THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Comprehending and Reflecting Personality in Dialog Systems

Ph.D. Thesis Defense

Zhiyuan Wen

Supervisor:

Prof. CAO Jiannong, PolyU

Committee:

Prof. LIU, Yan Fiona, PolyU

Prof. CHEUNG, Yiu-ming, HKBU

Prof. MA Xiaojuan, HKUST

Outline

- Background & Motivation
- Research Problem
- Research Challenges
- Research Framework
- Research Contributions
  - Affective Dialog Encoder for Personality Recognition in Conversation
  - DesPrompt: Personality-descriptive Prompt Tuning for Few-shot Personality Recognition
  - Personality-affected Emotion Generation in Dialog Systems
  - Decode with Template: Content Preserving Sentiment Transfer
- Conclusions & Future Directions

Background

Development of dialog systems has revolutionized human-computer interactions

- Long history of development
  - Eliza (rule-based chatbot in 1966) → Statistical-based Model → Retrieval-based Systems → Generative Models → ChatGPT (State-of-the-art)
  - Output pre-defined responses by rules → providing human-like conversation experience
- Plenty of applications
  - Virtual assistant/digital human
  - Education/Healthcare robot
  - Entertainment: emotional support, chit-chat, NPC in games,…

Eliza Chatbot

Tencent Cloud AI Digital Human


LuxAI Educational Robot


First AI NPC by Nvidia

Motivation

- Most dialog systems are trained on massive dialog content from various speakers to **only learn the general conversation patterns**
  - Deliver general responses regardless of users' characteristics
  - Lack stable and interpretable emotional expression
  - Generate content in inconsistent language styles

- Personality is a set of **relatively stable** individual traits and characteristics that define a person's **distinctive pattern** of thinking, feeling, and acting[1]
  - Represent characteristics of users
  - Encompass tendency of emotional expressions
  - Affect word usages and language tones in responses

The Big Five Personality Traits

Comprehending and reflecting personality for personalization, emotional intelligence, and language style consistency in dialog systems

[1] Mischel, W., Shoda, Y., & Ayduk, O. (2007). Introduction to personality: Toward an integrative science of the person. John Wiley & Sons.

Research Problem

- How to comprehend personality, equip personality to dialog systems, and reflect personality in responses?
  - Understand the personality manifested in dialog context
    - Analyze cues from conversation content to infer users' personality traits
  - Incorporate personality traits into dialog systems
    - Specify and quantitatively model the personality trait for dialog systems
  - Ensure responses consistently reflect the specified personality trait
    - Affect and adjust language style, tone, and content to match the specified personality

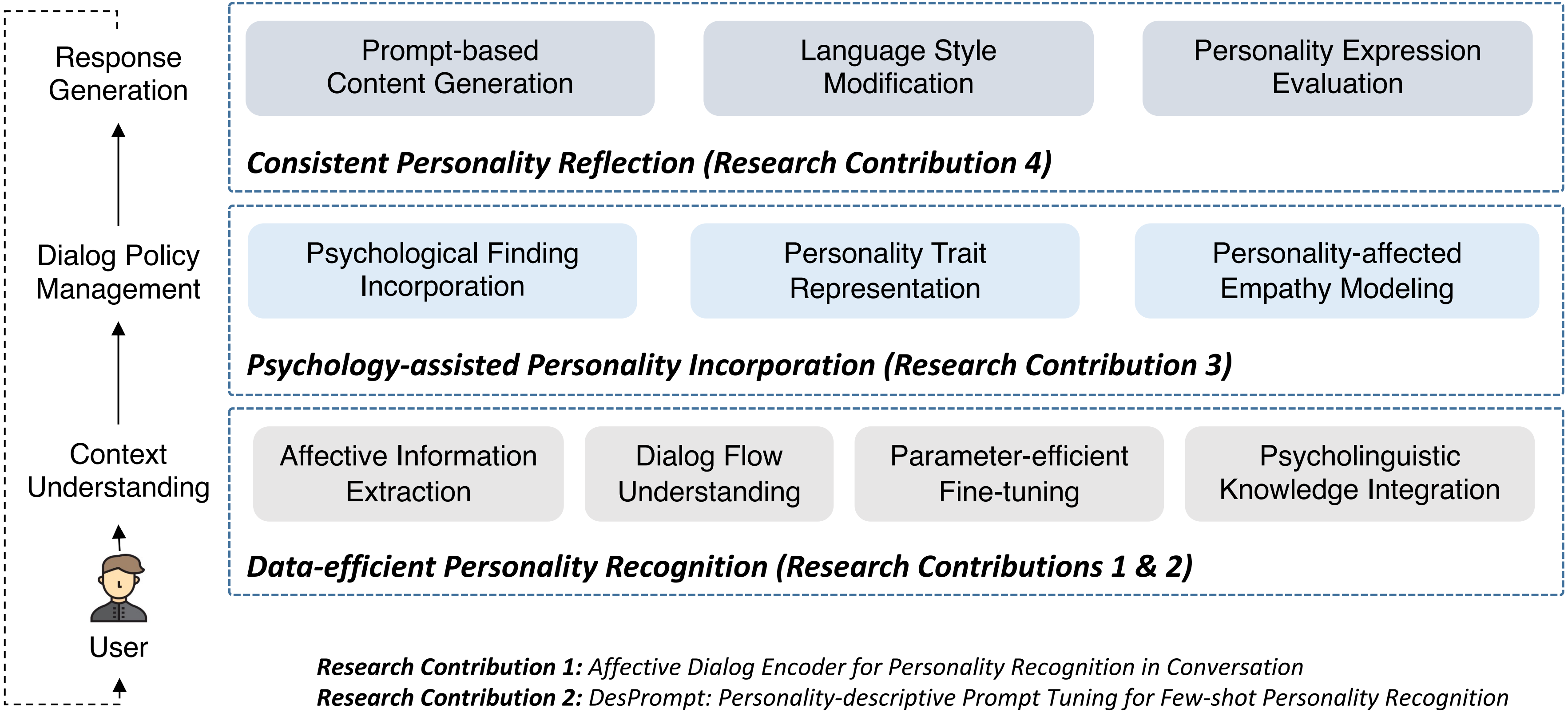
Research Challenges

- Comprehending personality in insufficient data
  - Understanding **long-term** patterns (personalities) from **short-term** conversations is difficult
  - Dialog content with precise personality annotations is rare

- Integrating psychological findings into neural network (NN)-based models
  - Personality is defined and analyzed in psychology; psychological findings on personality provide theoretical evidence for integrating personality into dialog systems
  - Findings in **small groups** may be unsuitable for NN-based models trained on the **massive general corpus**

- Reflecting personality consistently across various dialog contexts
  - Identifying and effectively controlling factors influenced by personality **consistently** across **different dialog contexts** is difficult

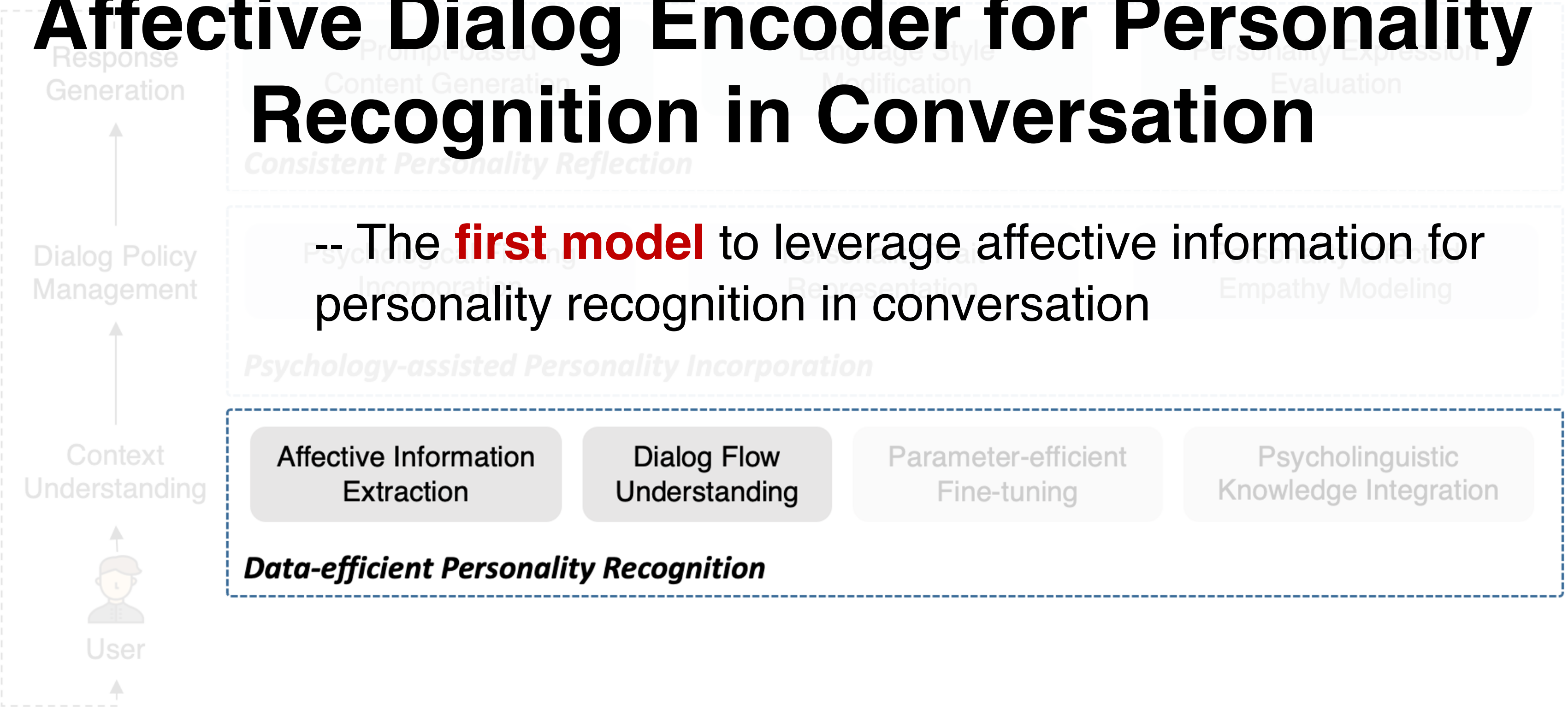
Comprehending and Reflecting Personality in Dialog Systems



Research Contribution 1: *Affective Dialog Encoder for Personality Recognition in Conversation*
Research Contribution 2: *DesPrompt: Personality-descriptive Prompt Tuning for Few-shot Personality Recognition*
Research Contribution 3: *Personality-affected Emotion Generation in Dialog Systems*
Research Contribution 4: *Decode with Template: Content Preserving Sentiment Transfer*

Affective Dialog Encoder for Personality Recognition in Conversation

-- The **first model** to leverage affective information for personality recognition in conversation

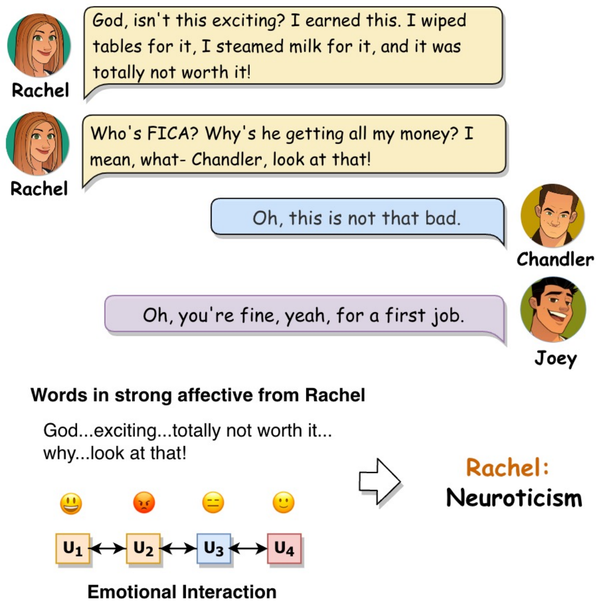


How to recognize the personality of a speaker with limited utterances in conversation?

- **Limitations in existing studies**
  - Analyze single utterances, overlook the structures of dialog flow
  - Focus on content understanding, neglect affective expressions

- **Intuition:** extracting information from multiple aspects in limited utterances for personality recognition
  - affective expressions of speakers (inspired in psychology findings)
  - emotional interactions among speakers in dialog flows

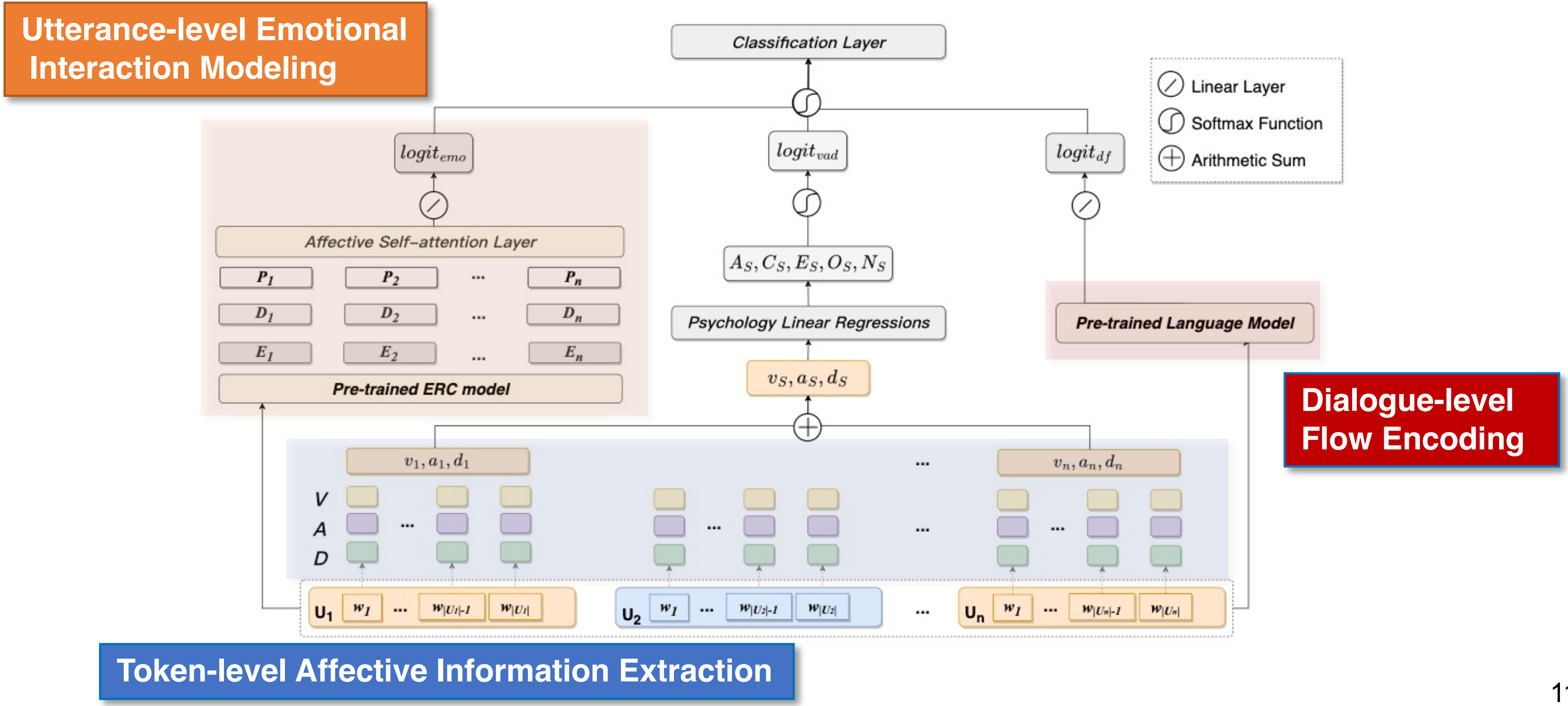
- **Challenges:** Accurately obtaining real-time affective annotations of utterances is impractical (**reason of the first**)
  - Automatic annotation with Pre-trained Emotion Recognition in Conversation (ERC) model
  - Token-level affective embeddings



- **Problem Formulation**
  - **Given:** dialog content $\boldsymbol{X} = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_n}\}$ among a speaker $\boldsymbol{s}$ and other speakers, $\boldsymbol{u_i}$ is the $\boldsymbol{i}$-th utterance
  - **Assumption:** the personality of $\boldsymbol{s}$ can be inferred from the semantics in dialog content $\boldsymbol{X}$
  - **Objective:** recognize the binary big-five personality trait of $\boldsymbol{s}$, denoted as $\boldsymbol{y}$. $\boldsymbol{y}$ is represented as a 5-$\boldsymbol{d}$ binary vector **[A, C, E, O, N]** indicating *Agreeableness, Conscientiousness, Extraversion, Openness, and Neuroticism*

Affective Dialog Encoder for Personality Recognition in Conversation

- **Model Design of Affective Dialog Encoder**



Affective Dialog Encoder for Personality Recognition in Conversation

- **Experiment Settings**
  - **Dataset:** FriendsPersona & CPED (TV series scripts)
  - **Tasks** (Evaluated by Personality Recognition **F-scores**):
    - **Overall:** takes the content of whole dialog flow for personality recognition
    - **Flow:** inputs first **25%**, **50%**, **75%**, and the **whole dialog flow**, respectively for personality recognition
  - **Baseline models:**
    - RoBERTa (S): only the speaker' utterances
    - RoBERTa (S+C): the speakers' utterances + Context
    - RoBERTa (F): the whole dialog flow
  - **Ablation sub-models:**
    - ADE (VAD): only VAD affectivity in utterances
    - ADE (EMO): only emotion interaction modeling

Dataset	FriendsPersona	CPED
#Dialogues	711	11,835
#Uttrs per dialogue	11.8	11.2
#Unique Uttrs	8,157	109,455
Uttr Length	16.3	28.9
Label Distribution (Positive:Negative)	AGR(.43:.57) CON(.46:.54) EXT(.44:.56) OPN(.35:.65) NEU(.47:.53)	AGR(.58:.42) CON(.67:.33) EXT(.65:.35) OPN(.50:.50) NEU(.59:.41)

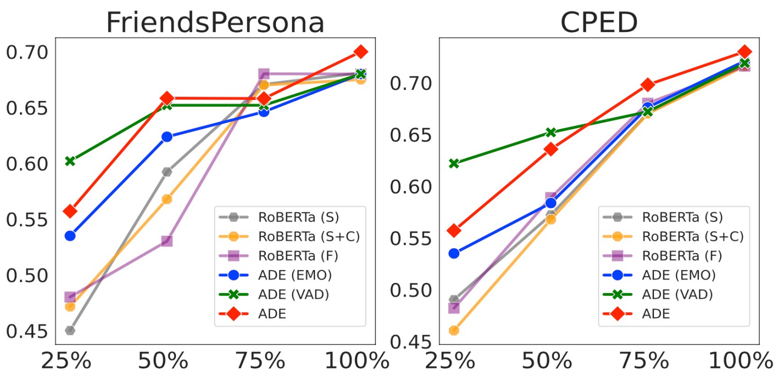
Dataset Statistics

- **Experiment Results and Findings**
  - Integrating affective information and modeling emotional interaction together enhance personality recognition in conversation
  - The impact of **utterance affective information** (**Green lines**) on personality recognition is evident in a shortage of dialog content
  - With **only one or two utterances**, ADE can instantly recognize the speaker's personality (by over 0.6 F-scores)

Dataset	Method	AGR	CON	EXT	OPN	NEU	Avg
FriendsPersona	RoBERTa (S)	.727 ± .004	.630 ± .004	.615 ± .001	.790 ± .001	.642 ± .001	.681
	RoBERTa (S+C)	.707 ± .017	.621 ± .015	.613 ± .029	.790 ± .017	.642 ± .021	.675
	RoBERTa (F)	.725 ± .002	.629 ± .001	.615 ± .001	.788 ± .007	.642 ± .002	.680
	ADE (VAD)	.743 ± .026	.633 ± .017	.616 ± .051	.790 ± .005	.635 ± .021	.685
	ADE (EMO)	.729 ± .027	.629 ± .001	.621 ± .014	.790 ± .001	.640 ± .007	.681
	ADE	**.748** ± .016	**.647** ± .040	**.626** ± .060	**.813** ± .008	**.665** ± .031	**.700**
CPED	RoBERTa (S)	.735 ± .004	.803 ± .003	.789 ± .004	.668 ± .012	.586 ± .007	.716
	RoBERTa (S+C)	.735 ± .002	.803 ± .002	.789 ± .001	.669 ± .009	.584 ± .003	.716
	RoBERTa (F)	.734 ± .004	.805 ± .008	.788 ± .006	.669 ± .015	.587 ± .011	.717
	ADE (VAD)	.734 ± .007	.803 ± .022	.789 ± .001	.664 ± .017	**.605** ± .012	.719
	ADE (EMO)	.733 ± .010	**.814** ± .005	.789 ± .009	.674 ± .014	.594 ± .019	.721
	ADE	**.759** ± .005	.812 ± .008	**.794** ± .013	**.698** ± .006	.601 ± .011	**.733**

Experiment results of Overall

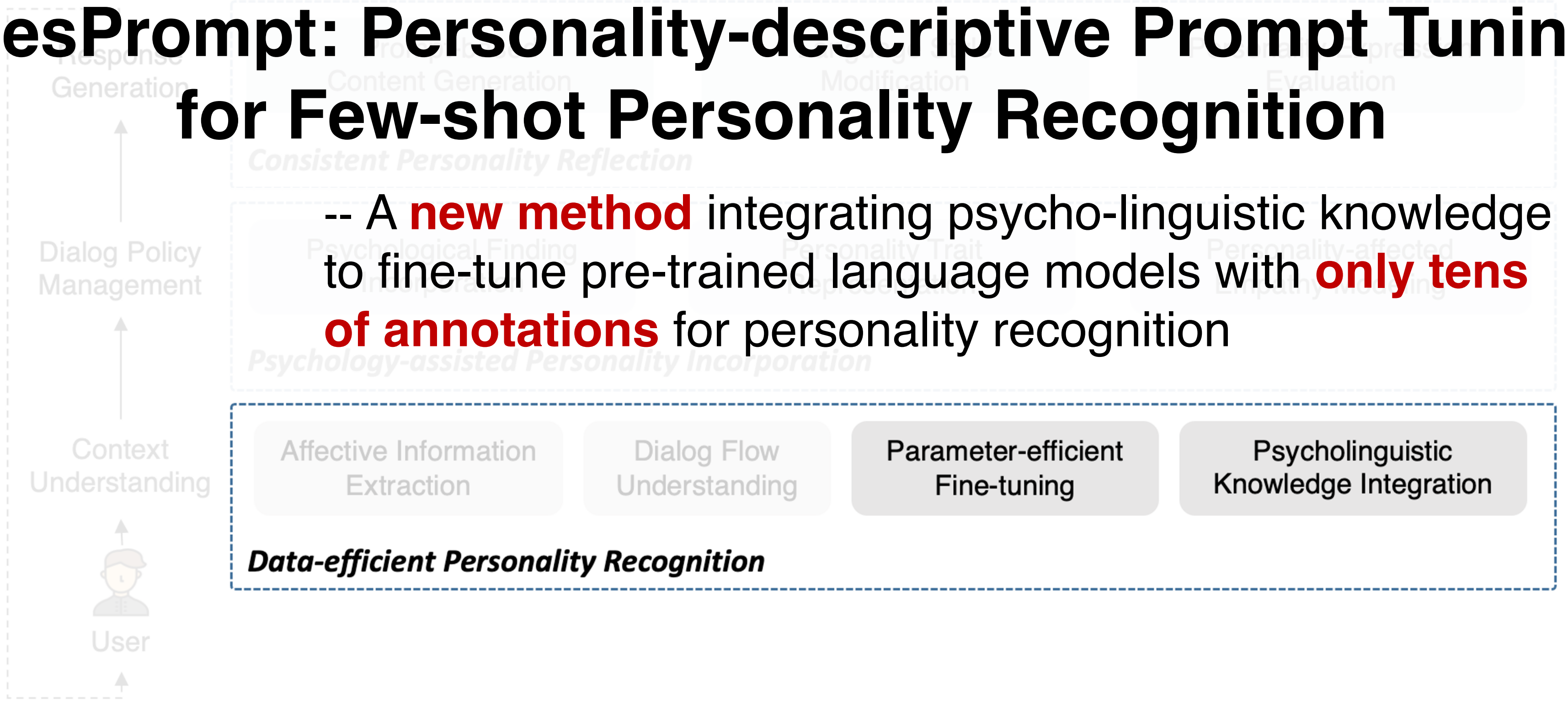
Red, **Green**, **Blue** lines are our methods



Experiment results of Flow

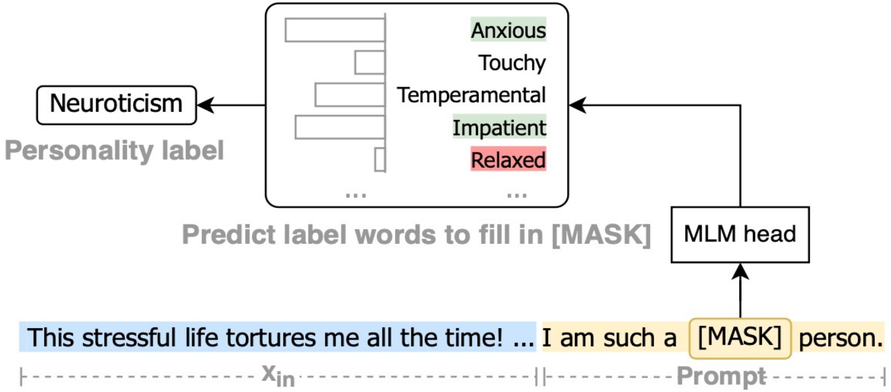
DesPrompt: Personality-descriptive Prompt Tuning for Few-shot Personality Recognition

-- A **new method** integrating psycho-linguistic knowledge to fine-tune pre-trained language models with **only tens of annotations** for personality recognition



How to recognize personality with limited labeled data for training?

- **Limitations of existing studies:**
  - Unsupervised statistical lexical analysis, lack of semantic understanding
  - Fine-tune pre-trained language models (PLMs), requires thousands of annotated samples
- **Intuition:** Encapsulate input with personality-descriptive prompts for parameter-efficient fine-tuning
  - Lexical hypothesis of personality
  - Prompt-based fine-tuning
- **Challenging issues:**
  - Finding **precise and commonly used** adjectives describing personality
  - Generating both **specific** (to each input) and **general** (commonly suitable) prompt content

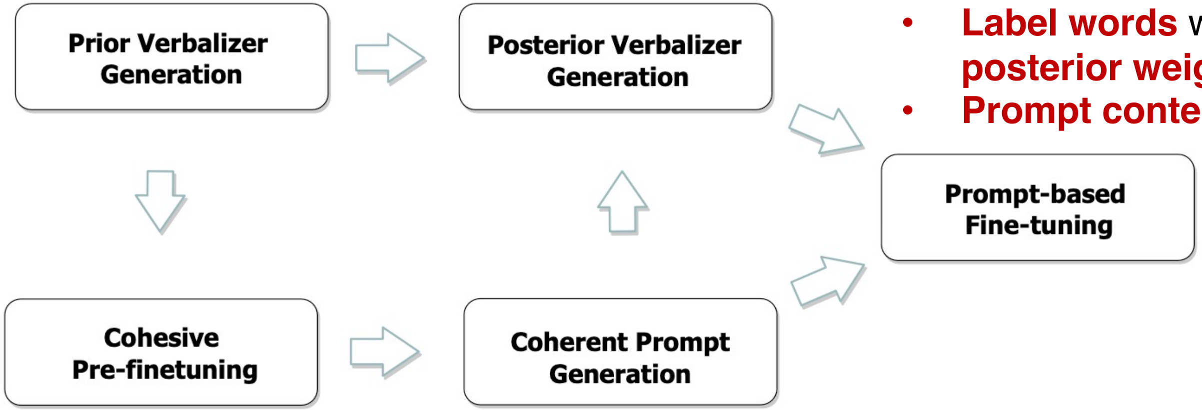


- **The overview of DesPrompt**

Label words:
- Trait-descriptive adjectives (psycho-linguistic findings)
- Synonyms and antonyms (knowledge graph)

Obtain the **posterior weight** for each label word

Prompt-based Fine-tuning PLM with:
- **Label words** with **their posterior weights**
- **Prompt content**



Pre-finetuning prompt generation model T5 to learn the appropriate context of label words

Generate **prompt content** coherent with the input

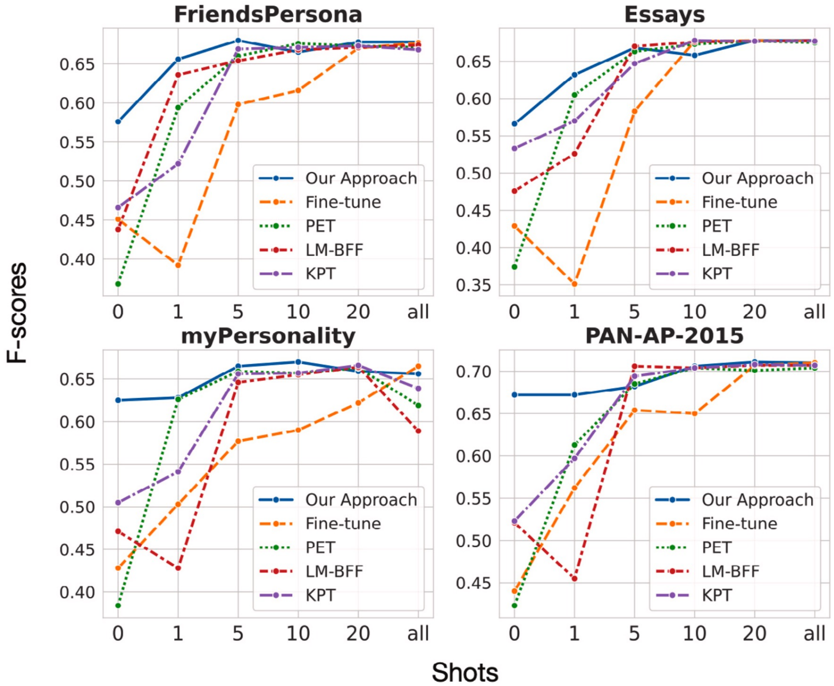
- **Experiment Settings**
  - **Dataset:** FriendsPersona & Essays & myPersonality & Pan-AP-2015
  - **Baseline models:**
    - Traditional fine-tuning: Fine-tune
    - State-of-the-art prompt-tuning: , PET, LM-BFF, KPT
  - **Tasks** (Evaluated by Personality Recognition F-scores):
    - Few-shot personality recognition

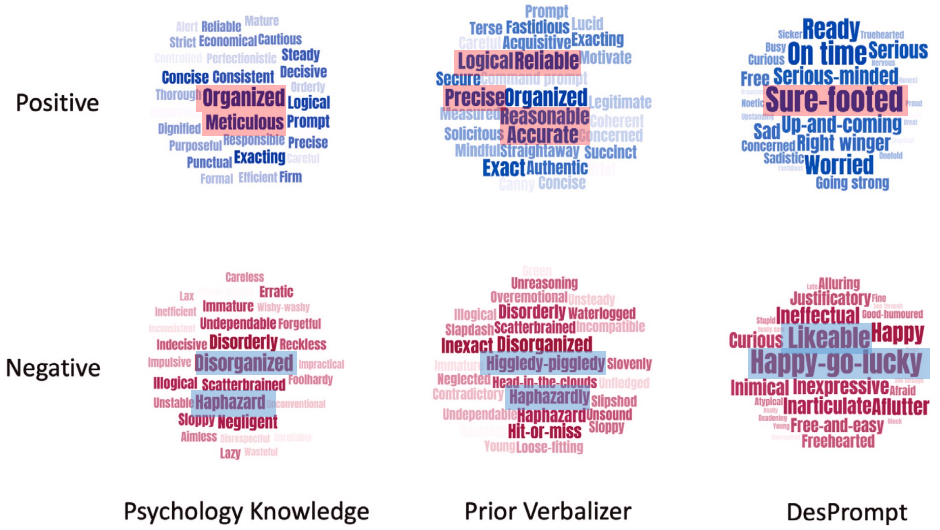
	FriendsPersona	Essays	myPersonality	PAN-AP-2015
Type	Conversation	Self-report essays	Facebook posts	Twitter posts
#Samples	711	2,467	425	658
Avg. length	48.30	662.40	321.48	464.05
AGR	0.43:0.57	0.47:0.53	0.47:0.53	0.46:0.54
CON	0.46:0.54	0.49:0.51	0.47:0.53	0.48:0.52
EXT	0.44:0.56	0.51:0.49	0.41:0.59	0.49:0.51
OPN	0.35:0.65	0.49:0.51	0.29:0.71	0.32:0.68
NEU	0.47:0.53	0.50:0.50	0.39:0.61	0.42:0.58

Basic statistics and label distributions (positive : negative) of the four datasets

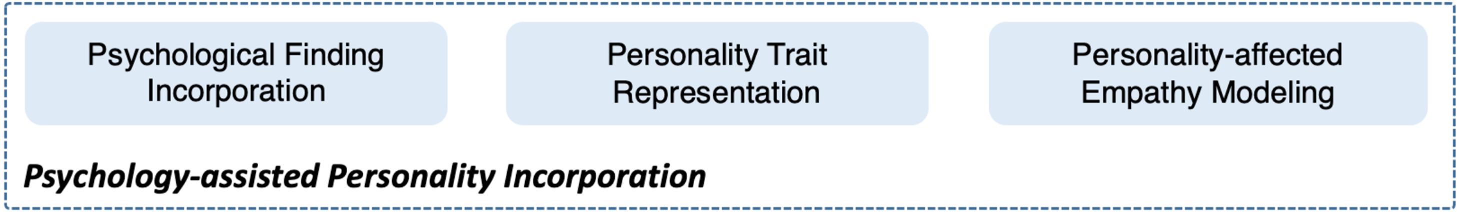
DesPrompt: Personality-descriptive Prompt Tuning for Few-shot Personality Recognition

- **Experiment Results**
  - **Quantitative:** significantly outperforms existing methods, especially in **zero-shot and few-shot** scenarios (**Blue** lines are our approach DesPrompt)
  - **Qualitative:** generates commonly used label words to describe personality





Personality-affected Emotion Generation in Dialog Systems



-- A **new research task** of Personality-affected Emotion Generation, a **new dataset** with emotion and personality annotations, and a **new model** integrated with psychology findings

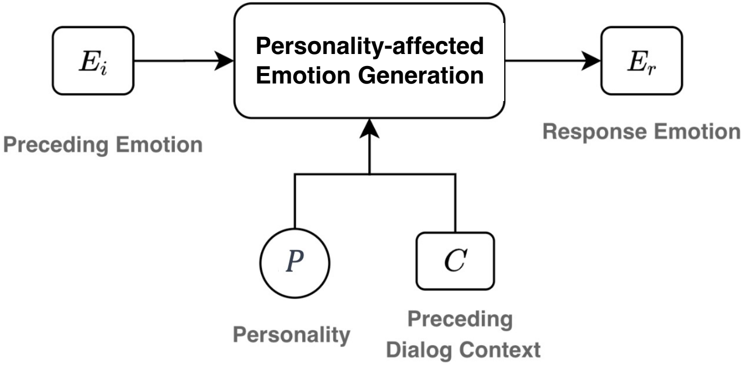
Personality-affected Emotion Generation in Dialog Systems

How to generate appropriate emotion for response to users?

- **Limitations of existing studies in emotional response generation:**
  - Render manually specified emotions rather than automatically generate emotions
  - Learn general empathetic patterns of common people, ignore individual differences

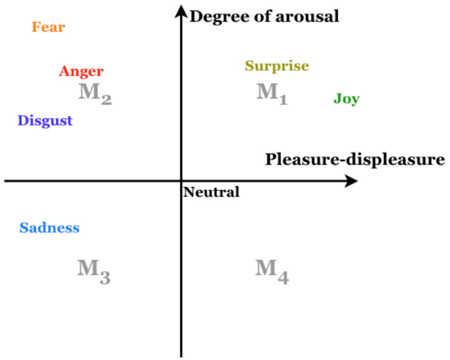
- **Intuition:** Equip dialog system with personality traits to facilitate automatic emotion generation
  - The speaker's **current emotion is derived from the preceding emotion** in conversation, and this process is **influenced by the speaker's personality** (psychological finding)

- **Challenge:** "one-to-many" nature of dialogues
  - Multiple emotions can be appropriate in a similar conversation context, only one can be selected for the response each time



Personality-affected Emotion Generation in Dialog Systems

- We raise a new task: **Personality-affected Emotion Generation**
- **Problem Definition**
  - **Given:**
    - the dialog context $\boldsymbol{C} = \{(\boldsymbol{U_1}, \boldsymbol{E_1}), (\boldsymbol{U_2}, \boldsymbol{E_2}), \ldots, (\boldsymbol{U_{n-1}}, \boldsymbol{E_{n-1}})\}$ including all the preceding $\boldsymbol{n-1}$ utterances, where $\boldsymbol{E_i}$ is the emotion label for each utterance $\boldsymbol{U_i}$
    - The specified personality trait $\boldsymbol{P}$ to the dialog system
  - **Objective:** generate an appropriate emotion $\boldsymbol{E_n}$ for the upcoming response $\boldsymbol{U_n}$ to the user
    - $\boldsymbol{E_n}$ should conform to the specified personality trait $\boldsymbol{P}$ in the current context $\boldsymbol{C}$



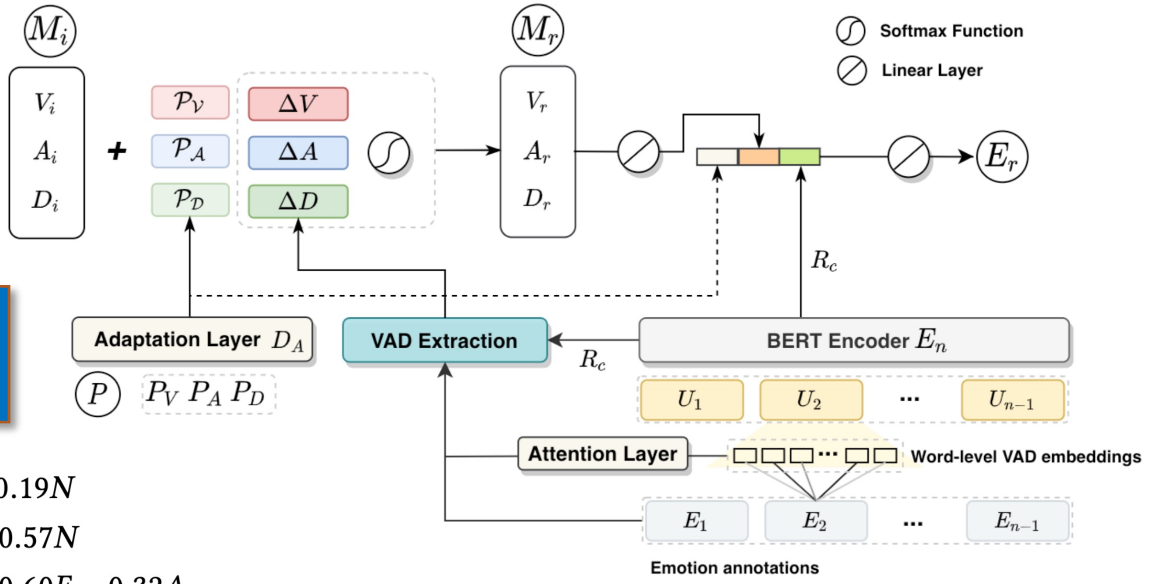
Emotions in the VAD Space

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.

The big-five personality traits and descriptions

Personality-affected Emotion Generation in Dialog Systems

- **Model Design**
  - Emotion generation: the mood state transition in the VAD space
  - Affective information in dialog content is the **variation**, personality is the **weights**
  - Coefficients in linear analysis from small groups (72 samples) → Trainable model parameters supervised by large-scale dialog data

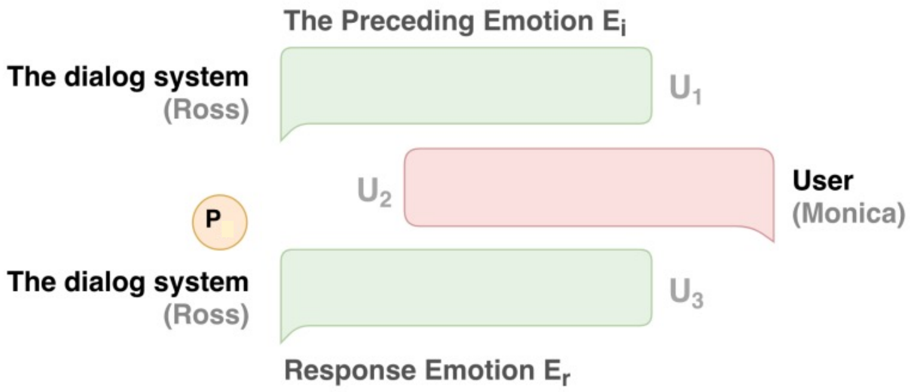


Correlations between:
- **Personalities in big-five**
- **Tendency of emotions in VAD**

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

Personality-affected Emotion Generation in Dialog Systems

- **P**ersonality **E**motion **L**ine **D**ataset (PELD)
  - An emotional dialog dataset of **6.5k** dialogues with personality annotations for speakers
  - Dialogue script of a famous TV series *Friends*



A triple example in PELD

Basic Statistics	**Train**	**Valid**	**Test**	**Total**
#Triple	5286	588	653	6527
#Unique Uttr.	9273	1529	1679	10468
Avg. Uttr. Length	9.26	9.33	8.95	9.32
#Emotions	**Train**	**Valid**	**Test**	**Total**
Anger	1857	238	247	2342
Disgust	316	30	30	376
Fear	1100	118	132	1350
Joy	2883	321	345	3549
Neutral	7066	782	880	8728
Sadness	1086	120	141	1347
Surprise	1550	155	184	1889
#Mood States	**Train**	**Valid**	**Test**	**Total**
Neutral	7066	782	880	8728
M_1	4433	476	529	5438
M_2	3273	386	409	4068
M_3	1086	120	141	1347
M_4	-	-	-	-
#Triples of Main Roles	**Train**	**Valid**	**Test**	**Total**
Chandler	864	107	117	1088
Joey	929	96	100	1125
Monica	847	95	111	1053
Phoebe	789	90	98	977
Rachel	934	97	128	1159
Ross	923	103	99	1123

Statistics in PELD

Personality-affected Emotion Generation in Dialog Systems

- **Experiment Settings:**
  - Dataset: PELD
  - Evaluation metric: F-scores of emotion generation (with statistical significance test)
  - Sub-models in ablation study: BERT, BERT-Mood, BERT-P, BERT-MT
- **Result:**
  - After integrating personality-affected mood state transition, our model achieved the best emotion generation performance (green values indicate the outperformances are statistically significant ($p < 0.05$))

Methods	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise	m-avg	w-avg
BERT	0.318	0.012	0.226	0.278	0.513	0.212	0.109	0.242	0.375
	0.05	0.02	0.29	0.05	0.03	0.03	0.03	0.03	0.03
BERT-Mood	0.252	0.113	0.227	0.248	0.468	0.288	0.107	0.242	0.344
	0.01	0.00	0.36	0.03	0.00	0.05	0.00	0.01	0.00
BERT-P	0.267	0.096	0.159	0.320	0.494	**0.299**	0.119	0.254	0.349
	0.05	0.04	0.05	0.01	0.05	0.03	0.01	0.05	0.03
BERT-MT	0.271	0.099	0.173	**0.334**	0.507	0.239	**0.127**	0.247	0.368
	0.05	0.02	0.40	0.03	0.04	0.03	0.02	0.39	0.04
Our Model	**0.323**	**0.167**	**0.229**	0.291	**0.545**	0.254	0.114	**0.269**	**0.392**

Emotion Generation F-scores

Prompt-based Content Generation

Language Style Modification

Personality Expression Evaluation

Consistent Personality Reflection

Decode with Template: Content Preserving Sentiment Transfer

-- A **new method** of incorporating lexical modification with semantic generation for effective language style modification

How to modify the language style (sentiment) without revising the remaining semantic content?

- **Limitations of existing studies:**
  - Instance-level lexical modification: disrupting the naturalness of the output content
  - Semantic disentanglement in latent space: poor sentiment transfer accuracy
- **Intuition:** Incorporate lexical modification with content generation from latent semantic space
  - Lexical modification: effective sentiment transfer
  - Semantic generation: natural content preservation

Positive to negative sentiment transfer	
Input:	I **love** this place , the service is always **great**!
Output:	I **hate** this place, the service is **bad**.

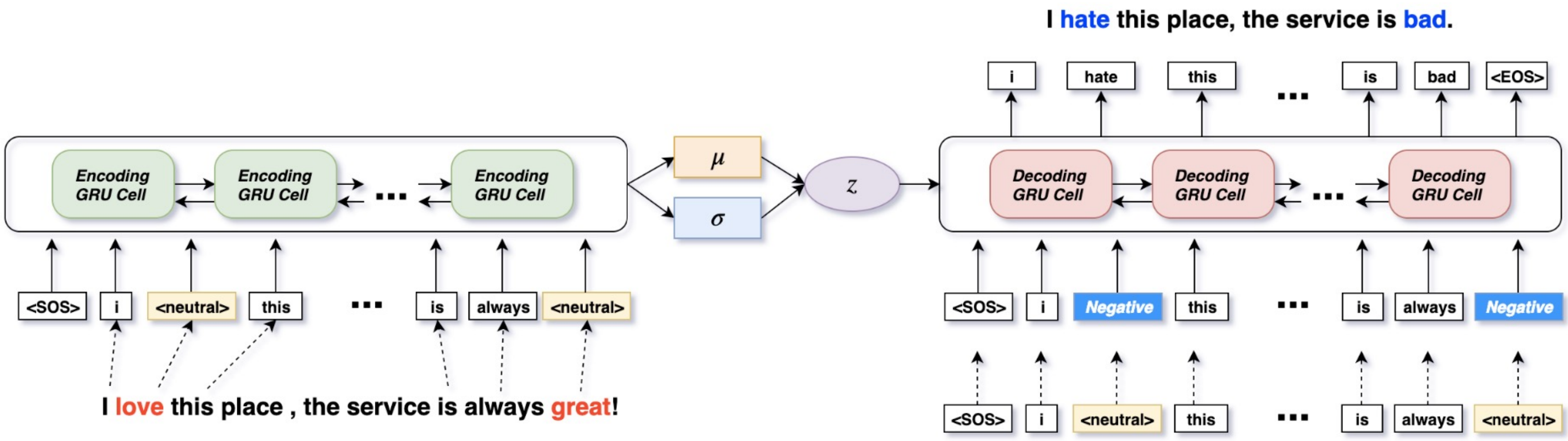
A toy example of Sentiment Transfer

Decode with Template: Content Preserving Sentiment Transfer

- **Problem Formulation:**
  - **Given:**
    - a set of sentences with sentiment labels $\boldsymbol{X} = \{(\boldsymbol{x}_1, \boldsymbol{y}_1), \ldots, (\boldsymbol{x}_n, \boldsymbol{y}_n)\}$, where $\boldsymbol{x}_i$ is a sentence whose sentiment label (either “positive” or “negative”) is indicated by $\boldsymbol{y}_i$
  - **Assumption:**
    - The sentiment of a sentence can be split from semantic content
  - **Objective:**
    - For each $\boldsymbol{x}_i$, generate a semantic coherent sentence $\widehat{\boldsymbol{x}}_i$:
      - rendering the sentiment $\widehat{\boldsymbol{y}}_i$ opposite to $\boldsymbol{y}_i$,
      - preserving the original content of $\boldsymbol{x}_i$

- **Challenge:** No parallel data (sentences with similar semantic content but different sentiments) to supervise the model for neither:
  - Sentiment transfer
  - Content preservation

- **Model Design of Decode with Template**
  - Identify and replace sentiment words with words:
    - in the opposite sentiment
    - coherent with the original context
  - Use sentiment-free template to preserve semantic content in generation



  - Supervision for sentiment transfer: classification with pre-trained sentiment classifier
  - Supervision for content preservation: reconstruction of original input

Decode with Template: Content Preserving Sentiment Transfer

- **Experiment Settings**
  - **Dataset:** Amazon & Yelp reviews
  - **Evaluation methods:**
    - Automatic evaluation & Human evaluation
    - Sentiment transfer accuracy, Content preservation, Naturalness
  - **Baseline models:**
    - Semantic disentanglement: Cross-Alignment Auto-Encoder (**CAAE**) , Control and Generation (**CtrlGen**) , Back-translation for Style Transfer (**BST**)
    - Instance-level lexical modification: **TemplateBased , DeleteAndRetrieve**
  - **Ablation study:**
    - w/o Template
    - w/o Content Representation
    - w/o Adversarial Training

Dataset	Sentiment	Train	Validation	Test
Yelp	Positive	270K	2000	500
	Negative	180K	2000	500
Amazon	Positive	277K	985	500
	Negative	278K	1015	500

Statistics in Amazon and Yelp datasets

Decode with Template: Content Preserving Sentiment Transfer

- **Results and Analysis**
  - Our method effectively transfers the sentiment while preserve the sentiment-free semantic content
  - Our method generates semantic coherent (with high Naturalness) sentences
  - The modified template (our key innovation) is a critical component to enhance content preservation

Yelp	ACC	BLEU	WMD
CAAE	0.772	4.9	11.655
CtrlGen	0.849	3.4	13.278
TemplateBased	0.849	16.3	4.122
DeleteAndRetrieve	0.903	11.3	7.651
BST	0.895	20.9	3.985
Our method	**0.930**	**25.2**	**3.126**
Amazon	**ACC**	**BLEU**	**WMD**
CAAE	0.587	5.1	10.354
CtrlGen	0.695	2.9	13.100
TemplateBased	0.703	25.6	3.290
DeleteAndRetrieve	0.640	21.3	4.058
BST	0.705	25.8	3.744
Our method	**0.752**	**27.9**	**3.2**

Automatic evaluation results

Yelp	Sentiment	Content	Naturalness
CAAE	2.379	1.605	2.506
CtrlGen	3.445	1.764	2.730
TemplateBased	3.304	3.998	2.489
DeleteAndRetrieve	2.501	3.584	3.500
BST	2.437	3.453	3.565
Our method	**3.449**	**4.173**	**3.709**
Amazon	**Sentiment**	**Content**	**Naturalness**
CAAE	2.643	1.455	2.834
CtrlGen	3.055	2.631	3.001
TemplateBased	**3.273**	3.400	2.340
DeleteAndRetrieve	2.309	3.220	3.554
BST	2.803	3.661	3.150
Our method	3.221	**3.845**	**3.669**

Human evaluation results

Yelp	Accuracy	BLEU	WMD
Our method	0.930	25.2	3.126
w/o Template	-	4.5	10.343
w/o Content Rep.	0.912	17.6	5.617
w/o Adversarial Training	0.884	22.2	3.170
Amazon	**Accuracy**	**BLEU**	**WMD**
Our method	0.752	27.9	3.281
w/o Template	-	3.7	13.600
w/o Content Rep.	0.751	20.1	4.399
w/o Adversarial Training	0.712	24.5	3.390

Ablation study results

Conclusions

- We pinpoint the limitations in **personalization**, **emotional intelligence**, and **language style consistency** within current dialogue systems

- In response to the research gap, we design a research framework for comprehending and reflecting personality in dialog systems and address issues within:
  - Personality recognition in conversation (with new methods)
  - Personality incorporation for emotion generation (with new dataset and task)
  - Language style modification (with new method)

- Our work takes a step towards creating more humanized conversational agents and improving conversational services such as empathetic companions, social chatbots, and AI-based mental therapy

Future Directions

- Expanding the range of personality traits in dialog systems
  - Big Five personality traits
  - Myers Briggs Type Indicator (MBTI)
  - Specifically designed traits
- Incorporating non-verbal cues into personality understanding and reflection
  - Facial expressions, tone of voice, gestures, physiological signals…
- Investigating ethical considerations in dialog systems
  - Ensures that the model's responses are not only accurate and coherent but also safe, ethical, and desirable from the perspective of users

List of Publications

[1] **Zhiyuan Wen,** Jiannong Cao, Ruosong Yang, Shuaiqi Liu, Jiaxing Shen, Maosong Sun. Personality-affected Emotion Generation in Dialog Systems. Minor revision submitted to IEEE Transactions on Information Systems: TOIS, 2023.

[2] **Zhiyuan Wen**, Jiannong Cao, Yu Yang, Haoli Wang, Ruosong Yang, Shuaiqi Liu. DesPrompt: Personality-descriptive prompt tuning for few-shot personality recognition. Information Processing & Management , 2023.

[3] **Zhiyuan Wen**, Jiannong Cao, Ruosong Yang, Shuaiqi Liu, and Jiaxing Shen. Automatically Select Emotion for Response via Personality-affected Emotion Transition. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021

[4] **Zhiyuan Wen**, Jiannong Cao, Ruosong Yang, Senzhang Wang. Decode with Template: Content Preserving Sentiment Transfer. Proceedings of the Twelfth Language Resources and Evaluation Conference: LREC 2020.

[5] Yu Yang*, **Zhiyuan Wen***, Jiannong Cao, Jiaxing Shen, Hongzhi Yin, Xiaofang Zhou. EPARS: Early Prediction of At-risk Students with Online and Offline Learning Behaviors. 25th International Conference on Database Systems for Advanced Applications: DASFAA 2020

[6] Shuaiqi Liu, Jiannong Cao, Zhongfen Deng, Wenting Zhao, Ruosong Yang, **Zhiyuan Wen**, S. Yu Philip. Neural Abstractive Summarization for Long Text and Multiple Tables. IEEE Transactions on Knowledge and Data Engineering: TKDE, 2023

[7] Shuaiqi Liu, Jiannong Cao, Ruosong Yang, **Zhiyuan Wen**. Key phrase aware transformer for abstractive summarization. Information Processing & Management IPM, 2022

[8] Ruosong Yang, Jiannong Cao, **Zhiyuan Wen**, and Jiaxing Shen. Automated post scoring: Evaluating posts with topics and quoted posts in online forum. World Wide Web Journal, 2022.

[9] Shuaiqi Liu, Jiannong Cao, Ruosong Yang, **Zhiyuan Wen**. Long Text and Multi- Table Summarization: Dataset and Method. Findings of the Association for Computational Linguistics: EMNLP 2022

[10] Shuaiqi Liu, Jiannong Cao, Ruosong Yang, **Zhiyuan Wen**. Generating a Structured Summary of Numerous Academic Papers: Dataset and Method. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence: IJCAI-22

[11] Shuaiqi Liu, Jiannong Cao, Ruosong Yang, **Zhiyuan Wen**. Highlight-Transformer: Leveraging Key Phrase Aware Attention to Improve Abstractive Multi-Document Summarization. Findings of the Association for Computational Linguistics: ACL- IJCNLP 2021

[12] Yu Yang, Jiannong Cao, Jiaxing Shen, Ruosong Yang, **Zhiyuan Wen**. Learning Analytics Based on Multilayer Behavior Fusion. Blended Learning. Education in a Smart Learning Environment: 13th International Conference: ICBL 2020.

[13] Ruosong Yang, Jiannong Cao, **Zhiyuan Wen**, Youzheng Wu, Xiaodong He. Enhancing Automated Essay Scoring Performance via Fine-tuning Pre-trained Language Models with Combination of Regression and Ranking. Findings of the Association for Computational Linguistics: EMNLP 2020

[14] Ruosong Yang, Jiannong Cao, **Zhiyuan Wen**. GGP: Glossary Guided Post-processing for Word Embedding Learning. Proceedings of the Twelfth Language Resources and Evaluation Conference: LREC 2020.


thank you