



PGTask: Introducing the Task of Profile Generation from Dialogues

Rui Ribeiro, Joao P. Carvalho, Luísa Coheur

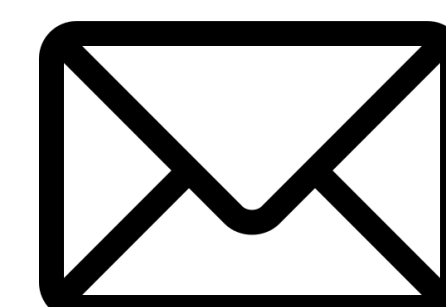
How can we create more personalized dialogue systems?

Speaker 1	Speaker 2
Profile Sentences: I have traveled and studied abroad. I got a job working in advertising last year. I met my best friend in kindergarten.	Profile Sentences: I love caring for people. I work as a nurses aide in a nursing home. I have two dogs and one cat.
Hey, are you a student? I traveled a lot, I even studied abroad.	No, I work full time at a nursing home. I am a nurses aide.
Nice, I just got a advertising job myself. Do you like your job?	Yes, I do. Caring for people is the joy of my life.
Nice, my best friend is a nurse , I knew him since kindergarten.	Very cool. Do you have pets?
No I do not, do you?	Yes. Two dogs and a cat. They are my babies.

Train	# Samples	34355
	Avg. Profile Sentences	1.06
	Avg. Utterance Words	13.13
Valid	Avg. Profile Sentence Words	7.14
	# Samples	4236
	Avg. Profile Sentences	1.06
Test	Avg. Utterance Words	13.36
	Avg. Profile Sentence Words	7.67
	# Samples	3760
Test	Avg. Profile Sentences	1.06
	Avg. Utterance Words	13.05
	Avg. Profile Sentence Words	7.17

Utterance	Profile Sentences
I enjoy hanging with my mother she is my best friend.	My mom is my best friend.
I am almost done, I only have two years left in law school.	I have got two more years in college. I study law.

The task is **more challenging** than just finding utterances that highly relate to each profile sentence. For instance, the profile sentence “**I like all genres of music**” is probably at the origin of the utterance “**Yes, sometimes I also listen to classical music**”, but we cannot extract that profile sentence from that single utterance.



rui.m.ribeiro@inesc-id.pt
github.com/ruinunca/PGTask

Benchmark results for future work

	Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-1	ROUGE-2	ROUGE-L	BERTScore
W/o FT	distilgpt2	5.59	0.30	0.00	0.00	6.86	0.93	5.80	84.66
	gpt2-small	4.87	0.40	0.00	0.00	6.08	0.63	5.20	84.21
	gpt2-medium	4.48	0.20	0.00	0.00	7.20	0.31	5.32	83.28
W/ FT	distilgpt2	44.42	13.18	5.60	0.00	35.68	14.12	35.39	92.35
	gpt2-small	61.30	32.30	20.62	9.44	50.07	28.31	50.00	94.39
	gpt2-medium	59.31	25.94	15.30	9.17	46.32	24.14	45.88	94.76

Acknowledgments:

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UIDB/50021/2020 and grant 2022.10640.BD, by the project CMU-PT MAIA with reference 045909, as well as by the Recovery and Resilience Plan (RRP) and Next Generation EU European Funds through project C644865762-00000008 Accelerat.AI.