Personalize empathic conversations

汇报人：黄正结

2022.10.26

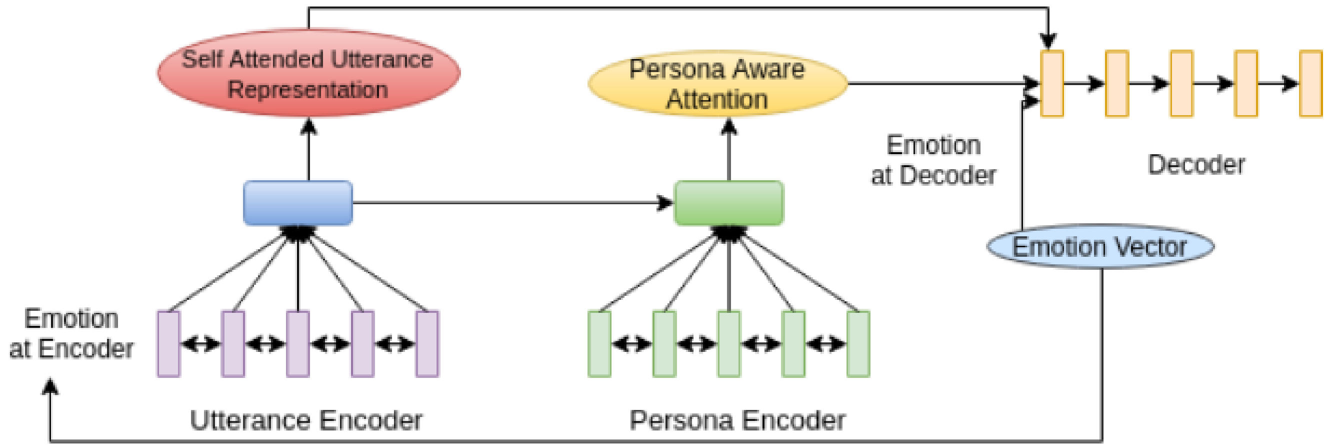
对话系统

在聊天机器人的研究重点是与用户进行聊天，选择合适的话题以适应对话环境，在这个过程中需要考虑诸多因素，主要有上下文语义、人物情绪、人物性格、人物偏好、常识知识等等

Persona aware Response Generation with Emotions 2020 IEEE

主要研究内容:

提出了一种新的角色感知注意力方法，能够将情感信息注入到反应中，使得对话系统可以在考虑人物角色信息和情感信息的同时，通过对话语境，生成具有同情心的、个性化的反应



数据来源:

使用EmpatheticDialogues数据集训练情感标注器，然后使用情感标注器对PersonaChat数据集进行进一步的标注

Table II: Classification scores of Emotion on Empathetic-Dialogue data. E-F1 denote the weighted average F1 score of emotion

Model	**E-F1**
LSTM	37.06
CNN	34.90
Bi-LSTM	39.87
BERT [34]	61.74
RoBERTa [35]	59.89

实验:

Table IV: Experimental results of different models. Here PAA represents Persona-Aware Attention, EE represents Emotion at Encoder, ED represents Emotion at Decoder

Model Description		Perplexity	BLEU	Rouge-L	Emotion Accuracy	Distinct-1	Distinct-2
Baseline Approaches	Seq2Seq	59.11	0.042	0.149	0.35	0.0125	0.0464
	Seq2Seq + Attn	58.23	0.047	0.151	0.38	0.0131	0.0472
	Seq2Seq + Attn + PAA	57.60	0.088	0.154	0.42	0.0163	0.0581
	Seq2Seq + Attn + EE	56.87	0.092	0.157	0.58	0.0155	0.0534
	Seq2Seq + Attn + ED	56.39	0.096	0.158	0.61	0.0158	0.0562
Proposed Approaches	**Seq2Seq + Attn + PAA + EE**	**55.59**	**0.099**	**0.162**	**0.65**	**0.0189**	**0.0844**
	Seq2Seq + Attn + PAA + ED	**52.68**	**0.108**	**0.169**	**0.67**	**0.0210**	**0.0923**

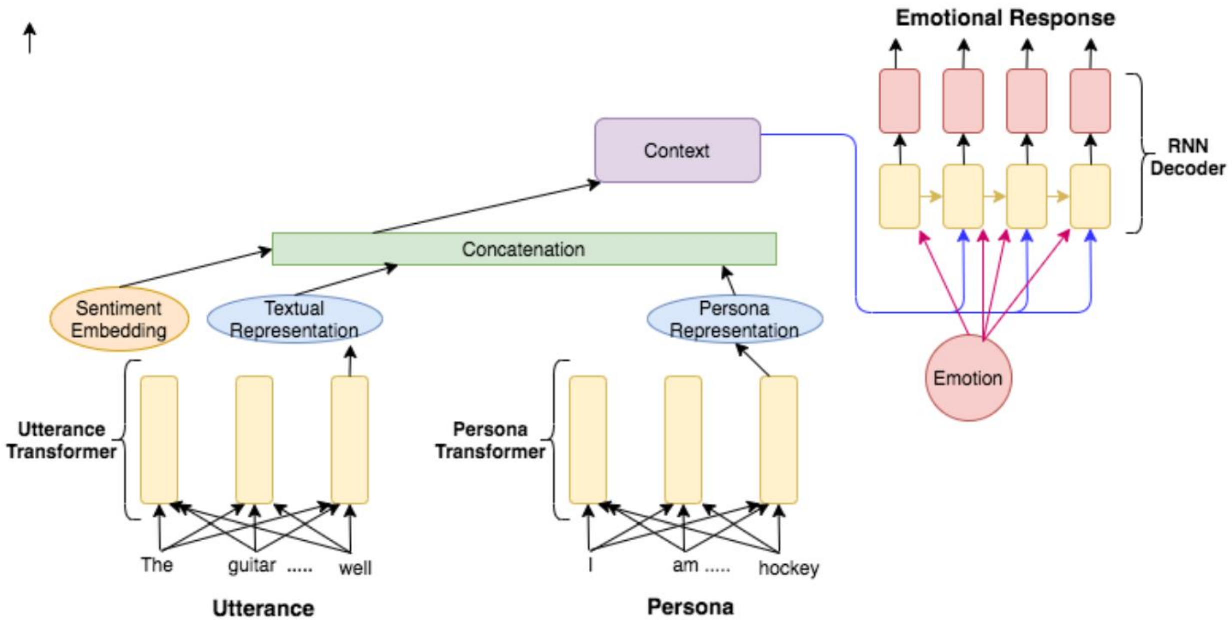
Table V: Results of Human Evaluation

Model Description		Fluency			Emotion		Persona Consistency	
		0	**1**	**2**	**0**	**1**	**0**	**1**
Baseline Approaches	Seq2Seq	27.36	45.83	26.81	75.93	24.07	77.20	22.80
	Seq2Seq + Attn	26.11	44.71	29.18	74.56	25.44	76.14	23.86
	Seq2Seq + Attn + PAA	23.41	42.96	33.63	73.81	26.19	51.64	48.36
	Seq2Seq + Attn + EE	24.17	43.11	32.72	59.33	40.67	70.88	29.12
	Seq2Seq + Attn + ED	23.05	42.88	34.07	57.49	42.51	70.31	29.69
Proposed Approaches	**Seq2Seq + Attn + PAA + EE**	**19.64**	**38.65**	**41.71**	**55.72**	**44.28**	**49.85**	**50.15**
	Seq2Seq + Attn + PAA + ED	**18.15**	**37.32**	**44.53**	**53.91**	**46.09**	**48.11**	**51.89**

SEPRG: Sentiment aware Emotion controlled Personalized Response Generation 2021 ACL

主要研究内容:

提出了一种新的基于Transformer的编码器-解码器框架，能够在响应中注入情感、情感和角色信息，使得对话系统可以在考虑人物角色信息和情感信息的同时，通过对话语境，生成具有同情心的、个性化的反应



SEPRG: Sentiment aware Emotion controlled Personalized Response Generation 2021 ACL

数据来源:

沿用Persona aware Response Generation with Emotions的数据集，但是将情感类别从原先的细粒度，重新标注为中性，积极，消极，重新标注方法为将grateful与sad等这类可以很好区分积极消极的情感自动标注，剩余情感进行人工注释

实验:

Model Description		Perplexity	BLEU-4	Rouge-L	Distinct-1	Distinct-2	Emotion Accuracy
Baseline Approaches	*Seq2Seq* (*Sutskever et al., 2014*)	56.11	0.089	0.196	0.0125	0.0464	0.358
	HRED (*Serban et al., 2017*)	55.63	0.096	0.201	0.0128	0.0469	0.376
	Seq2Seq + E + P (*Firdaus et al., 2020*)	54.13	0.103	0.189	0.0168	0.0549	0.657
	HRED + E + P	54.85	0.116	0.224	0.0174	0.0592	0.665
	Seq2Seq + E + P + S	53.61	0.115	0.203	0.0171	0.0555	0.673
	HRED + E + P + S	52.46	0.127	0.237	0.0186	0.0590	0.689
Proposed Approach	*Trans + E + P + S*	**51.92**	**0.143**	**0.266**	**0.0219**	**0.0987**	**0.715**
Ablation Study	*Trans*	53.47	0.118	0.239	0.0189	0.0883	0.678
	Trans + E + P	53.44	0.125	0.242	0.0193	0.0896	0.695

Table 3: Results of automatic evaluation. Here, E-Emotion, P-Persona, S-Sentiment, Trans-Transformers

Model Description		Fluency	Relevance	Emotion Appropriateness	Persona Consistency	Sentiment Coherence
Baseline Approaches	*Seq2Seq* (*Sutskever et al., 2014*)	2.98	2.65	38%	35%	33%
	HRED (*Serban et al., 2017*)	3.16	2.89	41%	39%	37%
	Seq2Seq + E + P (*Firdaus et al., 2020*)	3.29	3.02	53%	48%	42%
	HRED + E + P	3.45	3.18	59%	55%	47%
	Seq2Seq + E + P + S	3.33	3.21	56%	52%	51%
	HRED + E + P + S	3.52	3.45	61%	58%	55%
Proposed Approach	*Trans + E + P + S*	**3.89**	**3.66**	**65%**	**67%**	**62%**

Table 4: Results of human evaluation for the existing baselines and the proposed framework

SEPRG: Sentiment aware Emotion controlled Personalized Response Generation 2021 ACL

样例分析:

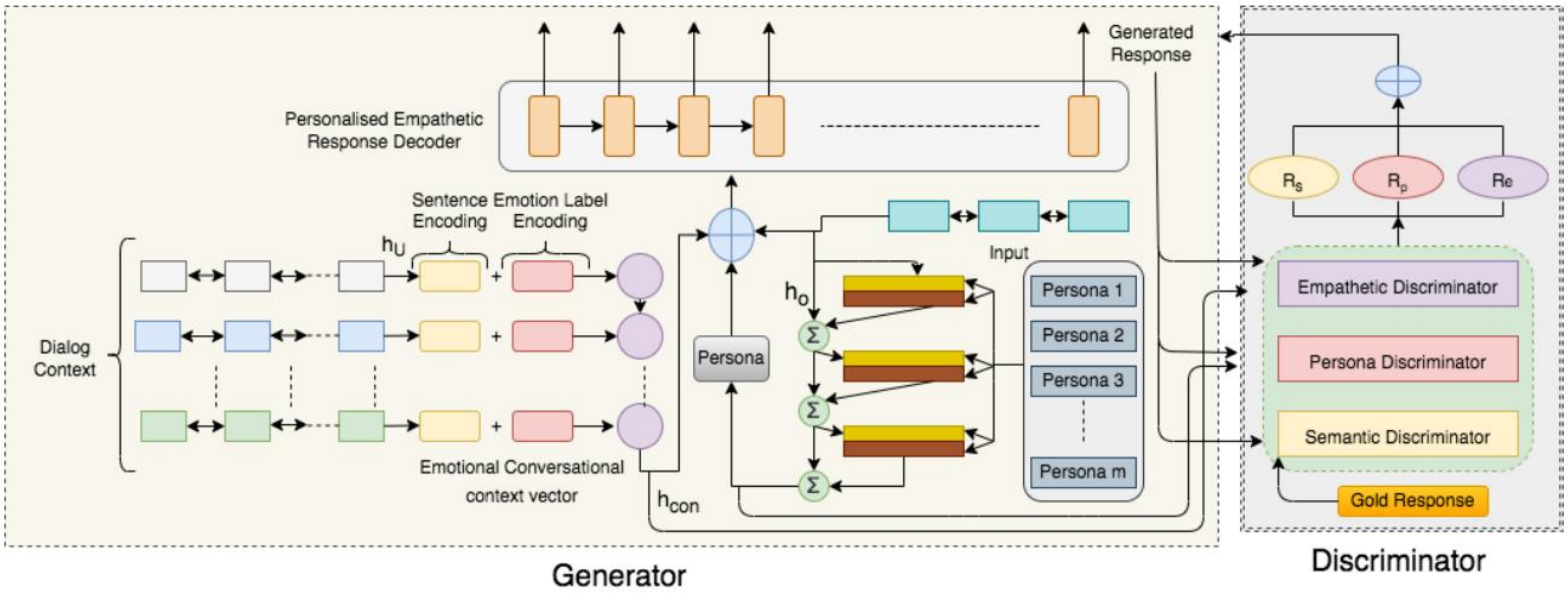
Dialog Input	Persona Information		Ground-Truth	Generated Response
[Person 1] What do you do for a career? [Person 2] I like to watch kids. [Person 1] I actually play guitar and do lot of welding.	Persona 1 I am primarily a meat eater. I am a guitar player. Welding is my career field. My parents don't know I am gay.	Persona 2 I've a sweet tooth. I'm a babysitter and drive a mercedes. I'm the middle child of 3 sibling. I'm getting married in six weeks.	What do you weld? houses? Neutral	**Seq2Seq:** What is weld? ¡UNK¿ ¡UNK¿ **HRED:** What you weld? **Seq2Seq+E+P:** Wow! What you weld mostly? (surprise) **HRED+E+P:** Really! Do you weld houses? (surprise) **Proposed:** Wow that's great, what do you weld mostly, is it houses? (surprise)
[Person 1] Hi! do you like turtles? [Person 2] I have two cats actually. [Person 1] I have a turtle his name is Speedy.	Persona 1 I don't pick up my toys. I have a pet turtle. I like to play with my dolls. I'm fourteen.	Persona 2 I love cats and have two cats. My favorite season is winter. I won a gold medal in the 1992 olympics. I've a hat collection of over 1000 hats.	I am a cat person. Positive	**Seq2Seq:** I have cats. **HRED:** I like cats mostly. **Seq2Seq+E+P:** Turtles are nice but I like cats. (joy) **HRED+E+P:** Nice name for a pet, but I love cats. (joy) **Proposed:** That is an adorable name for a turtle! but I am a cat person. (joy)

I enjoy writing and playing, do you?: A Personalized and Emotion Grounded Dialogue Agent using Generative Adversarial Network

2022 IEEE

主要研究内容:

提出了一种生成对抗方法EP-GAN (Empathy and Persona aware Generative Adversarial Network)，使得对话系统可以在考虑人物角色信息和情感信息的同时，通过对话语境，生成具有同情心的、个性化的反应



I enjoy writing and playing, do you?: A Personalized and Emotion Grounded Dialogue Agent using Generative Adversarial Network

2022 IEEE

数据来源:

使用EmpatheticDialogues数据集训练情感标注器，然后使用情感标注器对PersonaChat数据集进行进一步的标注

TABLE II: Emotion classification scores on EmpatheticDialogue data. The weighted average F1 score of emotion is denoted by E-F1.

Model	E-F1
LSTM	37.06
CNN	34.90
Bi-LSTM	39.87
BERT [83]	61.74
RoBERTa [84]	59.89
DistilBERT [85]	**64.95**
AlBERT [86]	63.11

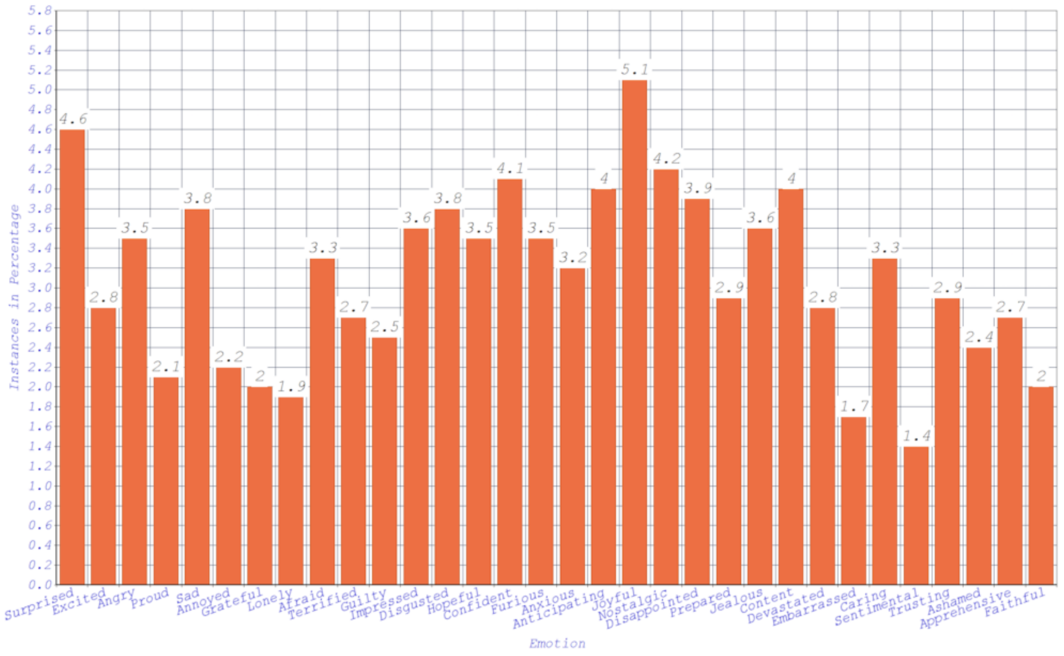


Fig. 2: Emotion Distribution of the PersonaChat Dataset.

I enjoy writing and playing, do you?: A Personalized and Emotion Grounded Dialogue Agent using Generative Adversarial Network

2022 IEEE

实验结果:

Model Description		Perplexity	BLEU-4	Rouge-L	Distinct-1	Distinct-2	Emotion Accuracy
Baseline Approaches	***Seq2Seq*** *[82]*	56.11	0.089	0.196	0.0125	0.0464	0.358
	HRED *[25]*	55.63	0.096	0.201	0.0128	0.0469	0.376
	Trans *[6]*	53.17	0.121	0.228	0.0186	0.0749	0.451
	SeqGAN *[80]*	55.61	0.098	0.203	0.0133	0.0470	0.381
	Seq2Seq + E + P *[21]*	54.13	0.103	0.189	0.0168	0.0549	0.657
	HRED + E + P	54.85	0.116	0.224	0.0174	0.0592	0.665
	Trans + E + P	52.87	0.132	0.241	0.0203	0.0839	0.681
	CoBERT *[22]*	51.09	0.138	0.258	0.0210	0.0894	0.693
Proposed Approach	***EP-GAN***	**51.92**	**0.143**	**0.266**	**0.0219**	**0.0987**	**0.715**
Ablation Study	***EP-GAN - SD***	53.47	0.118	0.239	0.0189	0.0883	0.678
	EP-GAN - ED	53.44	0.125	0.242	0.0193	0.0896	0.695
	EP-GAN - PD	52.39	0.129	0.249	0.0199	0.0953	0.683
	EP-GAN - SD + ED	52.26	0.130	0.251	0.0206	0.0976	0.708
	EP-GAN - SD + PD	52.12	0.135	0.257	0.0210	0.0971	0.688

Model Description		Fluency	Relevance	Emotion Appropriateness	Persona Consistency
Baseline Approaches	***Seq2Seq*** *[82]*	2.98	2.65	38%	35%
	HRED *[25]*	3.16	2.89	41%	39%
	Trans *[6]*	3.51	3.24	45%	59%
	SeqGAN *[80]*	3.25	2.95	42%	41%
	Seq2Seq + E + P *[21]*	3.29	3.02	53%	48%
	HRED + E + P	3.45	3.18	59%	55%
	Trans + E + P	3.68	3.51	61%	62%
	CoBERT *[22]*	3.72	3.59	63%	65%
Proposed Approach	***EP-GAN***	**3.89**	**3.66**	**65%**	**67%**

I enjoy writing and playing, do you?: A Personalized and Emotion Grounded Dialogue Agent using Generative Adversarial Network

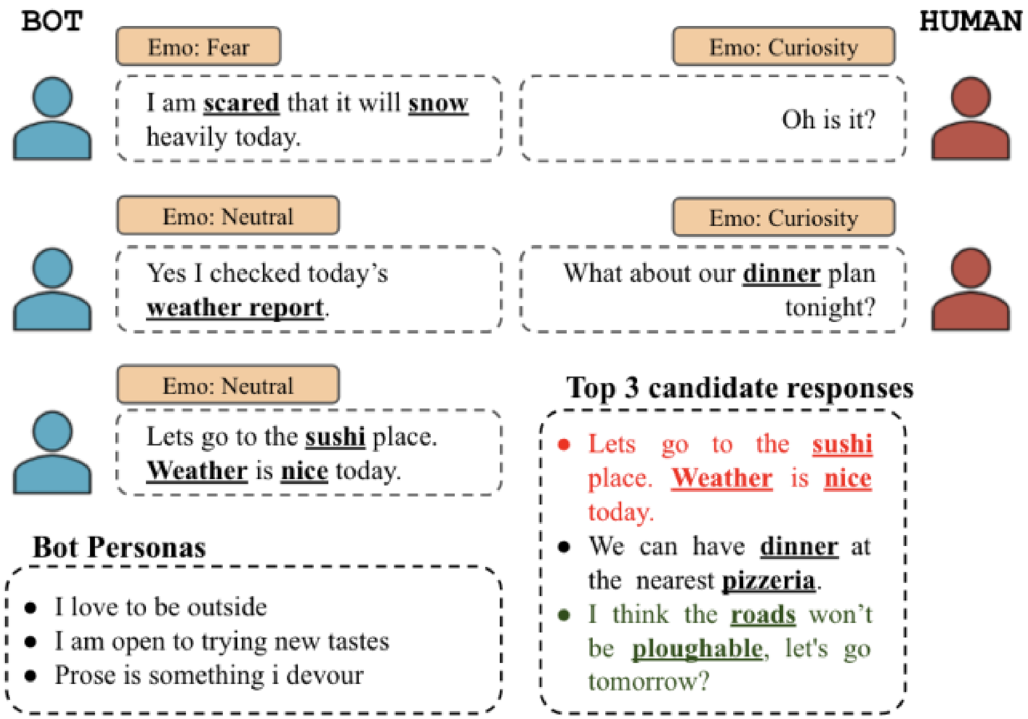
2022 IEEE

样例分析:

Dialog Input	Persona Information		Ground-Truth	Generated Response
[*Person 1*] What do you do for a career? [*Person 2*] I like to watch kids. [*Person 1*] I actually play guitar and do lot of welding.	**Persona 1** I am primarily a meat eater. I am a guitar player. Welding is my career field. My parents don't know I am gay.	**Persona 2** I've a sweet tooth. I'm the middle child of 3 sibling. I'm getting married in six weeks.	What do you weld? houses?	**Seq2Seq:** What is weld? ¡UNK¿ ¡UNK¿ **HRED:** What you weld? **Trans:** Welding houses are you? **SeqGAN:** What are you welding? **Seq2Seq+E+P:** Wow! What you weld mostly? (surprise) **HRED+E+P:** Really! Do you weld houses? (surprise) **Trans + E + P:** Oh my god! you weld houses? (surprise) **EP-GAN - ED:** Do you weld homes? **EP-GAN - PD:** That's nice, what you do? (surprise) **EP-GAN:** Wow that's great, what do you weld mostly, is it houses? (surprise)
[*Person 1*] Hi! do you like turtles? [*Person 2*] I have two cats actually. [*Person 1*] I have a turtle his name is Speedy.	**Persona 1** I don't pick up my toys. I have a pet turtle. I like to play with my dolls. I'm fourteen.	**Persona 2** I love cats and have two cats. My favorite season is winter. I've a hat collection of over 1000 hats.	I am a cat person.	**Seq2Seq:** I have cats. **HRED:** I like cats mostly. **Trans:** Cats are lovely. **SeqGAN:** I am into cats. **Seq2Seq+E+P:** Turtles are nice but I like cats. (joy) **HRED+E+P:** Nice name for a pet, but I love cats. (joy) **Trans + E + P:** Nice turtle name, but I am more in cats. (joy) **EP-GAN - ED:** I have cats only. **EP-GAN - PD:** I love to have pets. (joy) **EP-GAN:** That is an adorable name for a turtle! but I am a cat person. (joy)

主要研究内容:

试图通过提出一套融合策略来捕捉话语中人物角色、情感和隐含信息之间的交互。



数据来源：

基于Persona-Chat，增加了Emotion、Entailment、Concept Mining的标注

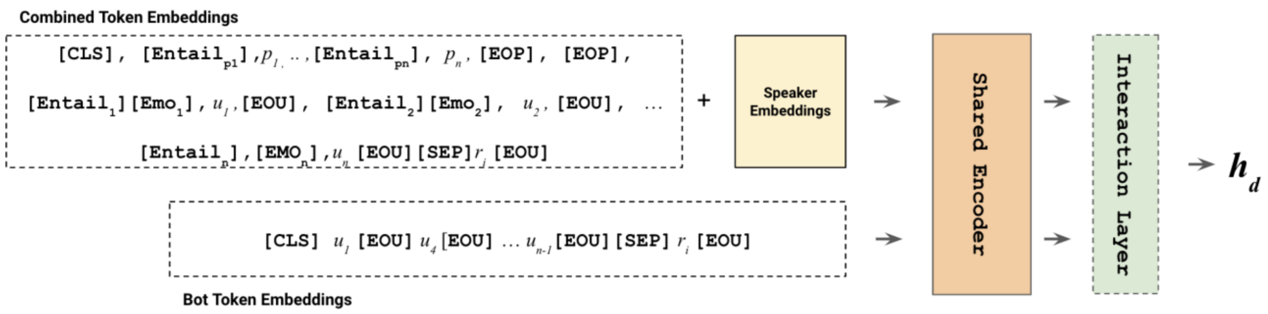
Emotion：使用GoEmotions数据集训练了一个基于RoBERTa的情感标注器，在测试集上Macro F1最高达到49.4%，为保证效果最终只保留预测准确率超过90%的情感类别

Entailment：将一个在SNLI数据上训练的RoBERTa模型和一个在DECODE数据（对话矛盾检测）上训练的RoBERTa模型的预测分布加权求和来标注{ entailment, neutral, contradiction }

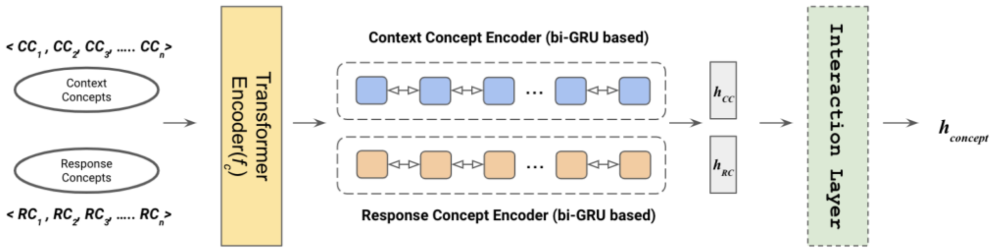
Concept Mining：从persona, utterances和responses中挖掘关键词和关键短语。假设在回答中出现的概念应该符合说话者的角色。因此，通过计算人物角色关键字和响应上下文关键字之间的逐点互信息得分的平均值，并拒绝低于阈值的Concept，来修剪一些响应上下文关键字

Using Multi-Encoder Fusion Strategies to Improve Personalized Response Selection arxiv

模型:



(a) Dual encoder pipeline consisting of combination of all the encoding strategies.



(b) Concept-flow interaction network, the output of this network $\mathbf{h_{concept}}$ can be concatenated with any of the BERT based dual encoder's output($\mathbf{h_d}$).

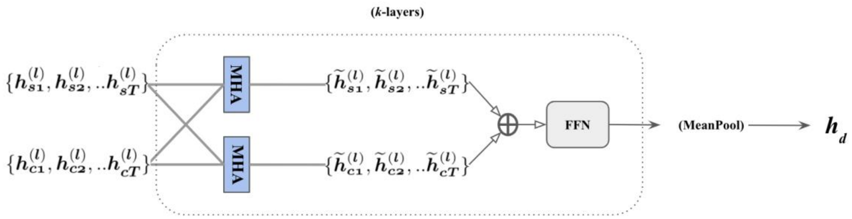


Figure 3: Interaction Layer

实验结果:

Model	Self Persona				Partner Persona			
	Original		Revised		Original		Revised	
	hits@1	**MRR**	**hits@1**	**MRR**	**hits@1**	**MRR**	**hits@1**	**MRR**
FT-PC (Mazaré et al., 2018)	-	-	60.7	-	-	-	-	-
DIM (Gu et al., 2019b)	78.8	86.7	70.7	81.2	64.0	76.1	63.9	76.0
TransferTransfo (Wolf et al., 2019)	80.7	-	-	-	-	-	-	-
FIRE (Gu et al., 2020c)	81.6	-	74.8	-	-	-	-	-
BERT-CRA (Gu et al., 2021b)	84.3	90.3	79.4	86.9	71.2	80.9	71.8	81.5
Baseline	84.4	90.7	79.4	87.6	71.2	81.1	71.4	81.5
BERT-EmA	84.6	90.9	79.8	87.7	71.4	81.2	71.4	81.6
BERT-P-EnA	85.3	91.2	80.5	87.9	71.7	81.3	71.3	81.4
BERT-EmA+BERT-P-EnA	85.8	91.4	80.7	88.0	72.3	81.5	71.7	81.5
BERT-EmA+BERT-P-EnA+CF (All)	**86.6***	**91.6***	**81.3***	**88.6***	**72.6***	**81.9***	**72.4***	**81.9***

样例分析:

personas	my favorite color is blue . <ent: neutral> I enjoy reading mysteries . <ent: neutral> I have seven children . <ent: entail> I grew up on a large farm . <ent: neutral>
context	A: hello how are you today? <emo:curiosity> <ent: neutral> B: I am well. how are you? <emo:curiosity> <ent: neutral> A: I am doing great just got back from the beach <emo:excitement> <ent: neutral> B: that is great. I live far from the beach . <emo:caring> <ent: neutral> A: I am very lucky we live beside the beach. what do you do for a living? <emo:curiosity> <ent: neutral> B: I keep busy with my seven children . <emo:excitement> <ent: neutral> A: wow that much have taken some adjusting I teach kindergarten . <emo:surprise> <ent: neutral>
golden response	do you teach mysteries to your children ? they are my favorite type of novel . <emo:curiosity>
BERT-CRA	that must be a lot of work but very rewarding i bet <emo:realization>
All	do you teach mysteries to your children ? they are my favorite type of novel . <emo:curiosity>

Table 4: Case study showing concept flow.

Automatically Select Emotion for Response via Personality-affected Emotion Transition 2021 ACL

主要研究内容:

为了与用户提供一致的情感互动，对话系统应该能够自动选择合适的情感，以应对人类这样的反应。然而，现有的大多数作品都侧重于在回应中呈现特定的情绪，或以同理心回应用户的情绪，但忽略了情绪表达的个体差异。这可能会导致不一致的情感表达和用户不感兴趣。

Basic Emotions	(Valence, Arousal, Dominance)
Anger	(-0.51, 0.59, 0.25)
Disgust	(-0.60, 0.35, 0.11)
Fear	(-0.62, 0.82, -0.43)
Joy	(0.81, 0.51, 0.46)
Neutral	(0.00, 0.00, 0.00)
Sadness	(-0.63, -0.27, -0.33)
Surprise	(0.40, 0.67, -0.13)

Table 1: Emotions in the VAD Space.

将情感通过映射到VAD空间的方式建模

Factor	Description
Openness	Openminded, imaginative, and sensitive.
Conscientiousness	Scrupulous, well-organized.
Extraversion	The tendency to experience positive emotions.
Agreeableness	Trusting, sympathetic, and cooperative.
Neuroticism	The tendency to experience psychological distress.

Table 2: The OCEAN personality traits and description (Costa and McCrae, 1992)

将人格建模为OCEAN得分

$$P_V = 0.21E + 0.59A + 0.19N$$
$$P_A = 0.15O + 0.30A - 0.57N$$
$$P_D = 0.25O + 0.17C + 0.60E - 0.32A$$

OCEAN得分映射到VAD空间

Automatically Select Emotion for Response via Personality-affected Emotion Transition 2021 ACL

个性化的情绪转化模型

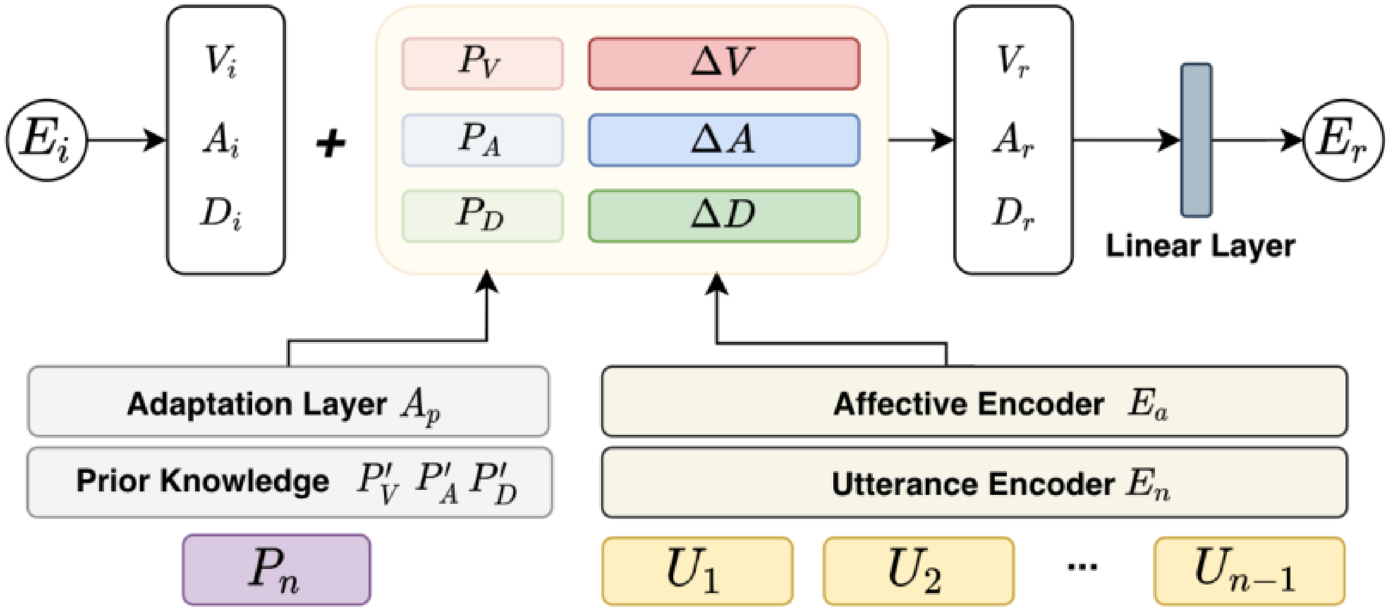


Figure 1: The Model Illustration

Automatically Select Emotion for Response via Personality-affected Emotion Transition

2021 ACL

数据集中主要角色的OCEAN值

Roles	**Personality Traits (O,C,E,A,N)**
Chandler	[0.648, 0.375, 0.386, 0.58, 0.477]
Joey	[0.574, 0.614, 0.297, 0.545, 0.455]
Monica	[0.713, 0.457, 0.457, 0.66, 0.511]
Phoebe	[0.6, 0.48, 0.31, 0.46, 0.56]
Rachel	[0.635, 0.354, 0.521, 0.552, 0.469]
Ross	[0.722, 0.489, 0.6, 0.533, 0.356]

Table 3: Personalities of *Friends* main roles in PELD.

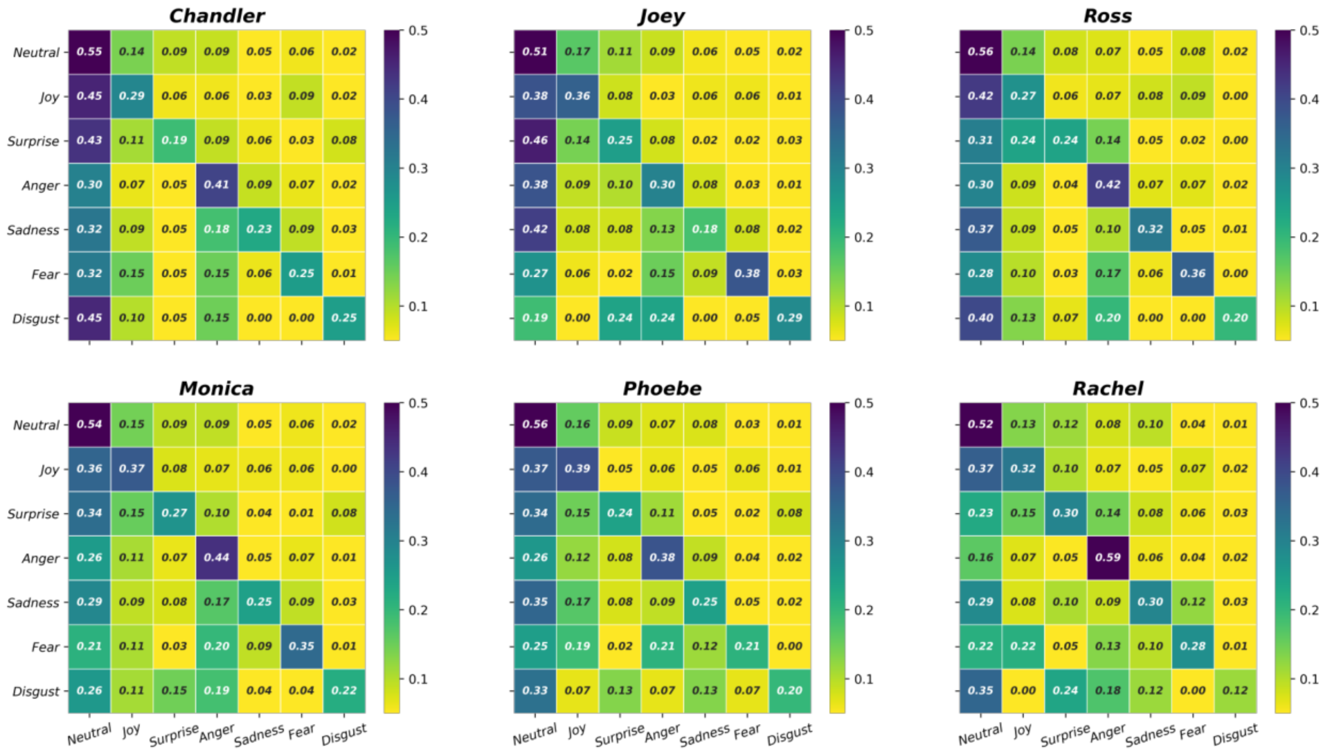


Figure 3: Emotion transition matrixes of the six main roles in PELD. Each row in a matrix shows the ratios of the current emotion E_i is transferred to the next emotion E_r.

情感分类任务实验结果

Methods	Negative	Neutral	Positive	m-avg	w-avg
RoBERTa	0.415	0.430	0.323	0.389	0.390
RoBERTa-P	0.401	**0.505**	0.176	0.361	0.430
PET-CLS	**0.492**	0.474	**0.327**	**0.431**	**0.445**

Table 7: Results for Sentiment Prediction.

Methods	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise	m-avg	w-avg
RoBERTa	0.218	0.000	0.107	0.214	0.453	0.122	0.126	0.177	0.287
RoBERTa-P	0.178	0.000	0.047	**0.265**	0.517	0.110	0.053	0.167	0.352
PET-VAD	0.190	**0.081**	0.115	0.188	0.474	0.000	**0.179**	0.175	0.309
PET-CLS	**0.320**	0.070	**0.140**	0.198	**0.528**	**0.155**	0.098	**0.203**	**0.424**

Table 6: Results for Emotion Prediction.

汇报结束，谢谢大家！