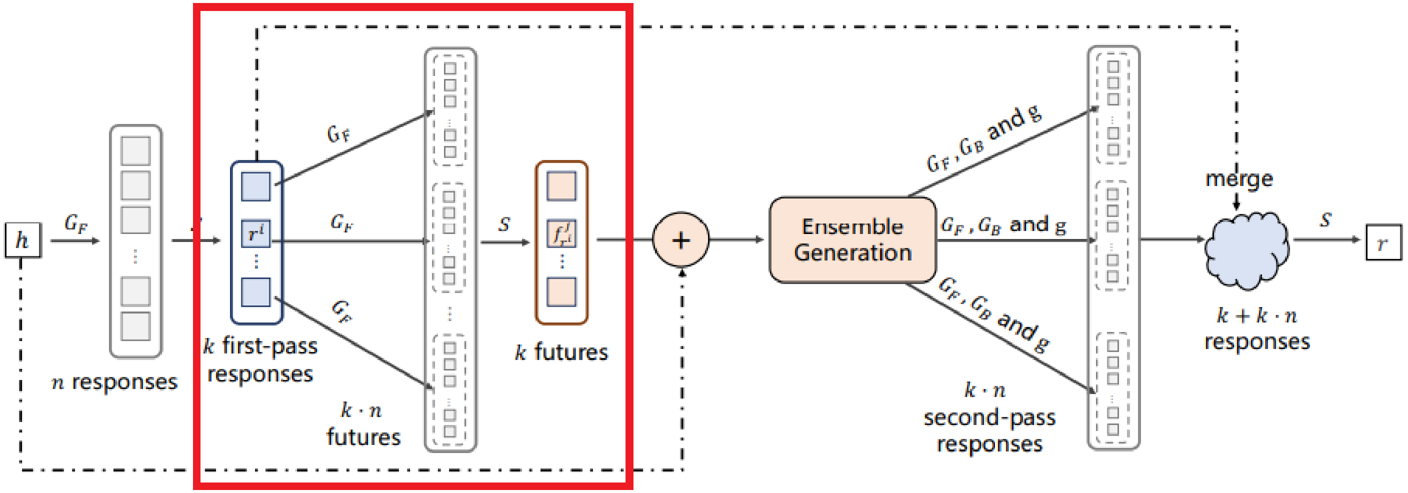


Figure 2: The overall framework of ProphetChat.

- Now, we have k first-pass responses. We use G_F (**exactly the**

Generating Futures.

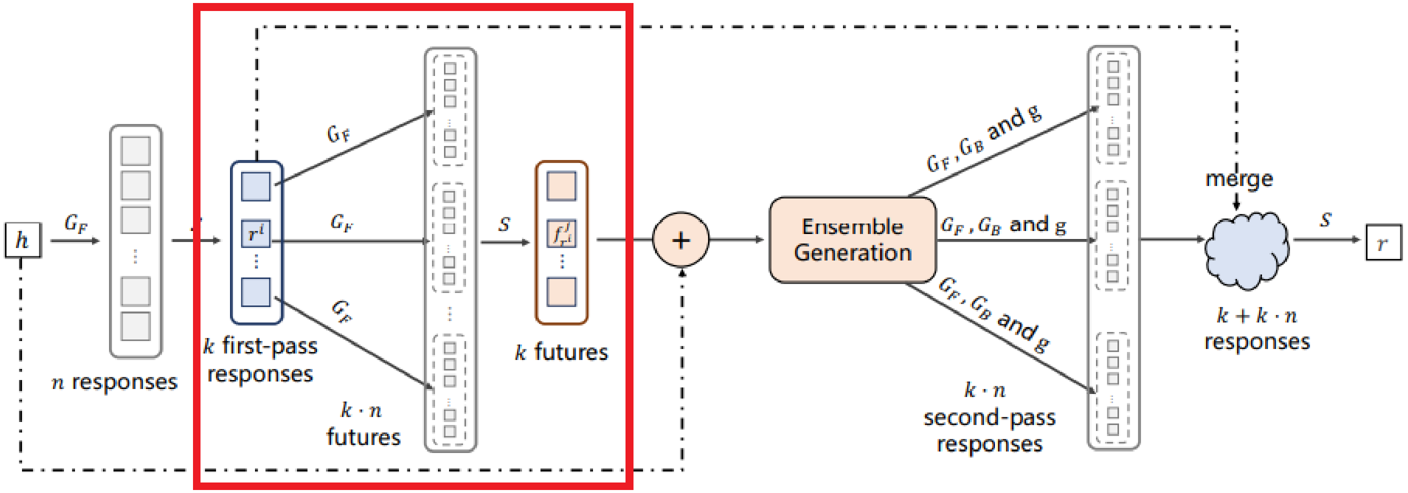


Figure 2: The overall framework of ProphetChat.

- Considering that the responses are not equal in quality, we

Ensemble Generation

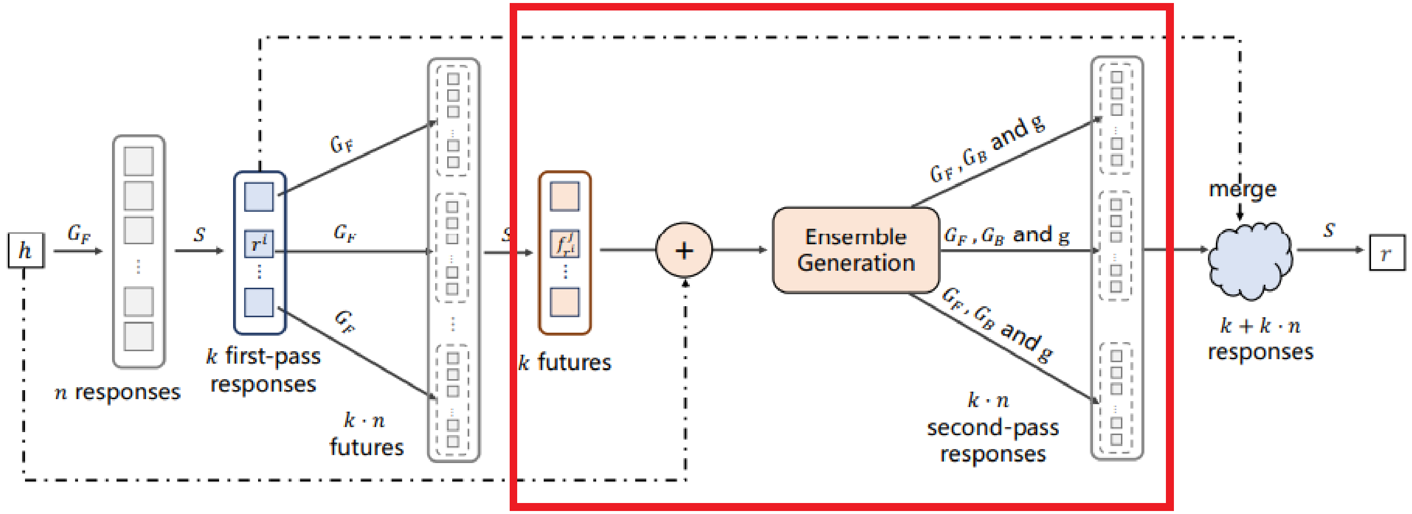


Figure 2: The overall framework of ProphetChat.

- The goal of ensemble generation is to generate to second-pass

Ensemble Generation.

- The response r using the per-step weighted ensemble of G_F and G_B conditioned on h and f:

$$P(\hat{r}_t|h, f, \hat{r}_{<t}; \theta_F, \theta_B, \theta_g) = w \cdot P(\hat{r}_t|h, \hat{r}_{<t}; \theta_F) + (1 - w) \cdot P(\hat{r}_t|f, \hat{r}_{<t}; \theta_B), \quad (1)$$

- w is a ***learnable*** weight:
  - Using a ***trainable*** gate g which takes the last hidden states from G_F and G_B as inputs and calculates an ensemble

Final Response Generation

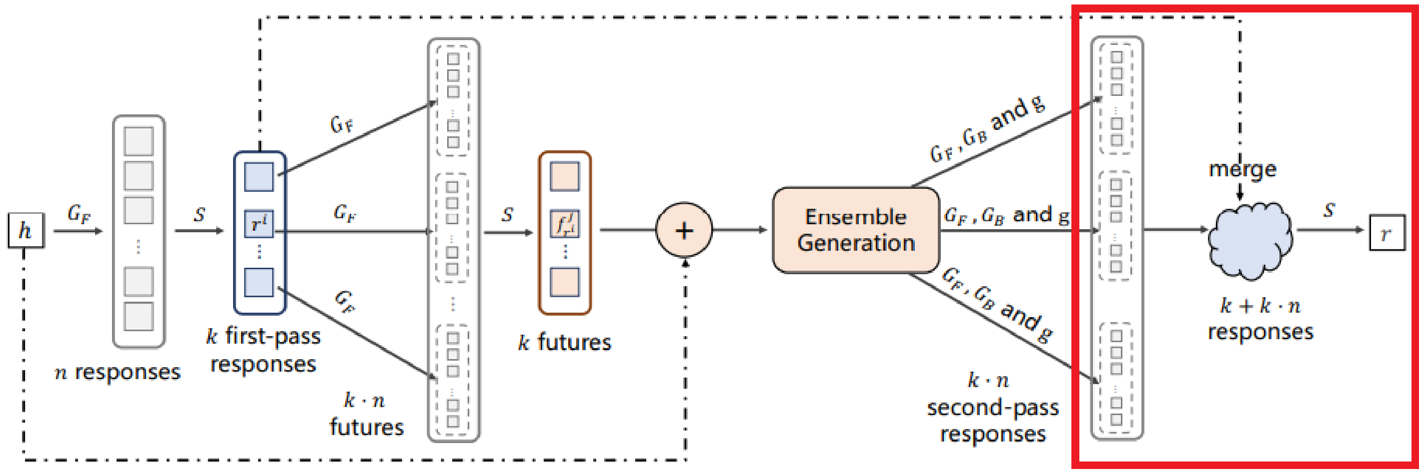


Figure 2: The overall framework of ProphetChat.

- To make full use of the k-best first-pass responses, we finally re-

Framework Summary

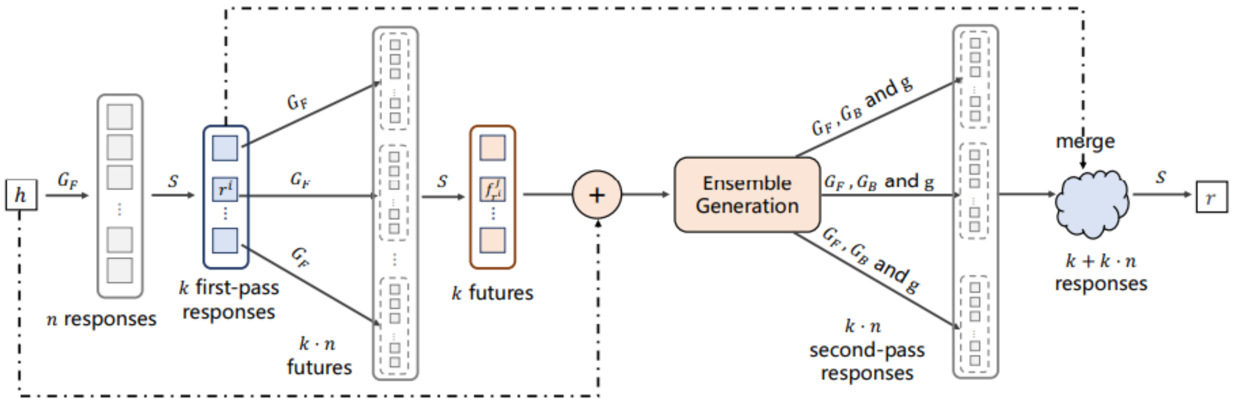


Figure 2: The overall framework of ProphetChat.

- There are 4 components in the framework:
  1. G_F: the forward model.
  2. G_B: the backward model.
  3. g: the gate used to compute the weight between G_F and G_B.
  4. S: the selector to evaluate the utterances.
- Generate responses → select k → generate futures → select k → generate more responses → select 1

Training of G_F and G_B

- The authors fine-tuned G_F and G_B as the same method as DialoGPT.
- G_F stands for the forward model and G_B stands for the backward (MMI) model.
- The two components are fine-tuned before training g and S.

Training of g

- Directly minimize the NLL of the gold response r^*:

$$\mathcal{L}_1(\theta_g) = -\sum_t \log P(r_t^* | h, f, r_{<t}^*; \theta_F, \theta_B, \theta_g), \tag{2}$$

- f (future) can be simulated from:

1. The gold response (teacher-forcing mode)
2. A sampled first-pass response (free-tunning mode)

- The paper creates a curriculum schedule that gradually switches

Training of S

- The goal is to directly optimize S to our ultimate goal which is to maximize the log-likelihood of the gold response.
- The paper adopts reinforcement learning to optimize S.
- The quality scores of first-hand responses (or futures) can naturally form a score distribution, then we can sample a response (or future) from them.
- Then, we can feed the sampled future and history into the ensemble generation model and calculate the log-likelihood of

Training of S.

- The self-critic baseline: select the response/future with the highest scores in each sampling step, and calculate the reward as R_b.
- The gradient can be estimated as:

$$\nabla_{\theta_S}\mathcal{L}_2(\theta_S) \approx -(R - R_b)\nabla_{\theta_S}[\log P(r^i|h;\theta_S) + \log P(f^j_{r^i}|r^i, h;\theta_S)]. \quad (3)$$

Training Summary

1. Use the same training objective to train G_F and G_B.
2. Traing g.
3. Train S.
4. Jointly tune g and S.

Results & Studies

Datasets

- DailyDialog: A dataset that contains our daily dialogues.
- PersonaChat: A multi-turn dataset that also consider the speaker's persona.
- For both datasets, the paper treats each consecutive three utterances as a triplet that represents history-response-future.

Variants and Ablations of ProphetChat

- $ProphetChat_{k=?}$: the model with the same model parameters but different beam sizes when simulating the futures.
- $ProphetChat_{first}$: the model only uses the first-pass responses in the final re-ranking process.
- ProphetChat w/o history: only use the futures to generate the response. (only use MMI)
- ProphetChat w/o selector: sample the responses and the futures randomly

Results

- The results of DailyDialog dataset.

Models	B-1	B-2	B-3	B-4	D-1	D-2	D-3	D-4	AVG	EXT	GRE
Posterior-GAN	37.65	14.25	4.90	1.66	0.91	5.13	13.21	22.23	0.530	0.472	0.313
RegDG	38.77	14.36	5.13	1.91	1.07	5.95	14.85	24.78	0.550	**0.493**	0.319
DialoGPT$_F$	34.63	12.89	4.81	1.75	5.19	29.00	55.09	73.17	0.623	0.468	0.370
DialoGPT$_{F,rerank}$	34.66	12.99	4.87	1.77	6.79	36.59	64.75	81.18	0.612	0.456	0.369
ProphetChat	**39.33**	**14.57**	**5.38**	**1.93**	6.53	35.93	64.18	80.66	**0.626**	0.470	**0.372**
ProphetChat$_{first}$	37.58	14.00	5.22	1.89	6.49	35.44	63.18	79.73	0.625	0.468	0.369
ProphetChat$_{second}$	39.10	14.55	5.41	1.96	6.47	36.15	65.14	**81.92**	0.616	0.465	0.366
ProphetChat$_{k=1}$	35.27	13.17	4.92	1.79	**6.80**	**37.07**	**65.25**	81.39	0.612	0.464	0.368
ProphetChat$_{k=2}$	36.43	13.57	5.06	1.84	6.70	36.80	65.01	81.45	0.618	0.466	0.370
ProphetChat$_{k=3}$	37.71	14.04	5.22	1.89	6.59	36.27	64.68	81.24	0.622	0.468	0.370
ProphetChat *w/o* history	32.45	11.51	4.03	1.39	5.27	30.07	56.94	74.35	0.601	0.442	0.347
ProphetChat *w/o* selector	35.74	13.08	4.75	1.68	5.07	28.53	54.98	73.12	0.623	0.464	0.364
ProphetChat *w/o* train	38.87	14.06	5.20	1.88	6.31	35.33	63.01	79.52	0.623	0.465	0.367

Results.

- The results of DailyDialog dataset.

Models	B-1	B-2	B-3	B-4	D-1	D-2	D-3	D-4	AVG	EXT	GRE
Posterior-GAN	44.13	16.57	5.73	1.91	0.41	2.18	5.34	9.91	0.647	0.489	0.380
RegDG	46.12	17.11	5.90	2.01	0.43	2.41	6.26	11.55	0.653	**0.512**	0.381
DialoGPT$_F$	45.84	16.91	6.07	2.12	2.26	14.89	32.28	48.35	0.657	0.480	0.383
DialoGPT$_{F,rerank}$	46.69	17.18	6.13	2.13	2.85	19.40	41.96	61.69	0.657	0.481	0.386
ProphetChat	47.55	**17.50**	**6.26**	**2.19**	3.01	20.01	42.32	61.58	**0.662**	0.484	**0.393**
ProphetChat$_{first}$	47.51	17.47	6.23	2.17	2.86	19.15	40.99	60.46	0.660	0.483	0.390
ProphetChat$_{second}$	46.43	17.03	6.05	2.10	**3.06**	**20.81**	**43.90**	**63.67**	0.661	0.484	0.391
ProphetChat$_{k=1}$	46.44	17.08	6.10	2.12	3.01	20.34	43.36	63.12	0.659	0.482	0.390
ProphetChat$_{k=2}$	46.92	17.20	6.15	2.14	3.05	20.18	42.91	62.56	0.659	0.483	0.391
ProphetChat$_{k=5}$	**47.66**	17.49	6.12	2.15	3.00	19.87	41.95	61.14	0.658	0.482	0.388
ProphetChat *w/o* history	42.47	15.23	5.30	1.81	2.42	15.84	34.74	52.70	0.637	0.461	0.369
ProphetChat *w/o* selector	46.38	16.98	6.05	2.11	2.30	15.44	34.35	52.32	0.656	0.477	0.382
ProphetChat *w/o* train	47.44	17.23	6.16	2.14	2.91	19.58	41.69	60.80	0.659	0.480	0.390

Results..

- Human evaluation results.

Models	Readability	kappa	Sensibleness	kappa	Specificity	kappa
Posterior-GAN	0.58	0.42	0.46	0.49	0.21	0.58
RegDG	0.60	0.45	0.51	0.59	0.27	0.50
DialoGPT$_F$	0.68	0.52	0.64	0.61	0.44	0.52
DialoGPT$_{F,rerank}$	0.69	0.50	0.69	0.60	0.45	0.64
ProphetChat	**0.71**	0.52	**0.75**	0.53	**0.49**	0.49

Models	Readability	kappa	Sensibleness	kappa	Specificity	kappa
Posterior-GAN	0.64	0.58	0.50	0.65	0.24	0.48
RegDG	0.65	0.56	0.53	0.55	0.28	0.63
DialoGPT$_F$	0.69	0.59	0.68	0.66	0.42	0.52
DialoGPT$_{F,rerank}$	0.70	0.44	0.72	0.52	0.48	0.52
ProphetChat	**0.72**	0.43	**0.77**	0.61	**0.53**	0.54

Results...

- TF_{rerank}: the model uses the history and gold response to generate the future.
- $FR_{k=?}$: origin method with different beam size.

Models	B-1	B-2	B-3	B-4	D-1	D-2	D-3	D-4	AVG	EXT	GRE
TF_{rerank}	**39.94**	**14.89**	**5.51**	**1.98**	**6.63**	**37.95**	**68.03**	**84.52**	**0.630**	**0.475**	**0.380**
$FR_{k=1}$	30.19	10.74	3.76	1.29	5.61	31.64	59.16	77.30	0.589	0.431	0.337
$FR_{k=2}$	38.10	13.52	4.72	1.61	6.50	37.12	66.98	83.88	0.588	0.430	0.332
$FR_{k=3}$	38.70	13.73	4.79	1.64	6.55	37.26	67.14	84.00	0.587	0.429	0.331
$FR_{k=5}$	39.13	13.87	4.83	1.65	6.54	37.28	67.13	83.91	0.587	0.429	0.330

Models	B-1	B-2	B-3	B-4	D-1	D-2	D-3	D-4	AVG	EXT	GRE
TF_{rerank}	42.89	15.51	**5.46**	**1.87**	**3.37**	**23.62**	**50.15**	**71.23**	0.616	0.446	**0.365**
$FR_{k=1}$	43.59	15.46	5.31	1.80	2.36	16.32	36.51	55.72	**0.630**	**0.450**	0.354
$FR_{k=2}$	44.71	15.75	5.37	1.81	2.88	20.77	44.85	65.21	0.629	0.447	0.354
$FR_{k=3}$	44.70	15.74	5.36	1.80	2.89	20.92	45.08	65.48	0.628	0.446	0.352
$FR_{k=5}$	**44.74**	**15.76**	5.37	1.81	2.91	20.99	45.13	65.49	0.627	0.446	0.352

Results....

- The final re-ranking process is sometimes a **remedy**:
  - Simulated futures are sometimes meaningless or include irrelevant informtaion, which has negative effect on second-pass responses.
- The proportions of the test cases where the final responses are picked from the **second-pass responses**:
  - DailyDialog: 40.4%
  - PersonaChat: 36.6%

Case Study

- Case 1

Case 1

History: That is cool. Do you get your books at the bookstore or online?
Gold Response: I usually try to use the library, but otherwise I get them online.
Gold Future: Are you going for elementary education or high school or college?

PropherChat: I usually get them online, but the bookstore is always nice to visit.
PropherChat First-Pass Response: I usually get them online. What about you?

Case Study

- Case 2

Case 2

History: Well, I'm a bit out of shape. I'm thinking about getting some exercises to keep fit.
Gold Response: Oh, that's good news for us.
Gold Future: So what do you provide?

PropherChat: That's a good idea! What kind of exercise do you like to do?
PropherChat First-Pass Response: What kind of exercise?
Simulated Future: Well, I'm thinking of doing some kind of body weight exercises. I don't know if that will

Conclusion

Conclusion

- Enhance the dialogue generation via simulated dialogue futures in **inference phase**.
- Different from MMI model, the backward model in ProphetChat contributes to the generation of second-pass responses.