Findings of EACL 2023

PLACES: Prompting Language Models for Social Conversation Synthesis

Maximillian Chen, Alexandros Papangelis, Chenyang Tao, Andy Rosenbaum, Seokhwan Kim, Yang Liu, Zhou Yu, Dilek Hakkani-Tur

Columbia Engineering
The Fu Foundation School of Engineering and Applied Science

amazon | science

Distilling Synthetic Conversations from Large Language Models

Random In-Context Examples

Topic: Relationships
Background info: Bob got engaged

<Conversation 0>
The following is a conversation between Alice and Bob about relationships. Bob recently got engaged. (Recipe)
Alice: Congrats on your engagement! When do you think you will have your wedding?
Bob: Thank you!! We're thinking of having it in November.
Alice: That's amazing! Will you pick a fancy destination?
...

<Conversation 2>

Topic: Pets
Background info: Alice likes cats. Bob like dogs.

Target Recipe

<Conversation 3>
The following is a conversation between Alice and Bob about pets. Alice likes cats but Bob likes dogs.

Result:
Alice: What kind of pet do you have?
Bob: I have a dog. Her name is Hope.
Alice: Are you an animal person?
Bob: I do love animals. I think they're so special.
Alice: I just adopted a kitten from a shelter. Her name is Mimi.
Bob: Aww, I hope she's doing well. You must be a really good pet parent.
Alice: Thanks! I just wanted to do what I could to make sure she was safe.
Bob: That's the greatest thing anyone can do for a pet. Well, I should probably go now.
Alice: OK! Have a good night.

- **Have you ever struggled with finding the right dataset?** Maybe your dataset doesn't exist, your existing dataset is privacy-restricted, or you aren't happy with your data quality.
- **You can distill high-quality conversations by prompting large language models!** By providing expert-written in-context examples and conversation "recipes," one can generate humanlike synthetic conversations with high semantic control.

Synthetic Conversation Quality

Source	Interesting	Coherent	Natural	Consistent
Human-Collected				
DailyDialog	3.44	4.51	4.85	4.57
Topical Chat	**4.55**	4.39	**4.92**	**4.87**
PLACES				
GPT-J 6B	3.96^*	4.49	4.86	4.36
GPT-NeoX 20B	3.81^*	4.40	4.63	4.35
OPT 30B	4.13^*	$\mathbf{4.61}^{*\dagger}$	4.82	4.63

- **Human evaluations** of synthetic conversations match or outperform human-annotated datasets

Dimension	DD-IC	TC-IC	HW-IC
Interesting	3.82	**4.35**	4.27^*
Coherent	4.48	4.56	$\mathbf{4.77}^{*+}$
Natural	4.54	4.69	$\mathbf{4.69}^*$
Consistent	4.76	**4.87**	4.86^*
On-Topic	0.91	0.88	$\mathbf{0.96}^{*+}$

- **In-context example selection matters**: expert-written examples (HW-IC) outperform crowdsourced conversation examples

Generalizing to Multiparty Conversations

Dimension	MPC	MELD	PLACES
Interesting	2.48	3.52	$\mathbf{4.14}^*$
Coherent	2.40	3.68	$\mathbf{4.65}^*$
Natural	2.69	3.69	$\mathbf{4.47}^*$
Consistent	2.96	3.83	$\mathbf{4.65}^*$
Comprehensible	2.48	3.83	$\mathbf{4.80}^*$
Balanced Engagement	3.45	4.00	$\mathbf{4.89}^*$

- **PLACES** generalizes to the multiparty conversation case, outperforming both MELD (*Friends* transcripts) and MPC (online chats)

Paper **Code**