

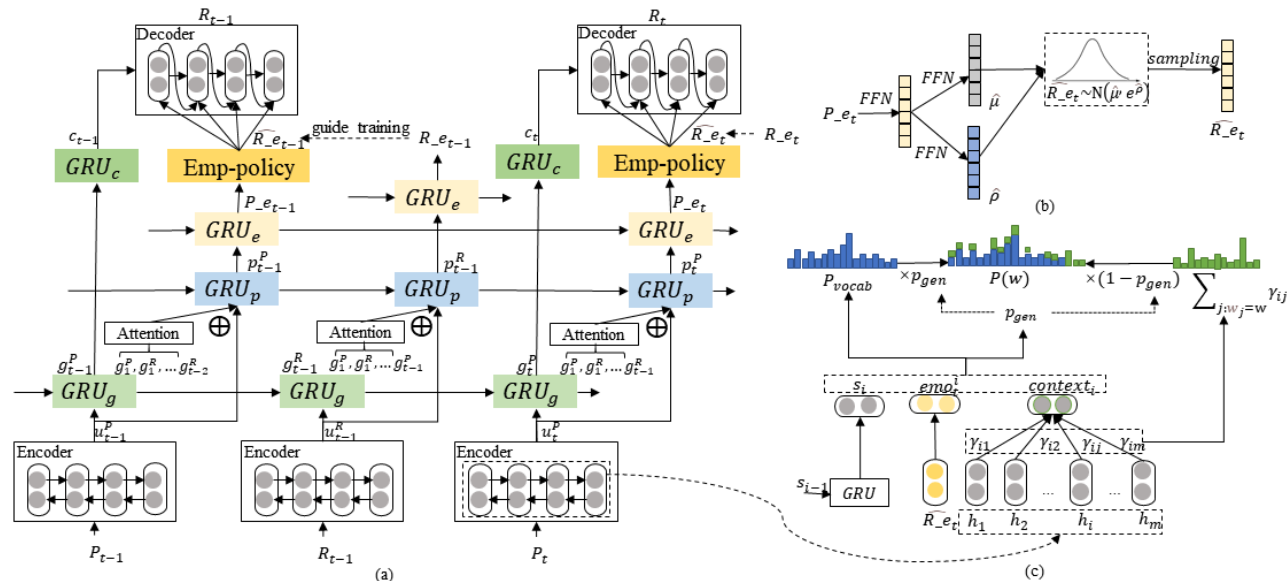
Generating Empathetic Responses through Emotion Tracking and Constraint Guidance

Jing LI, Donghong HAN, Zhishuai GUO, Baiyou QIAO,
Gang WU

Frontiers of Computer Science, DOI: [10.1007/s11704-023-2792-7](https://doi.org/10.1007/s11704-023-2792-7)

Problems & Ideas

- Problems of empathetic dialogue generation:
 - The previous methods ignore the continuity of parties' emotion. The emotional perception ability need to be further improved.
 - The appropriate response emotions and utterances are not unique. It makes the design of flexible and specific empathetic policy more difficult.
- Ideas: we present ETHREED model which uses hierarchical GRUs to model parties' emotion flow respectively and exploits a stochastic policy network to learn the transition process of parties' emotion.



The related diagrams of ETHREED, including (a) the overall architecture, (b) the policy network and (c) the pointer generation network

Main Contributions

- Contributions:
 - A novel empathetic dialogue generation model ETHREED is proposed which conducts hierarchical GRUs to model the continuous process of different parties and improves the emotional perception ability.
 - A stochastic policy network is conducted to learn the empathetic policy, which is optimized by the constraint-based guided policy search method. This preserves the flexibility of reinforcement learning when training on real data.
 - The experimental results on the Empathetic Dialogue dataset show that the responses generated by ETHREED are more diverse and empathetic.

Method	Params.	Automatic evaluation					Human rating			A/B test (ETHREED vs)			
		Acc	PPL	Dis-1	Dis-2	F _{BERT}	Emp.	Rel.	Flu.	Win(%)	Loss(%)	Tie(%)	κ
MIME	17.80M	0.3239	37.28	0.39	1.56	0.189	3.14	3.19	4.85	46.67	27.33	26.00	0.45
EmpDG	29.29M	0.3278	36.26	0.45	2.13	0.195	3.21	3.19	4.72	41.33	35.33	23.33	0.44
KEMP	32.54M	0.3852	34.56	0.61	2.69	0.223	3.20	3.16	4.79	43.33	31.67	25.00	0.46
EmpHi	10.92M	0.4046	33.49	0.98	4.64	0.187	3.07	3.05	4.67	40.67	37.67	21.67	0.36
ETHREED	11.98M	0.4294	27.74	1.14	4.76	0.226	3.25	3.23	4.85	-	-	-	-

All results of automatic and human evaluation. All results of automatic and human evaluation. The best performing results are highlighted. We reproduce the baselines based on the hyperparameters in their papers, repeat 5 runs with different seeds and average the results for each automatic metrics. Emp., Rel. and Flu. are short for Empathy, Relevance and Fluency. The Fleissia-Kappa κ for three rating aspects are 0.47, 0.51 and 0.41, which indicate three annotators reach a moderate agreement. κ of a/b test all achieve the fair or moderate agreement.