Detecting Speaker Personas from Conversational Texts

Jia-Chen Gu[1], Zhen-Hua Ling[1], Yu Wu[2], Quan Liu[1,3], Zhigang Chen[3], Xiaodan Zhu[4]

[1]National Engineering Laboratory for Speech and Language Information Processing, University of Science and Technology of China

[2]Microsoft Research Asia

[3]State Key Laboratory of Cognitive Intelligence, iFLYTEK Research

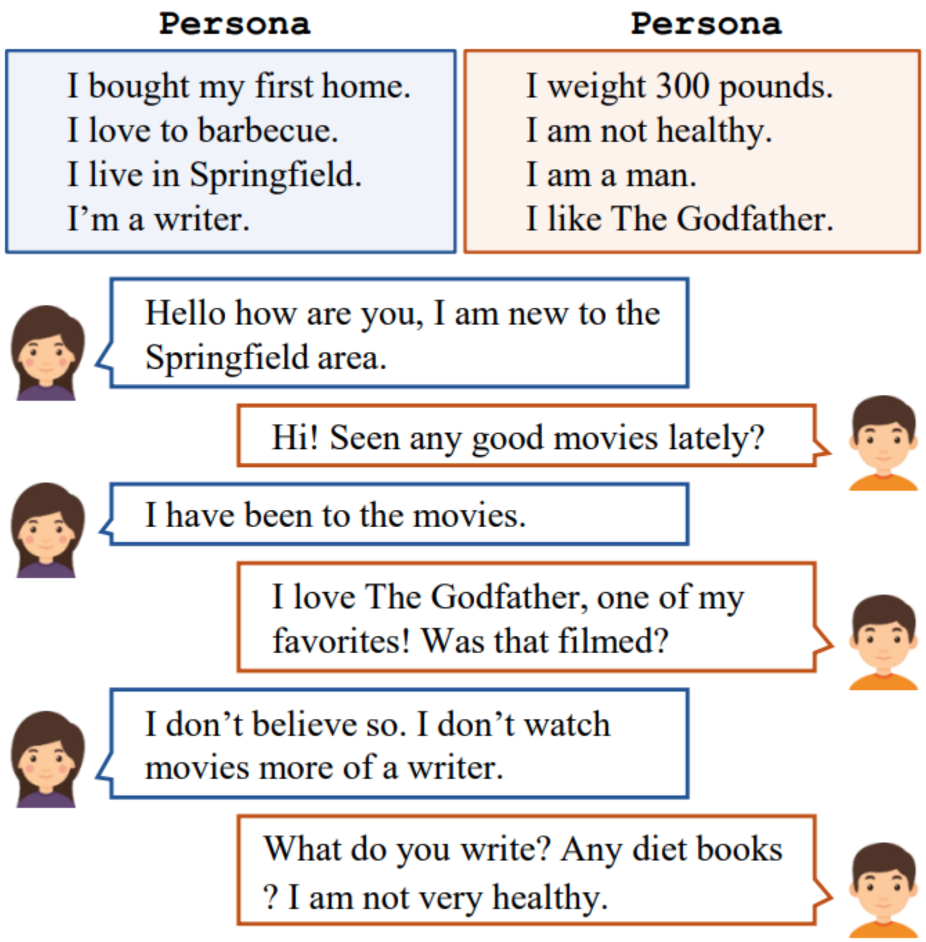
[4]ECE & Ingenuity Labs, Queen's University

Outline

- **Introduction**
- Speaker Persona Detection
- Persona Match on Persona-Chat (PMPC) Dataset
- Models
- Experiments
- Conclusion

Persona-Based Dialogue

- It is well-known that a user's persona can help machines to generate more appropriate and personalized responses.



Cold-Start Problem

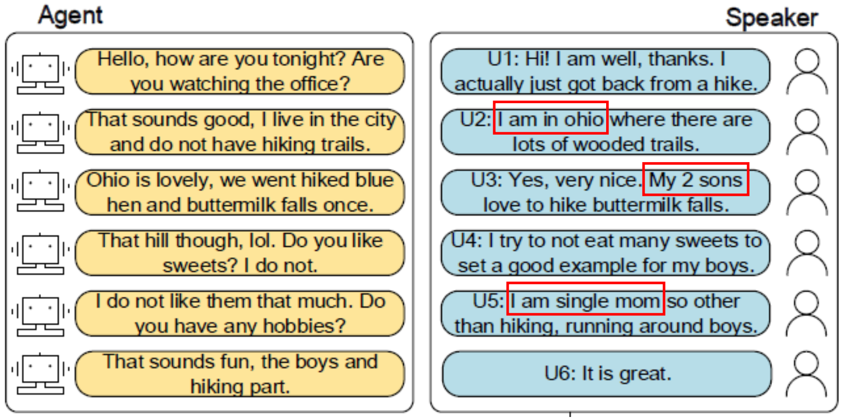
- These personas are pre-defined and difficult to obtain before a conversation.
- Speakers might not want to fill out a specific table to show its persona due to privacy issues.
- Hence, the cold-start problem may hinder the persona-aware response prediction in practice.

Outline

- Introduction
- **Speaker Persona Detection**
- Persona Match on Persona-Chat (PMPC) Dataset
- Models
- Experiments
- Conclusion

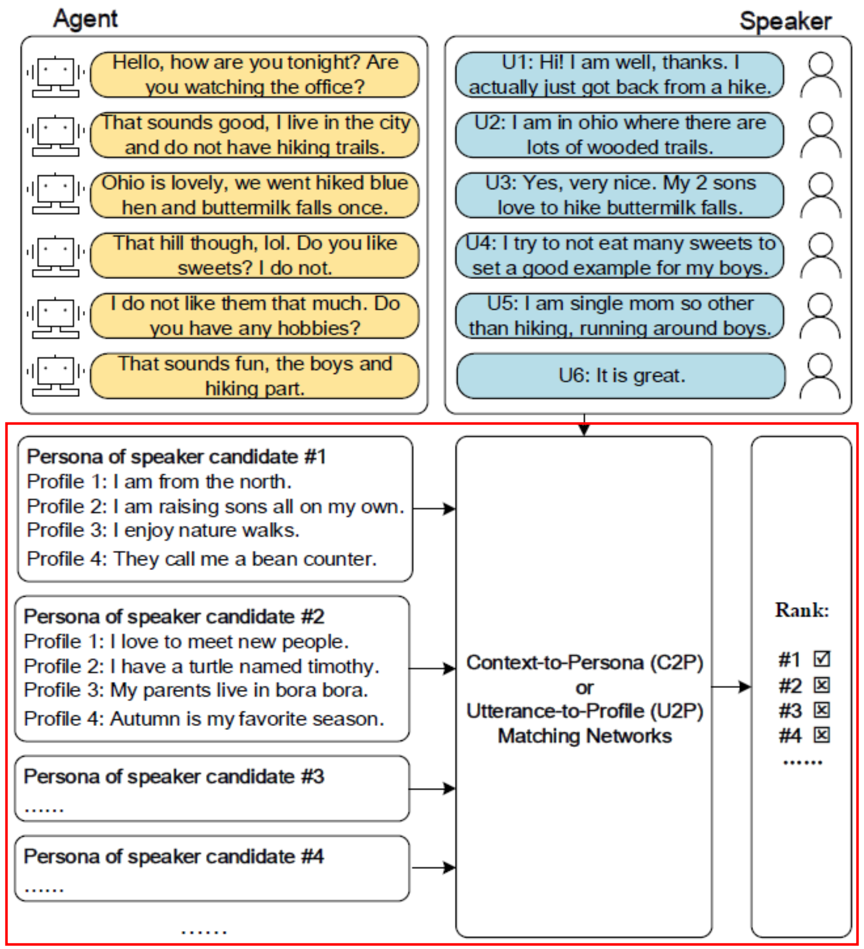
Task Description

- The personal information may be mentioned explicitly or implicitly during a conversation, which can be utilized to identify the speaker's persona.



Task Description

- If we can get the persona from early conversations, it can be utilized for future persona-aware response prediction.



Formalization

- The task of SPD is defined as selecting a best-matched persona from a list of candidates according to the conversational texts of the speaker. The candidate set is composed of one correct persona and *N* incorrect personas (distractors).

- Here, a persona description is composed of several profiles characterizing a person, which is unstructured and common in practice.

Challenges

- Long-term dependency among conversation utterances.
- A new many-to-many matching between two sets of sentences.
- Dynamic redundancy among conversation utterances and persona profiles.

Outline

- Introduction
- Speaker Persona Detection
- **Persona Match on Persona-Chat (PMPC) Dataset**
- Models
- Experiments
- Conclusion

Dataset Creation

- Based on an existing Persona-Chat dataset (Zhang et al., 2018).
- Steps:
  - Each dialogue in Persona-Chat was performed between two speakers, we can consider one of them as human speaker and the other as intelligent agent.
  - Exchange roles with each other.
  - Each dialogue can provide two matched context-persona pairs.
  - Adopt the revised version of dataset to make the SPD task more challenging.
- Two experimental settings
  - 9 and 99 distractors are used to construct the validation and test sets.

Dataset Statistics

		Train	Valid	Test
10@1	# distractors (N)	1	9	9
100@1	# distractors (N)	1	99	99
# matched context-persona pairs		18K	2K	2K
Avg. # utterances per context		7.35	7.80	7.76
Avg. # words per utterance		11.67	11.94	11.79
Avg. # profiles per persona		4.50	4.49	4.50
Avg. # words per profile		7.32	7.82	7.56

Outline

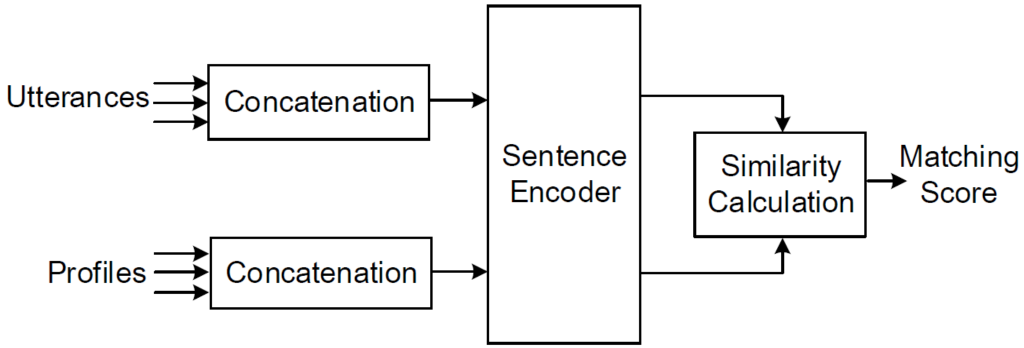
- Introduction
- Speaker Persona Detection
- Persona Match on Persona-Chat (PMPC) Dataset
- **Models**
- Experiments
- Conclusion

Models

- Frameworks
  - Sentence-encoding-based: BOW, BiLSTM and Transformer
  - Cross-attention-based: ESIM
  - Pretraining-based: BERT
- Matching granularity
  - Context-to-persona (C2P): established at a coarse granularity by concatenating two sets of sentences respectively.
  - Utterance-to-profile (U2P): established at a fine granularity by first obtaining the representation for each sentence and then derive the representations of contexts and personas through aggregation.

Sentence-Encoding-Based Models

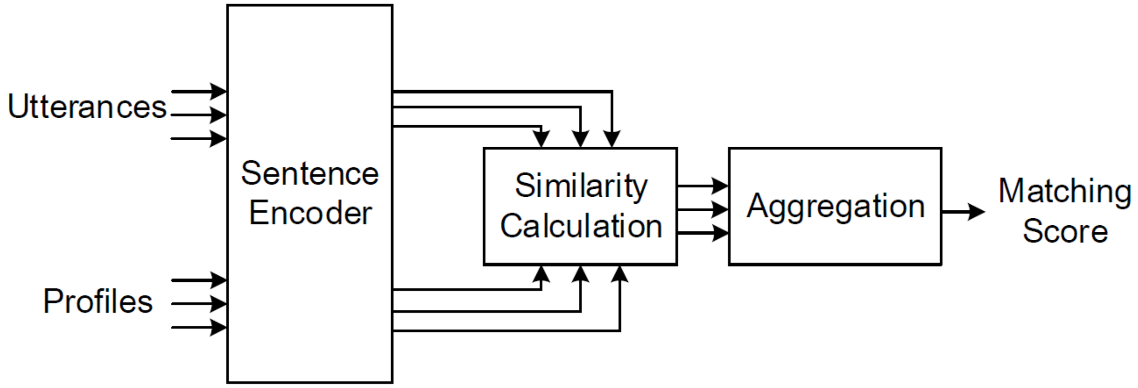
- C2P-BOW/BiLSTM/Transformer
  - BOW is employed to explore whether simple n-gram overlap could solve this task easily.
  - BiLSTM and Transformer are employed to discuss the impact of chronological (BiLSTM) or parallel (Transformer) encoding on this task.



(a) C2P-BOW/BiLSTM/Transformer

Sentence-Encoding-Based Models

- U2P-BOW/BiLSTM/Transformer
  - Each utterance and profile is encoded in parallel and separately by one of BOW, BiLSTM or Transformer encoder.
  - A similarity score is computed for each utterance-profile pair.
  - An aggregation is performed to obtain the matching score between the whole set of utterances and the whole set of profiles.



(b) U2P-BOW/BiLSTM/Transformer

Sentence-Encoding-Based Models

- Aggregation
  - Assumption: One utterance can only reflect one profile.
  - For a given utterance, its matching score with the persona is defined as the maximum matching score between it and all profiles.
  - Finally, we accumulate the matching scores of all utterances and derive the final matching score.

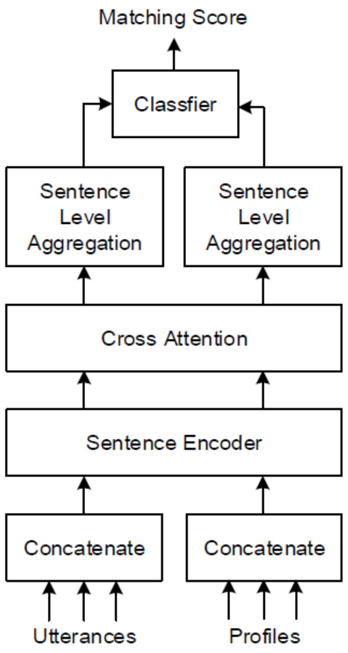
$$s_m = \max\{\max_n s_{mn}, 0\},$$

$$s = \sum_{m=1}^{n_c} s_m,$$

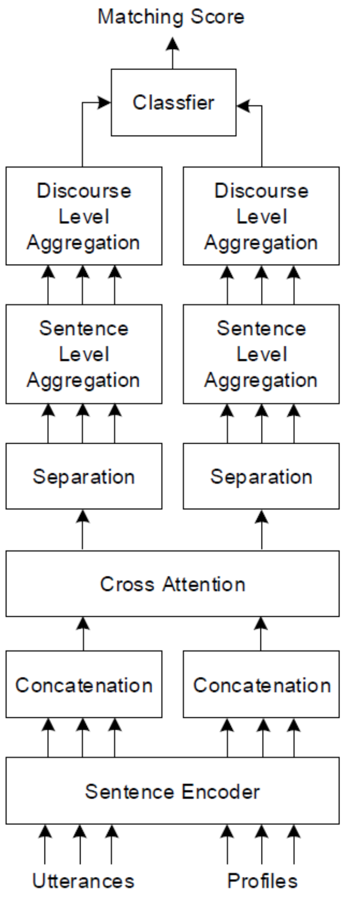
$$g(c, p) = \sigma(s),$$

Cross-Attention-Based Models

- C2P-ESIM
  - Similar to the original ESIM.
- U2P-ESIM
  - Separate encoding.
  - Concatenation for interacting and matching.
  - Separation for aggregation.



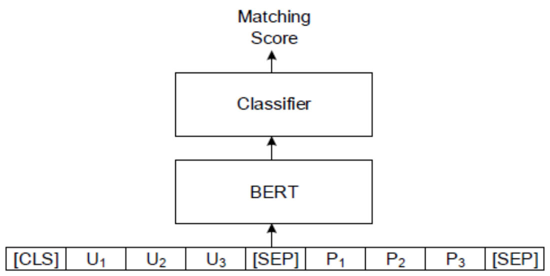
(a) C2P-ESIM



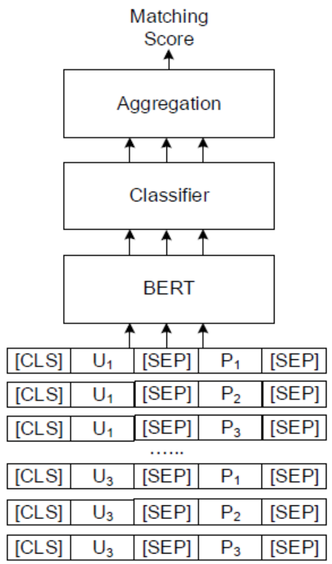
(b) U2P-ESIM

Pretraining-Based Models

- C2P-BERT
  - Utterances are concatenated to form the sentence A, and profiles are concatenated to form the sentence B.
- U2P-BERT
  - Finer interacting and matching between each utterance and each profile.
  - A specific utterance is used concatenated with all profiles.
  - Encode each utterance-profile pair.
  - Aggregation.



(a) C2P-BERT



(b) U2P-BERT

Outline

- Introduction
- Speaker Persona Detection
- Persona Match on Persona-Chat (PMPC) Dataset
- Models
- **Experiments**
- Conclusion

Metrics

- The recall of true positive replies by selecting *k* best-matched response from *n* available candidates for the given context and knowledge, denoted as $R_n@k$.
- Mean reciprocal rank , the average of reciprocal ranks of retrieval results among *n* available candidates, denoted as MRR_n.

Overall Performance

Model	$\mathbf{R}_{10}$@1	$\mathbf{MRR}_{10}$	$\mathbf{R}_{100}$@1	$\mathbf{MRR}_{100}$
C2P-BOW	34.7 ± 1.2	54.4 ± 0.9	8.9 ± 0.5	19.5 ± 0.5
U2P-BOW	46.5 ± 1.7	63.3 ± 1.3	16.9 ± 1.2	28.5 ± 1.2
C2P-BiLSTM	38.3 ± 1.2	57.7 ± 0.9	8.1 ± 0.8	19.2 ± 0.9
U2P-BiLSTM	57.4 ± 1.4	71.0 ± 1.4	24.0 ± 1.6	37.5 ± 1.6
C2P-Transformer	49.6 ± 3.7	65.3 ± 2.5	19.0 ± 1.5	30.5 ± 1.1
U2P-Transformer	56.2 ± 1.5	70.6 ± 1.1	22.9 ± 1.3	36.0 ± 1.3
C2P-ESIM	80.7 ± 0.5	87.7 ± 0.4	50.7 ± 1.4	62.8 ± 0.7
U2P-ESIM	81.6 ± 1.0	88.4 ± 0.6	54.5 ± 1.3	66.6 ± 0.7
C2P-BERT	87.4 ± 0.7	91.8 ± 0.4	64.7 ± 1.5	75.4 ± 0.8
U2P-BERT	90.4 ± 0.5	94.3 ± 0.2	79.1 ± 0.9	83.2 ± 0.5

All U2P models outperformed their C2P counterparts on all metrics.

Aggregation Method

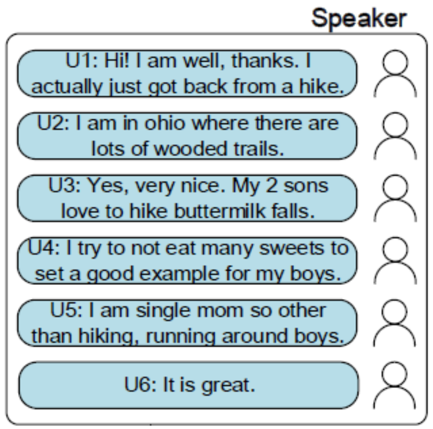
- Ablate aggregation operation over the set of profiles and utterances:
  - Max achieved better performance of aggregating profiles than Sum, supporting our assumption that one utterance reflect only one profile.
  - Sum achieved better performance of aggregating utterances than Max, indicating that multiple utterances should be considered when deriving the matching score for a context-persona pair.

Aggregation Strategy	**MRR**$_{10}$	**R**$_{10}$@1
Ps-*Max* & Us-*Sum*	71.0 ± 1.4	57.4 ± 1.4
Ps-*Max* & Us-*Max*	70.0 ± 1.3	53.7 ± 1.9
Ps-*Sum* & Us-*Max*	57.3 ± 0.8	37.1 ± 1.0
Ps-*Sum* & Us-*Sum*	67.5 ± 0.7	51.5 ± 1.1

Table 3: Evaluation results (%) of U2P-BiLSTM models with different aggregation strategies on the test set of PMPC ($N = 9$). Ps denotes Profiles and Us denotes Utterances. *Max* and *Sum* denote the aggregation operation used in Eq. (4) and Eq. (5).

Case Study

- Utterance-profile similarity scores for a matched context-persona pair, illustrating the interpretability of the aggregation operation.



Persona of speaker candidate #1
Profile 1: I am from the north.
Profile 2: I am raising sons all on my own.
Profile 3: I enjoy nature walks.
Profile 4: They call me a bean counter.

	U1	U2	U3	U4	U5	U6
P1	-0.07	-0.35	-0.22	-0.70	-1.05	-0.19
P2	-0.16	0.90	0.72	-0.20	0.38	-0.34
P3	0.83	1.14	1.00	-0.48	0.05	-0.10
P4	-0.92	-1.17	-0.89	-0.64	-2.21	-0.09
s_m	0.83	1.14	1.00	0.0	0.38	0.0

Table 4: Utterance-profile similarity scores for the matched context-persona pair shown in Figure 1. Here, Um and Pn denote the m-th utterance and the n-th profile respectively.

Time Complexity

- Parallel encoding of multiple sequences in U2P networks can improve the efficiency of RNN-based sentence encoders but can not benefit the BOW-based or Transformer-based ones.
- U2P-BERT took more time than C2P-BERT as the calculation of former is an order of magnitude higher than the latter.

Model	Time (s)	Parameters
C2P-BOW	7.1	90k
U2P-BOW	8.6	90k
C2P-BiLSTM	17.1	962K
U2P-BiLSTM	12.2	962K
C2P-Transformer	8.3	271K
U2P-Transformer	10.3	271K
C2P-ESIM	36.7	4.1M
U2P-ESIM	22.4	5.7M
C2P-BERT	121.3	110M
U2P-BERT	742.8	110M

Table 5: The inference time over the validation set of PMPC whose configuration of N was 9 using different models, together with their numbers of parameters.

Space Complexity

- C2P-BOW/BiLSTM/Transformer/BERT contained the same number of parameters with their U2P counterparts, since the additional aggregation in these U2P models consume only the calculation of Max or Sum functions, while do not require additional parameters.
- U2P-ESIM adopted an additional BiLSTM for discourse-level aggregation, and thus contained more parameters than C2P-ESIM。

Model	Time (s)	Parameters
C2P-BOW	7.1	90k
U2P-BOW	8.6	90k
C2P-BiLSTM	17.1	962K
U2P-BiLSTM	12.2	962K
C2P-Transformer	8.3	271K
U2P-Transformer	10.3	271K
C2P-ESIM	36.7	4.1M
U2P-ESIM	22.4	5.7M
C2P-BERT	121.3	110M
U2P-BERT	742.8	110M

Table 5: The inference time over the validation set of PMPC whose configuration of N was 9 using different models, together with their numbers of parameters.

Outline

- Introduction
- Speaker Persona Detection
- Persona Match on Persona-Chat (PMPC) Dataset
- Models
- Experiments
- **Conclusion**

Conclusion

- We propose the task of Speaker Persona Detection (SPD) and build a PMPC dataset for studying this task. The ability to learn speakers' personas can have wide applications in commercial chatbots, recommendation systems and other scenarios that involve conversations.
- It is beneficial to treat both contexts and personas as sets of multiple sequences in the many-to-many matching task.

Jia-Chen Gu

Zhen-Hua Ling

Yu Wu

Quan Liu

Zhigang Chen

Xiaodan Zhu

Thanks! Q&A

Code: https://github.com/JasonForJoy/SPD