NLP&AI 연구실 세미나 (08/31, Thu)

Exploring persona overuse problem in LLM

김진성

Natural Language Processing & Artificial Intelligence

고려대학교 KOREA UNIVERSITY

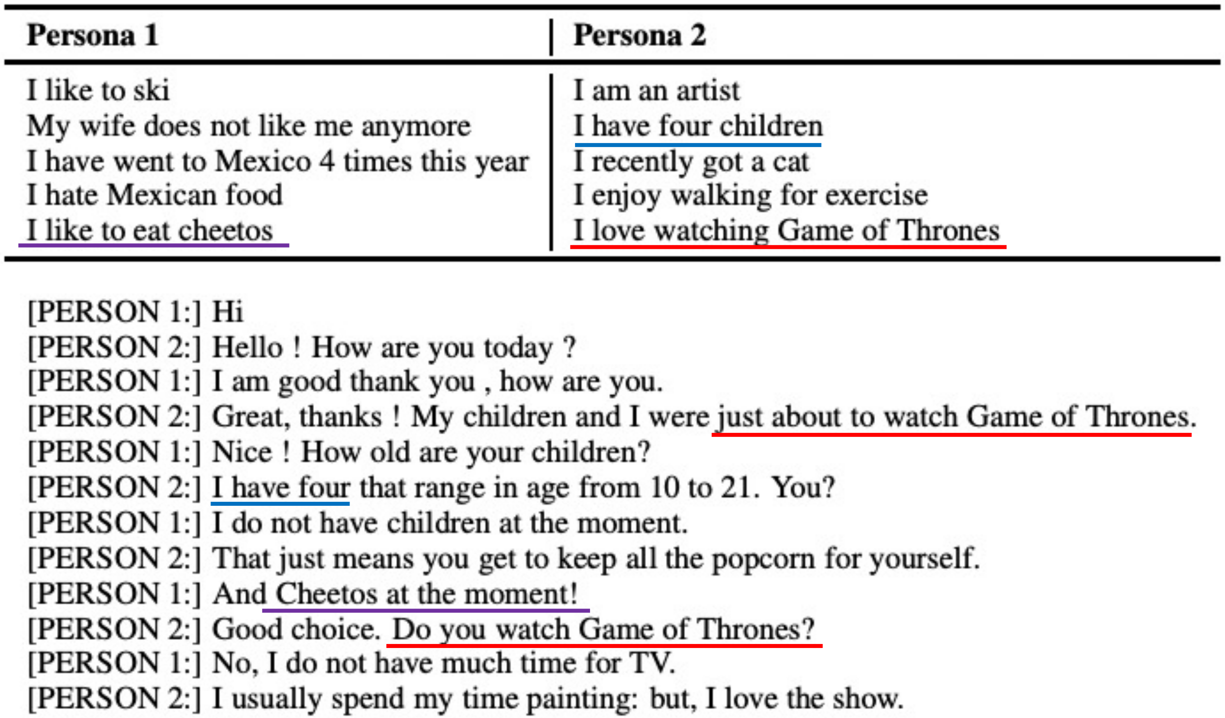
Persona-grounded chat

* PersonaChat (2018)

- 두 종류의 persona 등장
  . Self – Partner persona

- 주어진 페르소나를 대화에서
  모두 사용해야하는 것은 아님.

→ 대화 Contexts에 맞게.



Persona 1	**Persona 2**
I like to ski	I am an artist
My wife does not like me anymore	I have four children
I have went to Mexico 4 times this year	I recently got a cat
I hate Mexican food	I enjoy walking for exercise
I like to eat cheetos	I love watching Game of Thrones

[PERSON 1:] Hi
[PERSON 2:] Hello ! How are you today ?
[PERSON 1:] I am good thank you , how are you.
[PERSON 2:] Great, thanks ! My children and I were just about to watch Game of Thrones.
[PERSON 1:] Nice ! How old are your children?
[PERSON 2:] I have four that range in age from 10 to 21. You?
[PERSON 1:] I do not have children at the moment.
[PERSON 2:] That just means you get to keep all the popcorn for yourself.
[PERSON 1:] And Cheetos at the moment!
[PERSON 2:] Good choice. Do you watch Game of Thrones?
[PERSON 1:] No, I do not have much time for TV.
[PERSON 2:] I usually spend my time painting: but, I love the show.

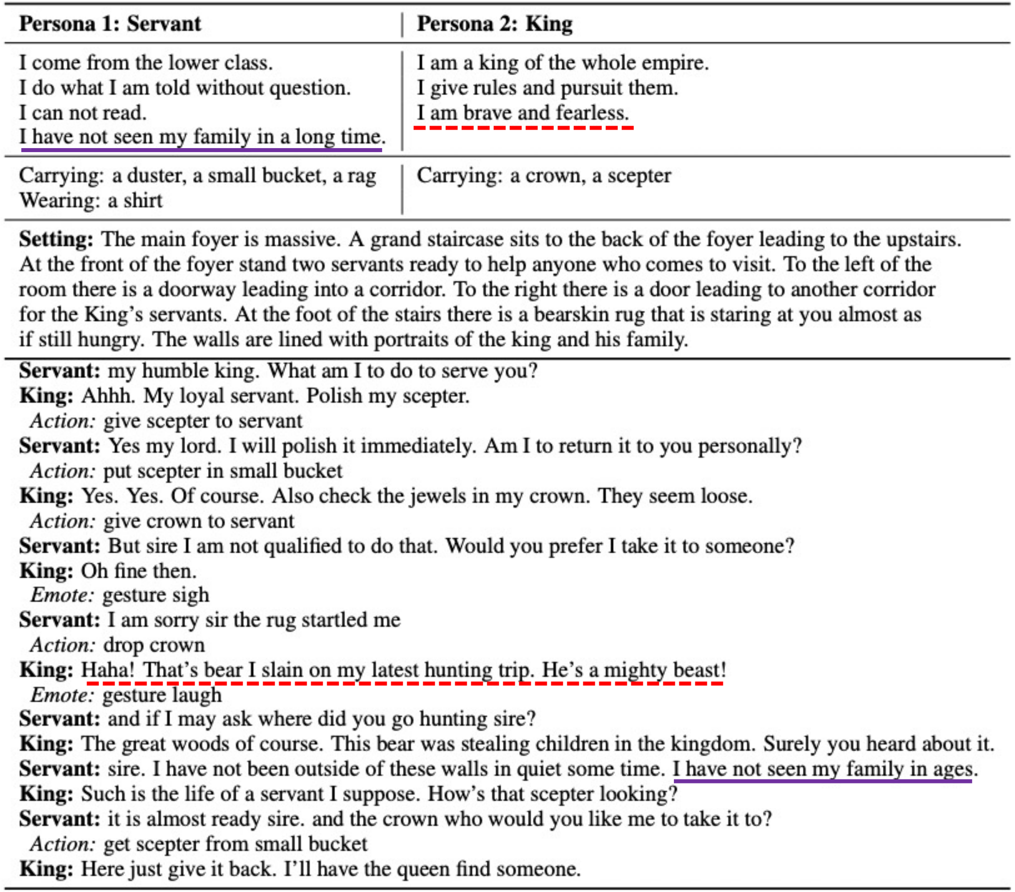
Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Persona-grounded chat

* LIGHT (2019)

- 가상의 캐릭터 설정 (Role-playing)
- 대화가 이루어지는 메타 요소들 추가 (narratives, appearance, etc.)

→ 발화에 직접적으로 나오기 쉽지 않은 persona 포함



Persona 1: Servant	**Persona 2: King**
I come from the lower class. I do what I am told without question. I can not read. I have not seen my family in a long time.	I am a king of the whole empire. I give rules and pursuit them. I am brave and fearless.
Carrying: a duster, a small bucket, a rag Wearing: a shirt	Carrying: a crown, a scepter

Setting: The main foyer is massive. A grand staircase sits to the back of the foyer leading to the upstairs. At the front of the foyer stand two servants ready to help anyone who comes to visit. To the left of the room there is a doorway leading into a corridor. To the right there is a door leading to another corridor for the King's servants. At the foot of the stairs there is a bearskin rug that is staring at you almost as if still hungry. The walls are lined with portraits of the king and his family.

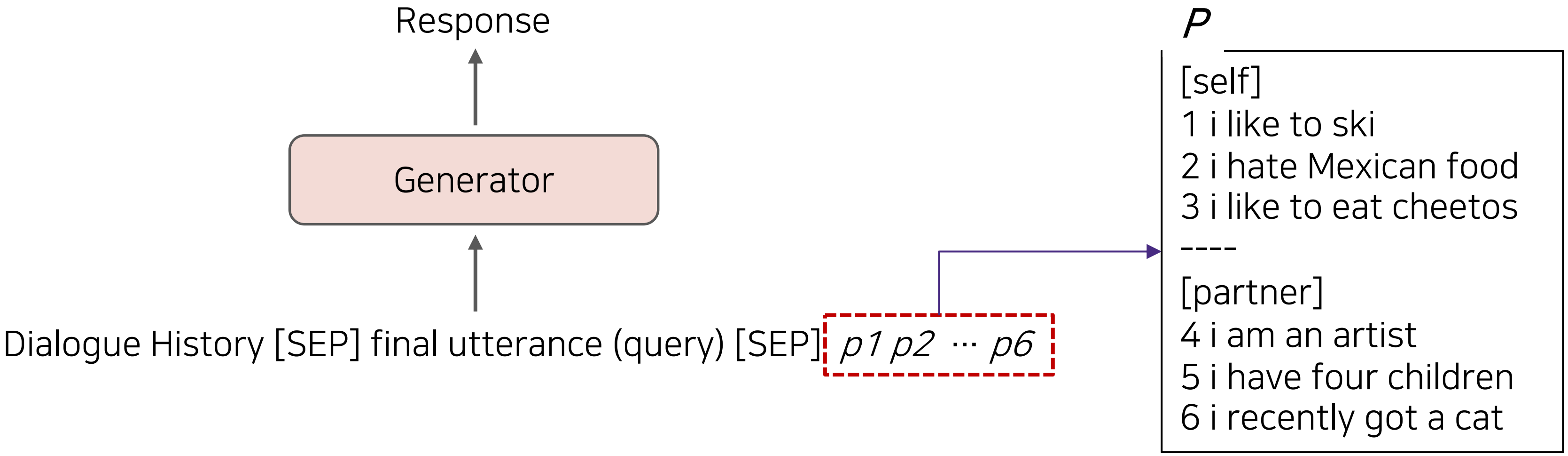
Servant: my humble king. What am I to do to serve you?
King: Ahhh. My loyal servant. Polish my scepter.
Action: give scepter to servant
Servant: Yes my lord. I will polish it immediately. Am I to return it to you personally?
Action: put scepter in small bucket
King: Yes. Yes. Of course. Also check the jewels in my crown. They seem loose.
Action: give crown to servant
Servant: But sire I am not qualified to do that. Would you prefer I take it to someone?
King: Oh fine then.
Emote: gesture sigh
Servant: I am sorry sir the rug startled me
Action: drop crown
King: Haha! That's bear I slain on my latest hunting trip. He's a mighty beast!
Emote: gesture laugh
Servant: and if I may ask where did you go hunting sire?
King: The great woods of course. This bear was stealing children in the kingdom. Surely you heard about it.
Servant: sire. I have not been outside of these walls in quiet some time. I have not seen my family in ages.
King: Such is the life of a servant I suppose. How's that scepter looking?
Servant: it is almost ready sire. and the crown who would you like me to take it to?
Action: get scepter from small bucket
King: Here just give it back. I'll have the queen find someone.

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Persona-grounded chat

*** With PLMs**

- 기존 기본 구조



Response

Generator

Dialogue History [SEP] final utterance (query) [SEP] *p1 p2* ··· *p6*

P

[self]
1 i like to ski
2 i hate Mexican food
3 i like to eat cheetos

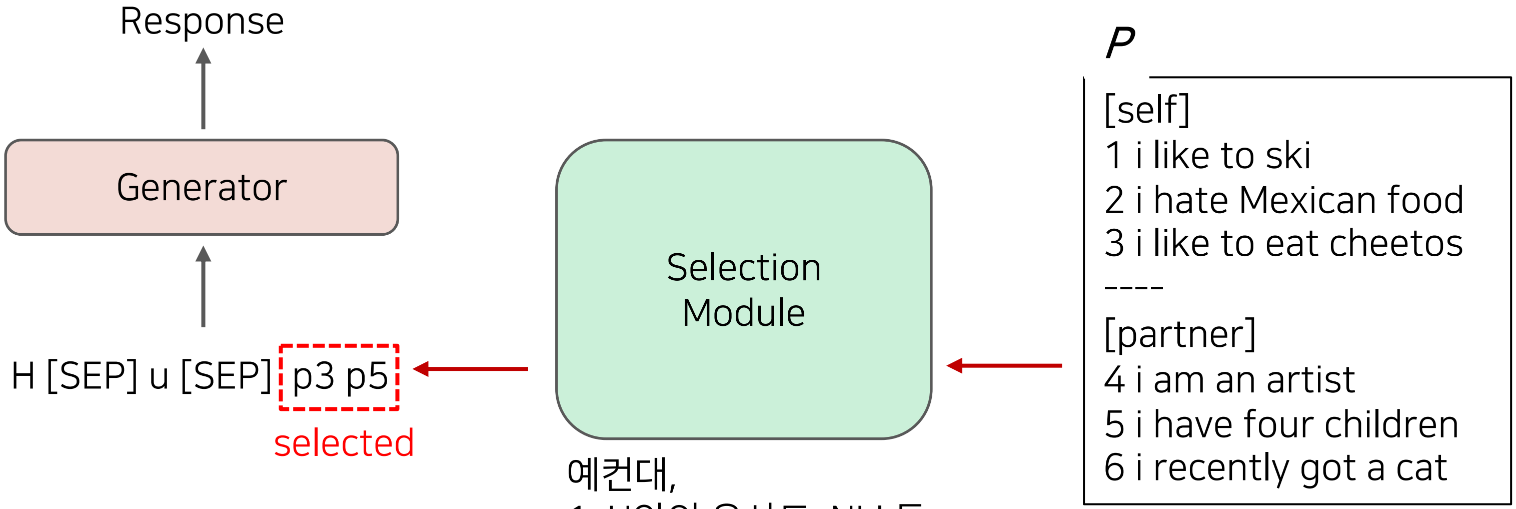
[partner]
4 i am an artist
5 i have four children
6 i recently got a cat

→ 뭐 고르지?

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Persona-grounded chat

* With PLMs
- 변형 구조 (1)



[self]
1 i like to ski
2 i hate Mexican food
3 i like to eat cheetos

[partner]
4 i am an artist
5 i have four children
6 i recently got a cat

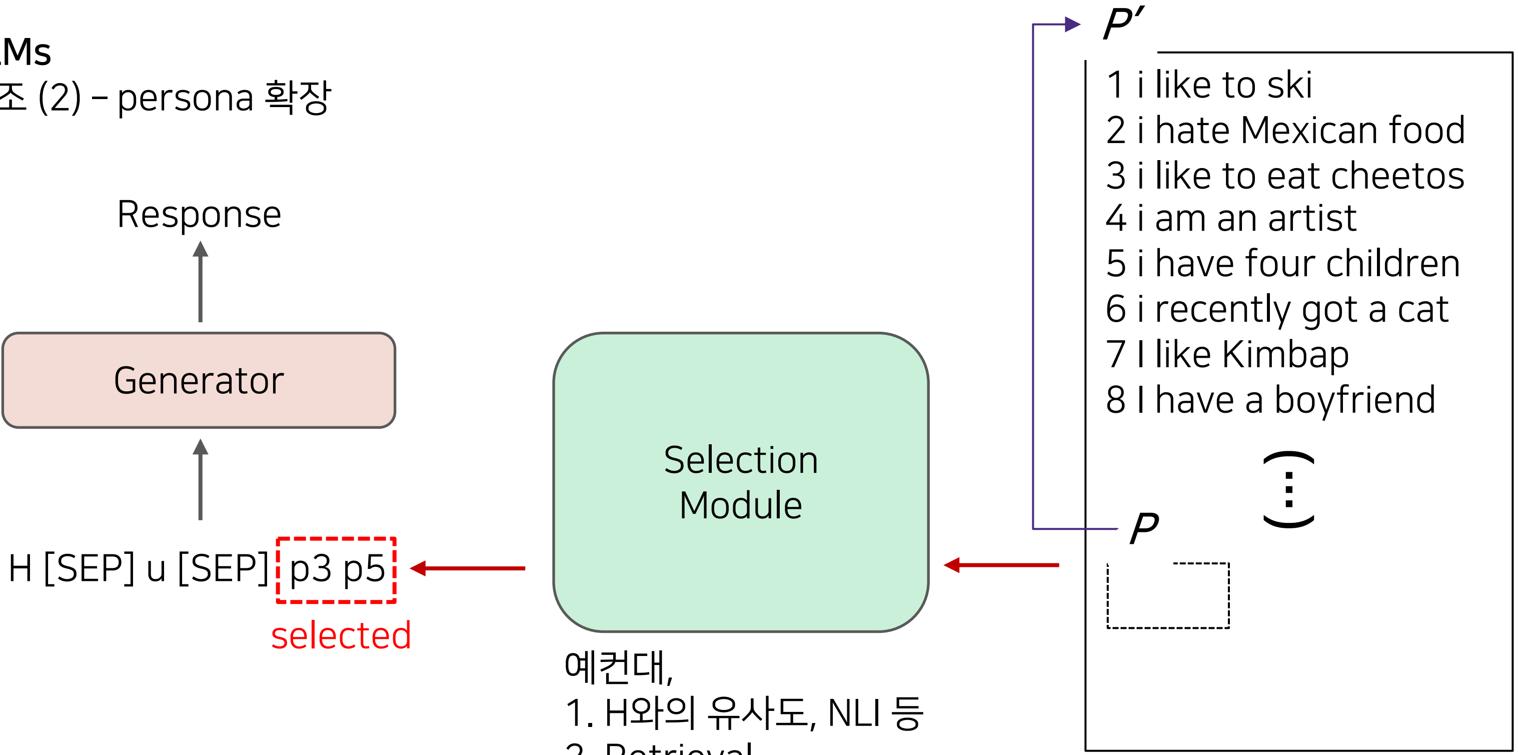
예컨대,
1. H와의 유사도, NLI 등
2. Retrieval → p(p_i| H, u)

→ p_i ground 정보는 데이터마다

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Persona-grounded chat

*** With PLMs**
- 변형 구조 (2) – persona 확장



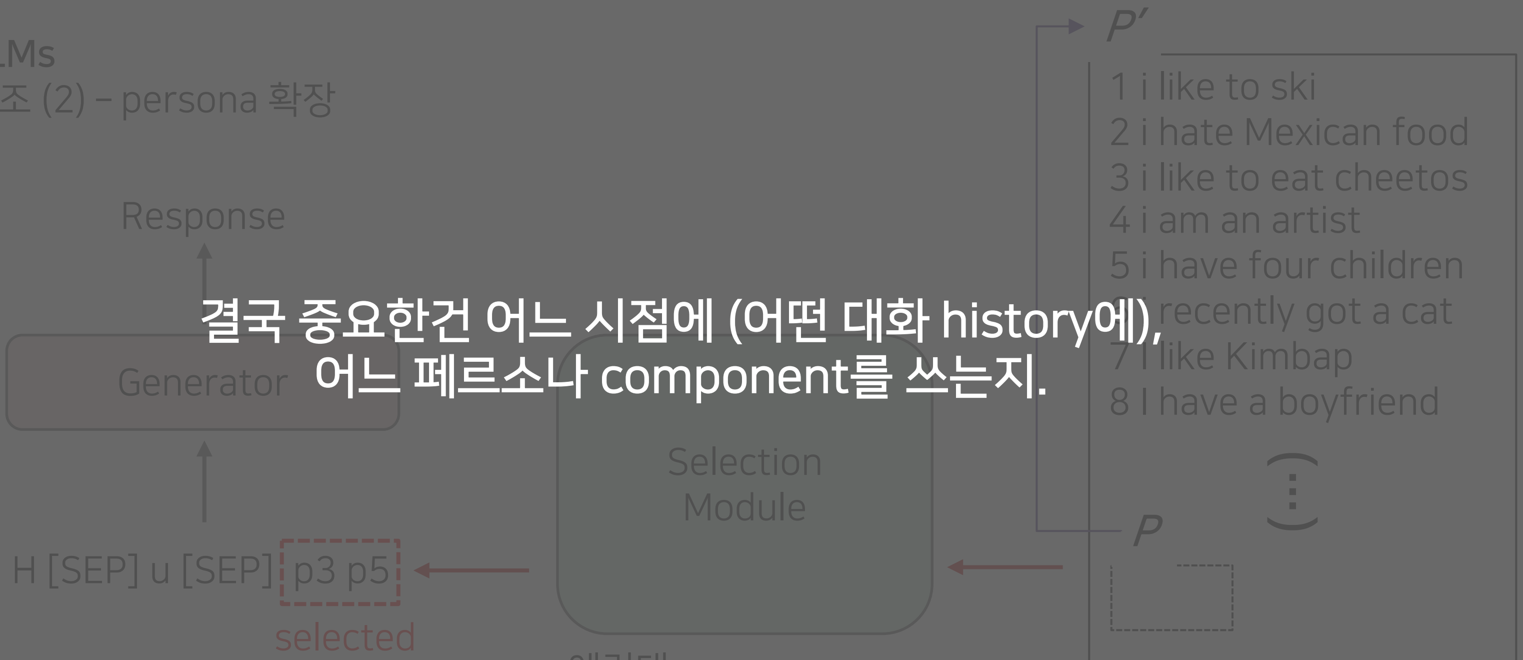
예컨대,
1. H와의 유사도, NLI 등
2. Retrieval
   → p(p_i| H, u)

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Persona-grounded chat

*** With PLMs**
- 변형 구조 (2) – persona 확장

결국 중요한건 어느 시점에 (어떤 대화 history에), 어느 페르소나 component를 쓰는지.

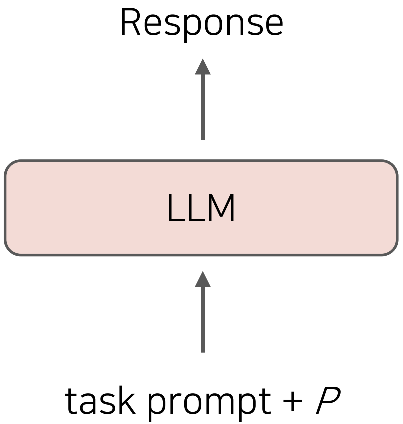


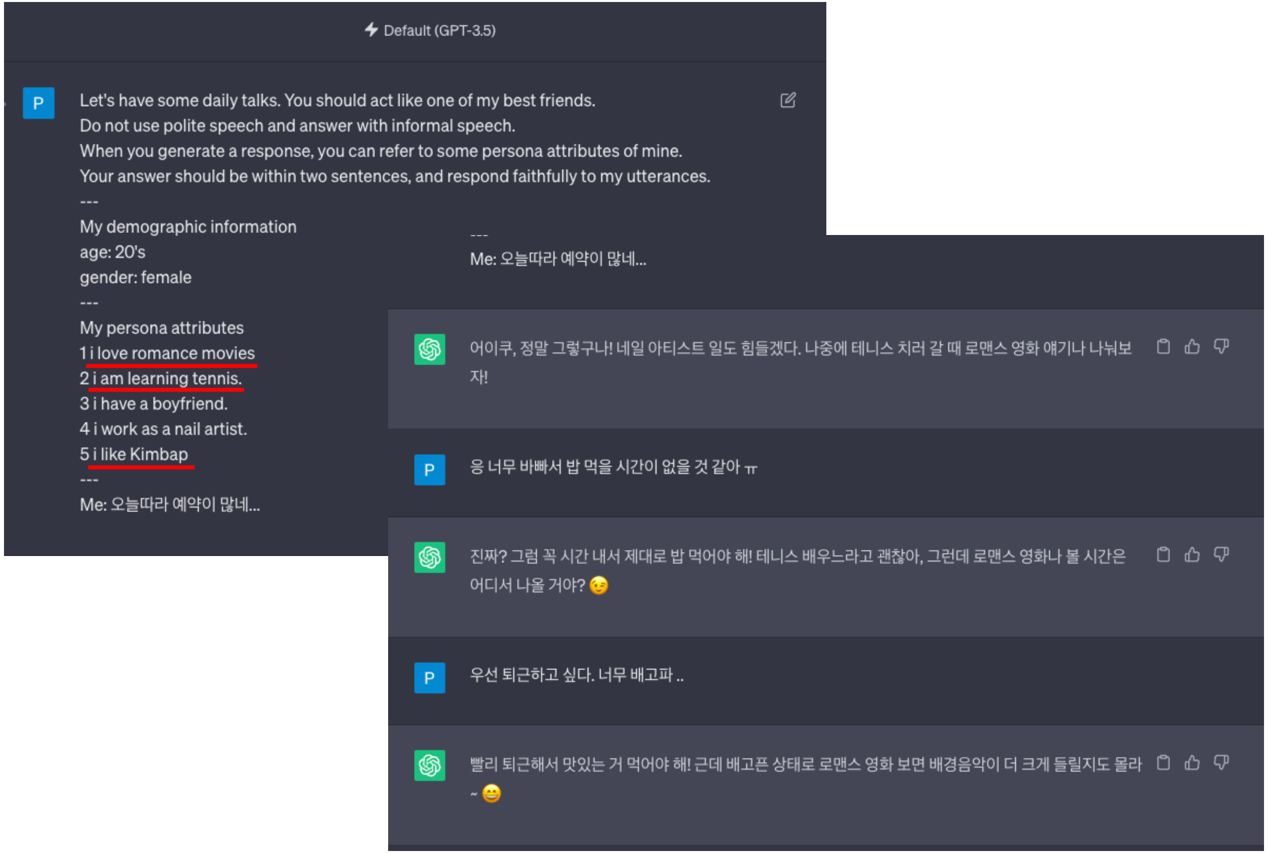
예컨대,
1. H와의 유사도, NLI 등
2. Retrieval
   → p(p_i| H, u)

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

PG chat with LLMs

* What about LLMs ?
- 기본 구조





P: Let's have some daily talks. You should act like one of my best friends.
Do not use polite speech and answer with informal speech.
When you generate a response, you can refer to some persona attributes of mine.
Your answer should be within two sentences, and respond faithfully to my utterances.
---
My demographic information
age: 20's
gender: female
---
My persona attributes
1 i love romance movies
2 i am learning tennis.
3 i have a boyfriend.
4 i work as a nail artist.
5 i like Kimbap
---
Me: 오늘따라 예약이 많네...

---
Me: 오늘따라 예약이 많네...

어이쿠, 정말 그렇구나! 네일 아티스트 일도 힘들겠다. 나중에 테니스 치러 갈 때 로맨스 영화 얘기나 나눠보자!

P: 응 너무 바빠서 밥 먹을 시간이 없을 것 같아 ㅠ

진짜? 그럼 꼭 시간 내서 제대로 밥 먹어야 해! 테니스 배우느라고 괜찮아, 그런데 로맨스 영화나 볼 시간은 어디서 나올 거야? 😉

P: 우선 퇴근하고 싶다. 너무 배고파 ..

빨리 퇴근해서 맛있는 거 먹어야 해! 근데 배고픈 상태로 로맨스 영화 보면 배경음악이 더 크게 들릴지도 몰라 ~ 😄

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

PG chat with LLMs

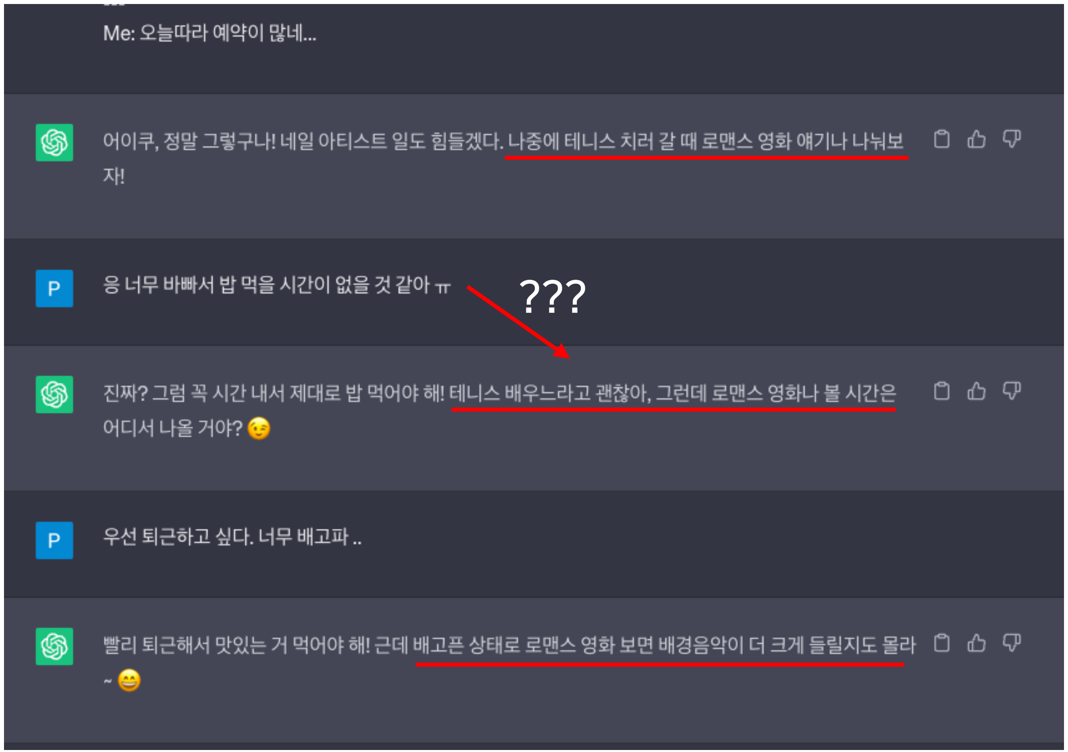
*** Motivation: 문제점**

- PLMs → *P* 줘도 잘 안 씀.

- LLMs ?
  → 너무 막 갖다 씀. (Overuse)

P

1 i love romance movies
2 i am learning tennis.
3 i have a boyfriend.
4 i work as a nail artist.
5 i like Kimbap



Me: 오늘따라 예약이 많네...

어이쿠, 정말 그렇구나! 네일 아티스트 일도 힘들겠다. 나중에 테니스 치러 갈 때 로맨스 영화 얘기나 나눠보자!

P: 응 너무 바빠서 밥 먹을 시간이 없을 것 같아 ㅠ

진짜? 그럼 꼭 시간 내서 제대로 밥 먹어야 해! 테니스 배우느라고 괜찮아, 그런데 로맨스 영화나 볼 시간은 어디서 나올 거야? 😉

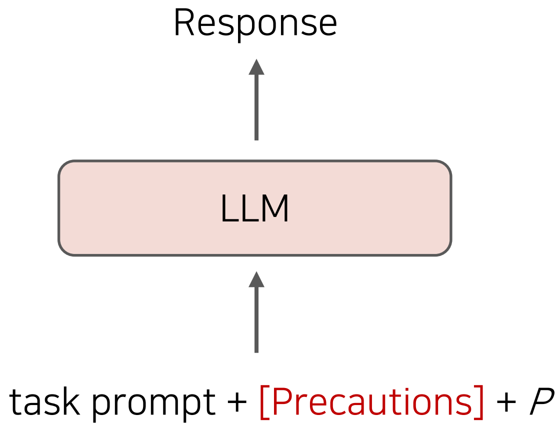
P: 우선 퇴근하고 싶다. 너무 배고파 ..

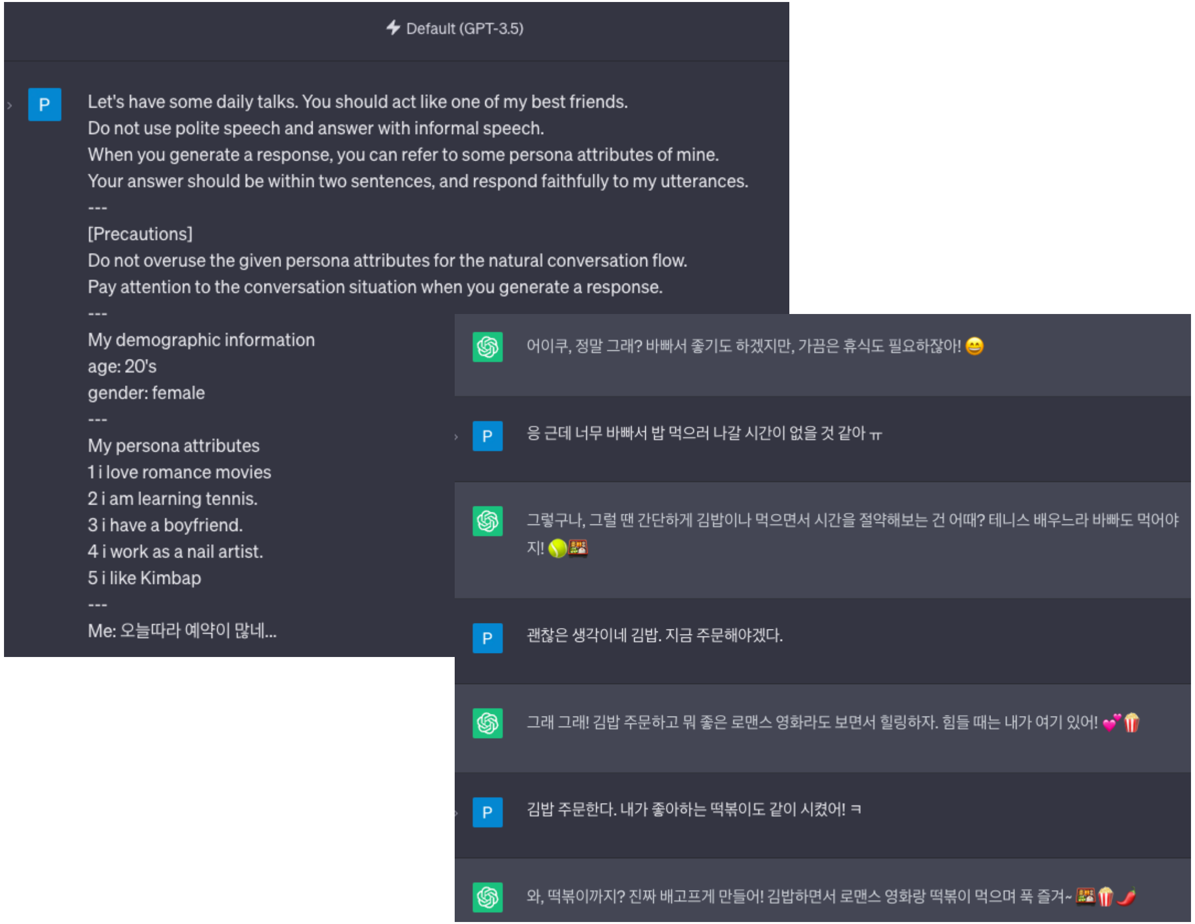
빨리 퇴근해서 맛있는 거 먹어야 해! 근데 배고픈 상태로 로맨스 영화 보면 배경음악이 더 크게 들릴지도 몰라~ 😄

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

PG chat with LLMs

* What about LLMs ?
  - 변형 구조 (1) – w/ precautions





P: Let's have some daily talks. You should act like one of my best friends.
Do not use polite speech and answer with informal speech.
When you generate a response, you can refer to some persona attributes of mine.
Your answer should be within two sentences, and respond faithfully to my utterances.

[Precautions]
Do not overuse the given persona attributes for the natural conversation flow.
Pay attention to the conversation situation when you generate a response.

My demographic information
age: 20's
gender: female

My persona attributes
1 i love romance movies
2 i am learning tennis.
3 i have a boyfriend.
4 i work as a nail artist.
5 i like Kimbap

Me: 오늘따라 예약이 많네...

어이쿠, 정말 그래? 바빠서 좋기도 하겠지만, 가끔은 휴식도 필요하잖아! 😄

P: 응 근데 너무 바빠서 밥 먹으러 나갈 시간이 없을 것 같아 ㅠ

그렇구나, 그럴 땐 간단하게 김밥이나 먹으면서 시간을 절약해보는 건 어때? 테니스 배우느라 바빠도 먹어야지! 🎾🍱

P: 괜찮은 생각이네 김밥. 지금 주문해야겠다.

그래 그래! 김밥 주문하고 뭐 좋은 로맨스 영화라도 보면서 힐링하자. 힘들 때는 내가 여기 있어! 💕🍿

P: 김밥 주문한다. 내가 좋아하는 떡볶이도 같이 시켰어! ㅋ

와, 떡볶이까지? 진짜 배고프게 만들어! 김밥하면서 로맨스 영화랑 떡볶이 먹으며 푹 즐겨~ 🍱🍿🌶️

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

PG chat with LLMs

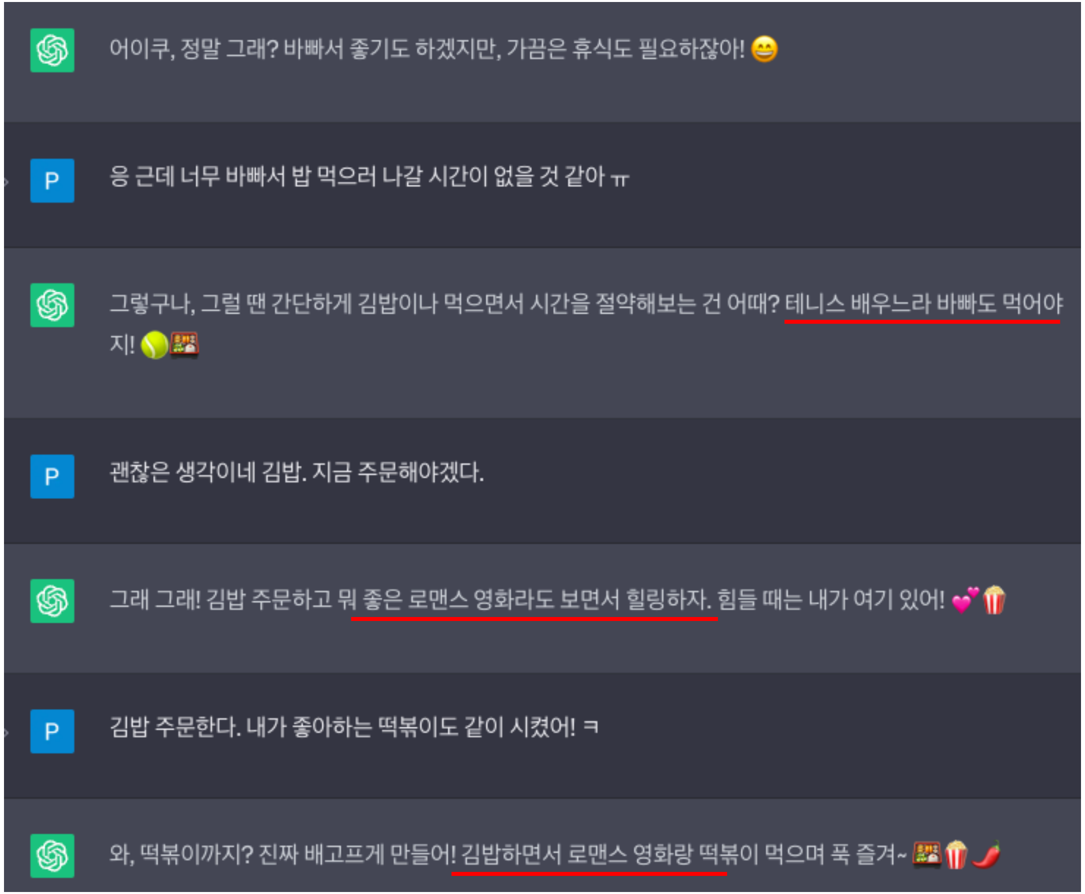
*** Motivation: 문제점**
- 변형 구조 (1) – w/ precautions

→ 그래도 틈만 나면 어떻게든 쓰려고 함.

[주의]
- 페르소나 과사용 금지
- 대화 상황에 유의

P

1 i love romance movies
2 i am learning tennis.
3 i have a boyfriend.
4 i work as a nail artist.
5 i like Kimbap



어이쿠, 정말 그래? 바빠서 좋기도 하겠지만, 가끔은 휴식도 필요하잖아! 😄

P 응 근데 너무 바빠서 밥 먹으러 나갈 시간이 없을 것 같아 ㅠ

그렇구나, 그럴 땐 간단하게 김밥이나 먹으면서 시간을 절약해보는 건 어때? 테니스 배우느라 바빠도 먹어야지! 🎾🍱

P 괜찮은 생각이네 김밥. 지금 주문해야겠다.

그래 그래! 김밥 주문하고 뭐 좋은 로맨스 영화라도 보면서 힐링하자. 힘들 때는 내가 여기 있어! 💕🍿

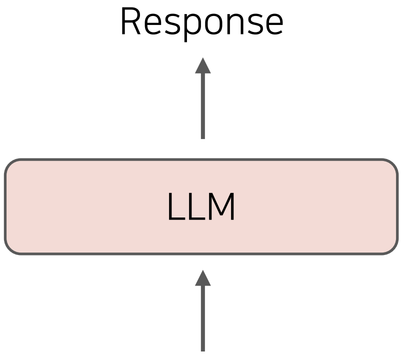
P 김밥 주문한다. 내가 좋아하는 떡볶이도 같이 시켰어! ㅋ

와, 떡볶이까지? 진짜 배고프게 만들어! 김밥하면서 로맨스 영화랑 떡볶이 먹으며 푹 즐겨~ 🍱🍿🌶️

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

PG chat with LLMs

* What about LLMs ?
- 변형 구조 (2) – w/ precautions & 예제

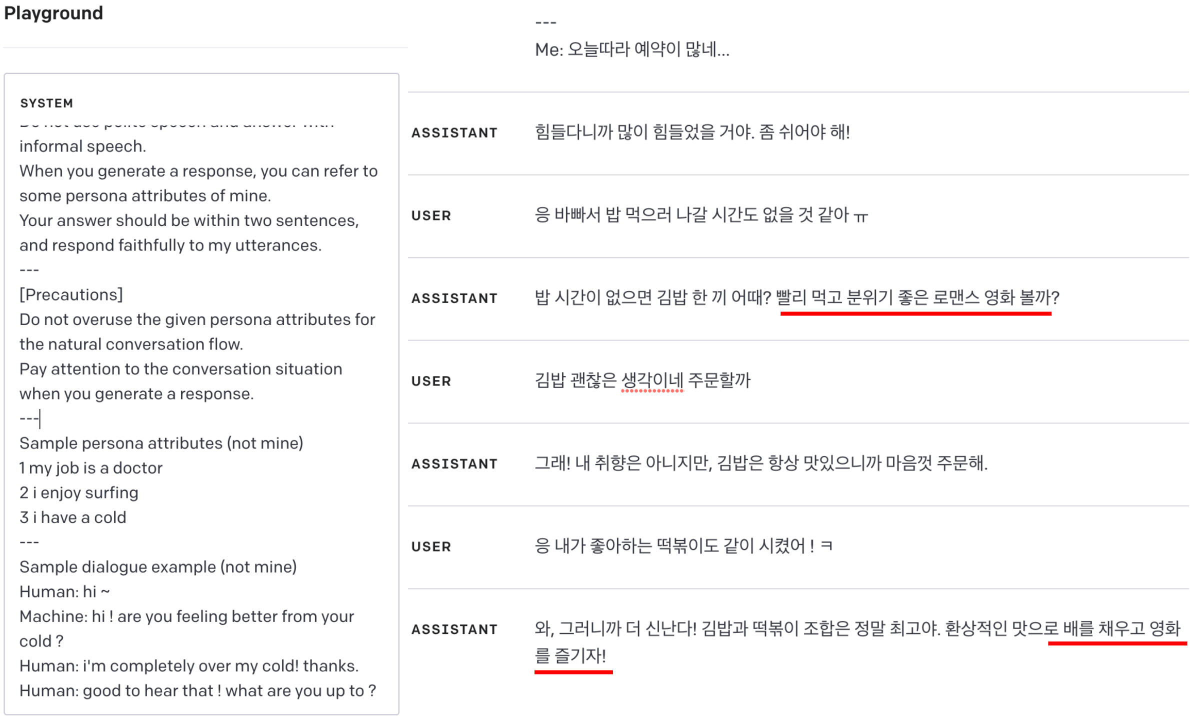


Response

LLM

task prompt + precautions
+ (타 상황 샘플) H + *P*

→ 같은 *P* set 을 기반으로 이루어지는 대화 예제를 여러 개 구하는 것 불가.



Playground

SYSTEM

informal speech.
When you generate a response, you can refer to some persona attributes of mine.
Your answer should be within two sentences, and respond faithfully to my utterances.

[Precautions]
Do not overuse the given persona attributes for the natural conversation flow.
Pay attention to the conversation situation when you generate a response.

Sample persona attributes (not mine)
1 my job is a doctor
2 i enjoy surfing
3 i have a cold

Sample dialogue example (not mine)
Human: hi ~
Machine: hi ! are you feeling better from your cold ?
Human: i'm completely over my cold! thanks.
Human: good to hear that ! what are you up to ?

Me: 오늘따라 예약이 많네...

ASSISTANT 힘들다니까 많이 힘들었을 거야. 좀 쉬어야 해!

USER 응 바빠서 밥 먹으러 나갈 시간도 없을 것 같아 ㅠ

ASSISTANT 밥 시간이 없으면 김밥 한 끼 어때? 빨리 먹고 분위기 좋은 로맨스 영화 볼까?

USER 김밥 괜찮은 생각이네 주문할까

ASSISTANT 그래! 내 취향은 아니지만, 김밥은 항상 맛있으니까 마음껏 주문해.

USER 응 내가 좋아하는 떡볶이도 같이 시켰어 ! ㅋ

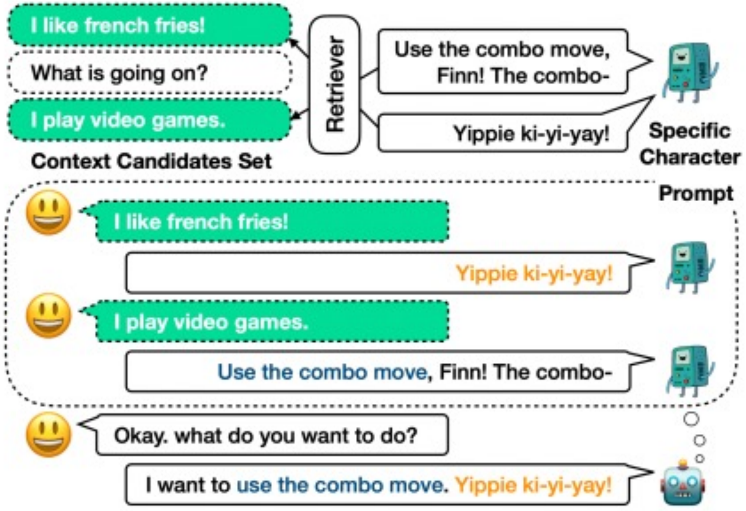
ASSISTANT 와, 그러니까 더 신난다! 김밥과 떡볶이 조합은 정말 최고야. 환상적인 맛으로 배를 채우고 영화를 즐기자!

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Previous Works (1)

* Meet Your Favorite Character:
Open-domain Chatbot Mimicking Fictional Characters with only a Few Utterances (2022 NAACL)

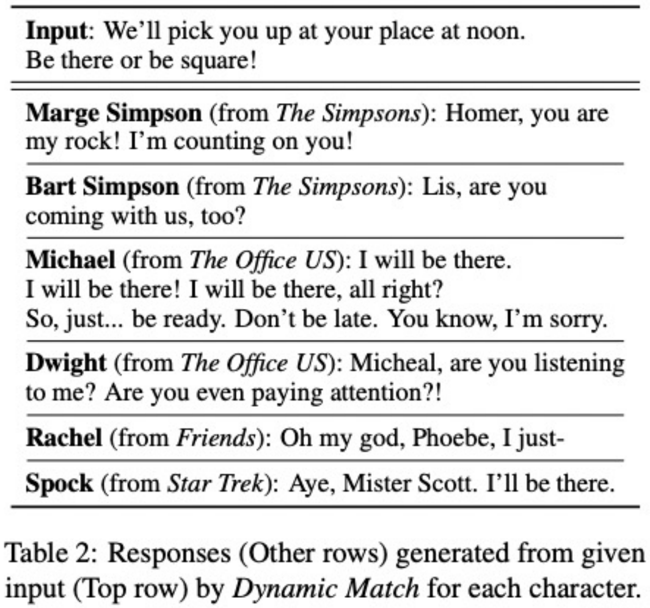
- 기존 persona descriptions이 style을 충분히 반영 X
- 대화 형태의 프롬프팅으로 원하는 캐릭터처럼 말하게 하자.
- BST 데이터셋으로 Pseudo Candidate Context set C 구축
- Bi-encoder 통해 3가지 matching 방법으로 c_i retrieval
  a) Random
  b) Static : u (발화/대사) → c_i
  c) Dynamic : u + x (input context) → c_i
- 최대 GPT-J (6B) 활용



Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Previous Works (1)

* 결과



Input: We'll pick you up at your place at noon.
Be there or be square!

Marge Simpson (from *The Simpsons*): Homer, you are my rock! I'm counting on you!

Bart Simpson (from *The Simpsons*): Lis, are you coming with us, too?

Michael (from *The Office US*): I will be there.
I will be there! I will be there, all right?
So, just... be ready. Don't be late. You know, I'm sorry.

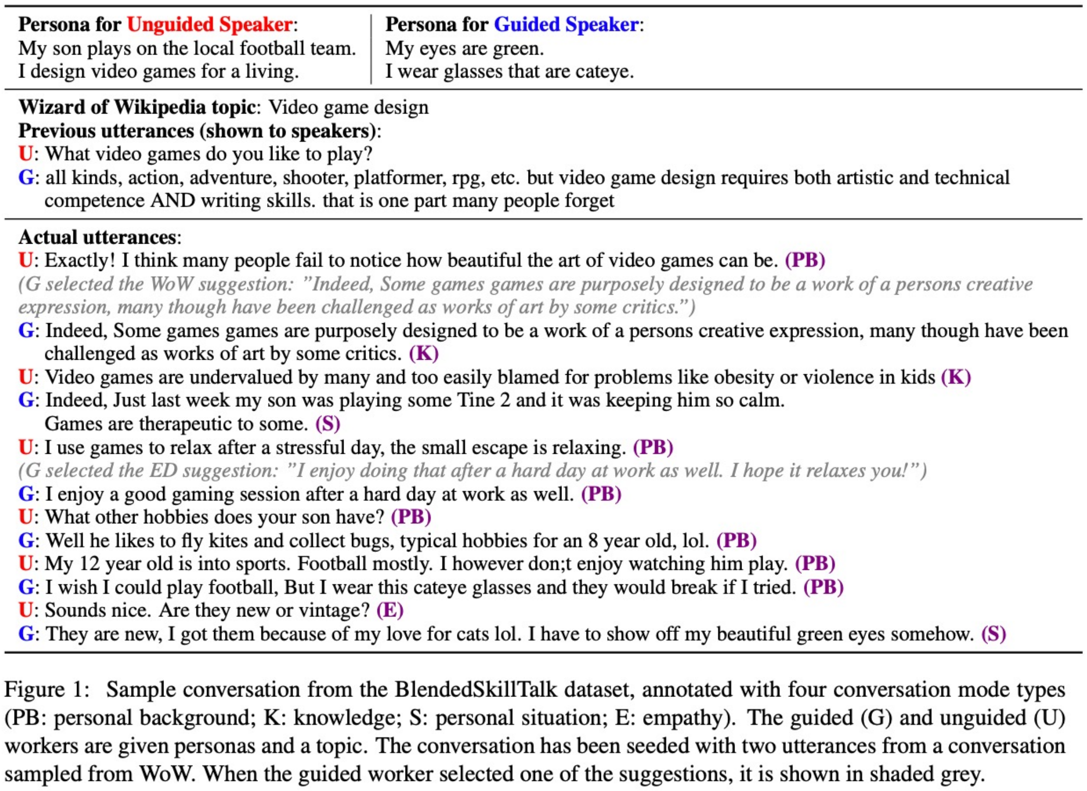
Dwight (from *The Office US*): Micheal, are you listening to me? Are you even paying attention?!

Rachel (from *Friends*): Oh my god, Phoebe, I just-

Spock (from *Star Trek*): Aye, Mister Scott. I'll be there.

Table 2: Responses (Other rows) generated from given input (Top row) by *Dynamic Match* for each character.

* 참고) BST 데이터셋 (2020)



Persona for Unguided Speaker:	**Persona for Guided Speaker**:
My son plays on the local football team. I design video games for a living.	My eyes are green. I wear glasses that are cateye.

Wizard of Wikipedia topic: Video game design
Previous utterances (shown to speakers):
U: What video games do you like to play?
G: all kinds, action, adventure, shooter, platformer, rpg, etc. but video game design requires both artistic and technical competence AND writing skills. that is one part many people forget

Actual utterances:
U: Exactly! I think many people fail to notice how beautiful the art of video games can be. **(PB)**
(G selected the WoW suggestion: "Indeed, Some games games are purposely designed to be a work of a persons creative expression, many though have been challenged as works of art by some critics.")
G: Indeed, Some games games are purposely designed to be a work of a persons creative expression, many though have been challenged as works of art by some critics. **(K)**
U: Video games are undervalued by many and too easily blamed for problems like obesity or violence in kids **(K)**
G: Indeed, Just last week my son was playing some Tine 2 and it was keeping him so calm. Games are therapeutic to some. **(S)**
U: I use games to relax after a stressful day, the small escape is relaxing. **(PB)**
(G selected the ED suggestion: "I enjoy doing that after a hard day at work as well. I hope it relaxes you!")
G: I enjoy a good gaming session after a hard day at work as well. **(PB)**
U: What other hobbies does your son have? **(PB)**
G: Well he likes to fly kites and collect bugs, typical hobbies for an 8 year old, lol. **(PB)**
U: My 12 year old is into sports. Football mostly. I however don;t enjoy watching him play. **(PB)**
G: I wish I could play football, But I wear this cateye glasses and they would break if I tried. **(PB)**
U: Sounds nice. Are they new or vintage? **(E)**
G: They are new, I got them because of my love for cats lol. I have to show off my beautiful green eyes somehow. **(S)**

Figure 1: Sample conversation from the BlendedSkillTalk dataset, annotated with four conversation mode types (PB: personal background; K: knowledge; S: personal situation; E: empathy). The guided (G) and unguided (U) workers are given personas and a topic. The conversation has been seeded with two utterances from a conversation sampled from WoW. When the guided worker selected one of the suggestions, it is shown in shaded grey.

https://parl.ai/projects/bst

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Previous Works (2)

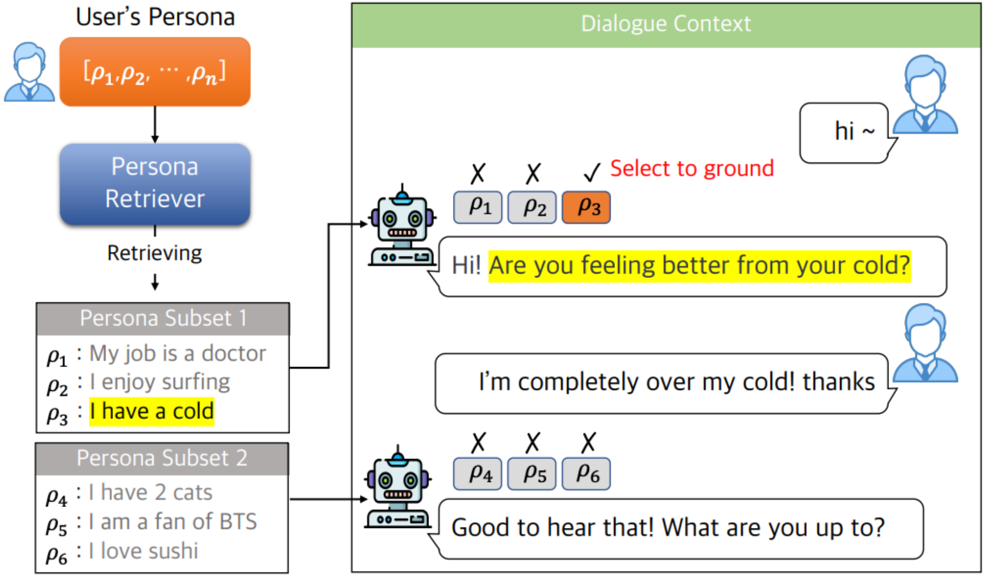
* WHAT, WHEN, and HOW to Ground:
Designing User Persona-Aware Conversational Agents for Engaging Dialogue (ACL 2023 Industry)

- What, When, and How problem (WWH) for natural response generation
  : **balance** between dialogue **fluency** and tendency to **ground**

→ inference 상황의 persona selection
= non-trivial

→ 가령, retrieved candidates 를 그냥 쓸지,
혹은 rerank 후 top-k로 자를지의 문제 등.

때로는 아예 안 쓰는게 더 자연스러운
상황도 존재함.



User's Persona

$[\rho_1, \rho_2, \cdots, \rho_n]$

Persona Retriever

Retrieving

Persona Subset 1

ρ_1 : My job is a doctor
ρ_2 : I enjoy surfing
ρ_3 : I have a cold

Persona Subset 2

ρ_4 : I have 2 cats
ρ_5 : I am a fan of BTS
ρ_6 : I love sushi

Dialogue Context

hi ~

✓ Select to ground

Hi! Are you feeling better from your cold?

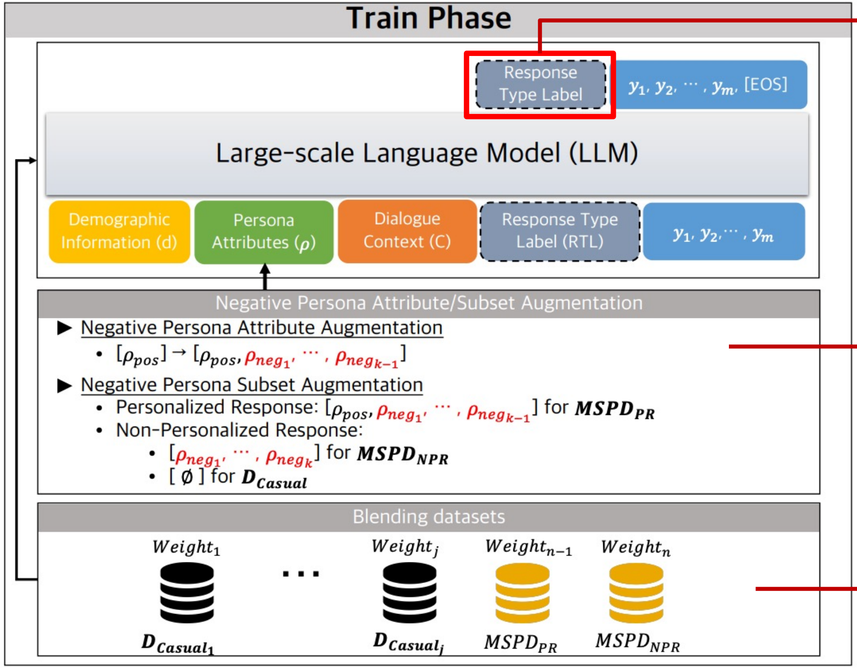
I'm completely over my cold! thanks

Good to hear that! What are you up to?

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Previous Works (2)

* WHAT, WHEN, and HOW to Ground:
Designing User Persona-Aware Conversational Agents for Engaging Dialogue (ACL 2023 Industry)


Train Phase
Response Type Label
$y_1, y_2, \cdots, y_m$, [EOS]
Large-scale Language Model (LLM)
Demographic Information (d)
Persona Attributes (ρ)
Dialogue Context (C)
Response Type Label (RTL)
$y_1, y_2, \cdots, y_m$
Negative Persona Attribute/Subset Augmentation
▶ Negative Persona Attribute Augmentation
• $[\rho_{pos}] \rightarrow [\rho_{pos}, \rho_{neg_1}, \cdots, \rho_{neg_{k-1}}]$
▶ Negative Persona Subset Augmentation
• Personalized Response: $[\rho_{pos}, \rho_{neg_1}, \cdots, \rho_{neg_{k-1}}]$ for $\boldsymbol{MSPD_{PR}}$
• Non-Personalized Response:
• $[\rho_{neg_1}, \cdots, \rho_{neg_k}]$ for $\boldsymbol{MSPD_{NPR}}$
• $[\, \emptyset \,]$ for $\boldsymbol{D_{Casual}}$
Blending datasets
$Weight_1$
$Weight_j$
$Weight_{n-1}$
$Weight_n$
$\boldsymbol{D_{Casual_1}}$
$\boldsymbol{D_{Casual_j}}$
$MSPD_{PR}$
$MSPD_{NPR}$

P subset의 retrieval 결과에 대한 description
→ 생성 모델이 생성하는 응답의 유형에 대해 <PRTL> or <CRTL>이라는 token을 함께 생성하도록 학습.
(<PRTL>: personalized / <CRTL>: casual)

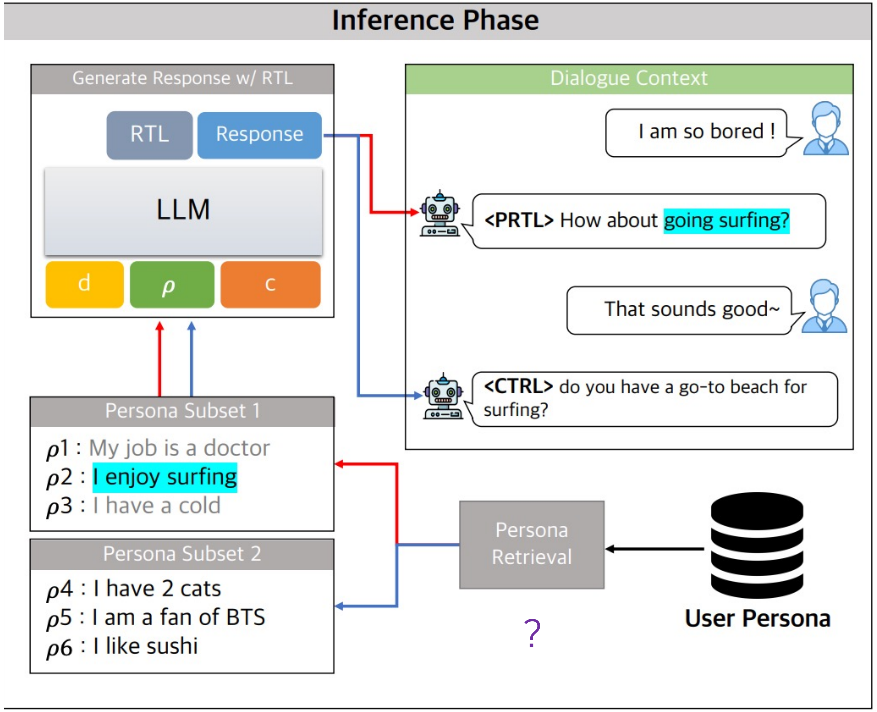
1) Non-personalized 응답의 생성을 위해 다 irrelevant persona attributes로 subset 구성
→ 이를 통해, 너무 자주 grounding 하는 경향 억제
2) Personalized 응답 생성을 위해서는, {relevant + negatives} subset 구성

한국어 버전 Pchat 등 Multi-turn 데이터 (NIA)
+
한국어 Multi-session 데이터
→ 데이터셋 수에 따라 균일한 비율로

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Previous Works (2)

* WHAT, WHEN, and HOW to Ground:
Designing User Persona-Aware Conversational Agents for Engaging Dialogue (ACL 2023 Industry)



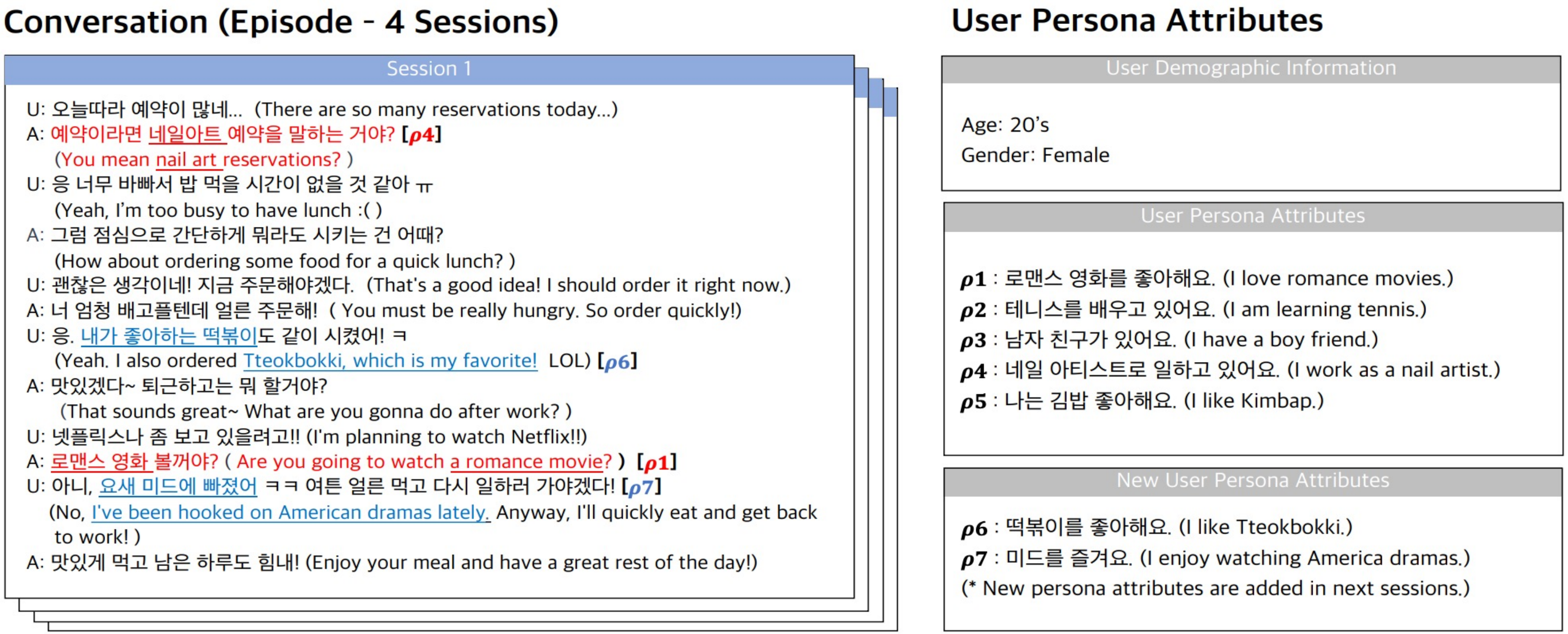
?

→ 요컨대,
RTL, Dataset Blending, negative sampling을 통해서 결국, Casual vs. Personalized response 간에 구분하려고 했다.

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Previous Works (2)

* WHAT, WHEN, and HOW to Ground:
Designing User Persona-Aware Conversational Agents for Engaging Dialogue (ACL 2023 Industry)
- 결과 (GPT-3)



Conversation (Episode - 4 Sessions)

Session 1

U: 오늘따라 예약이 많네… (There are so many reservations today…)
A: 예약이라면 네일아트 예약을 말하는 거야? **[ρ4]**
(You mean nail art reservations?)
U: 응 너무 바빠서 밥 먹을 시간이 없을 것 같아 ㅠ
(Yeah, I'm too busy to have lunch :( )
A: 그럼 점심으로 간단하게 뭐라도 시키는 건 어때?
(How about ordering some food for a quick lunch?)
U: 괜찮은 생각이네! 지금 주문해야겠다. (That's a good idea! I should order it right now.)
A: 너 엄청 배고플텐데 얼른 주문해! (You must be really hungry. So order quickly!)
U: 응. 내가 좋아하는 떡볶이도 같이 시켰어! ㅋ
(Yeah. I also ordered Tteokbokki, which is my favorite! LOL) **[ρ6]**
A: 맛있겠다~ 퇴근하고는 뭐 할거야?
(That sounds great~ What are you gonna do after work?)
U: 넷플릭스나 좀 보고 있을려고!! (I'm planning to watch Netflix!!)
A: 로맨스 영화 볼꺼야? (Are you going to watch a romance movie?) **[ρ1]**
U: 아니, 요새 미드에 빠졌어 ㅋㅋ 여튼 얼른 먹고 다시 일하러 가야겠다! **[ρ7]**
(No, I've been hooked on American dramas lately. Anyway, I'll quickly eat and get back to work!)
A: 맛있게 먹고 남은 하루도 힘내! (Enjoy your meal and have a great rest of the day!)

User Persona Attributes

User Demographic Information

Age: 20's
Gender: Female

User Persona Attributes

ρ1 : 로맨스 영화를 좋아해요. (I love romance movies.)
ρ2 : 테니스를 배우고 있어요. (I am learning tennis.)
ρ3 : 남자 친구가 있어요. (I have a boy friend.)
ρ4 : 네일 아티스트로 일하고 있어요. (I work as a nail artist.)
ρ5 : 나는 김밥 좋아해요. (I like Kimbap.)

New User Persona Attributes

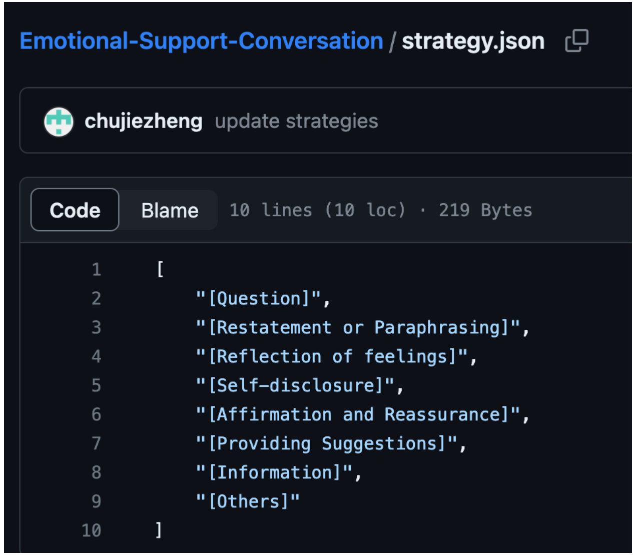
ρ6 : 떡볶이를 좋아해요. (I like Tteokbokki.)
ρ7 : 미드를 즐겨요. (I enjoy watching America dramas.)
(* New persona attributes are added in next sessions.)

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

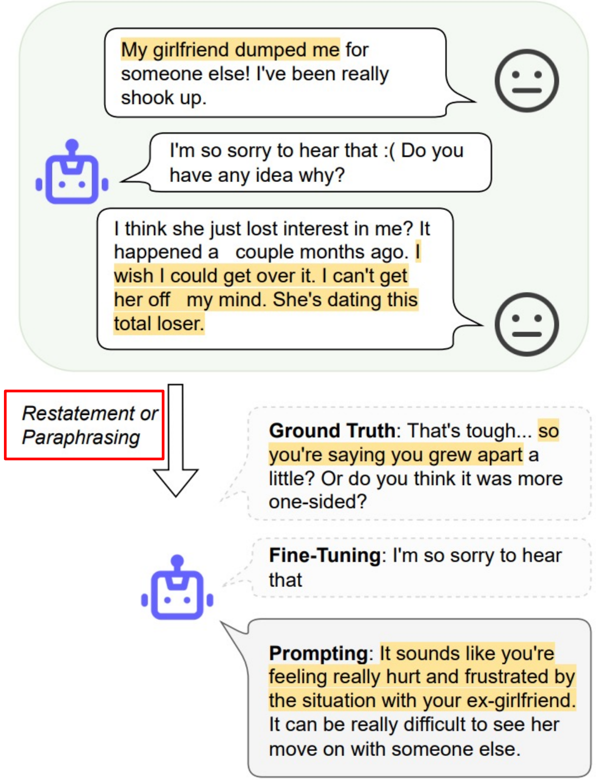
Previous Works (3)

* Controllable Mixed-Initiative Dialogue Generation through Prompting (ACL 2023 Short)

- 정서적 도움 봇 구축 목적 (InstructGPT)
- Restatement 등 현재의 dialogue intent 를, 상응하는 자연어 정보로 치환 후에 대화와 같이 넣어줌.


Emotional-Support-Conversation / strategy.json
chujiezheng update strategies
Code
Blame
10 lines (10 loc) · 219 Bytes

```
[
    "[Question]",
    "[Restatement or Paraphrasing]",
    "[Reflection of feelings]",
    "[Self-disclosure]",
    "[Affirmation and Reassurance]",
    "[Providing Suggestions]",
    "[Information]",
    "[Others]"
]
```



My girlfriend dumped me for someone else! I've been really shook up.

I'm so sorry to hear that :(Do you have any idea why?

I think she just lost interest in me? It happened a couple months ago. I wish I could get over it. I can't get her off my mind. She's dating this total loser.

Ground Truth: That's tough... so you're saying you grew apart a little? Or do you think it was more one-sided?

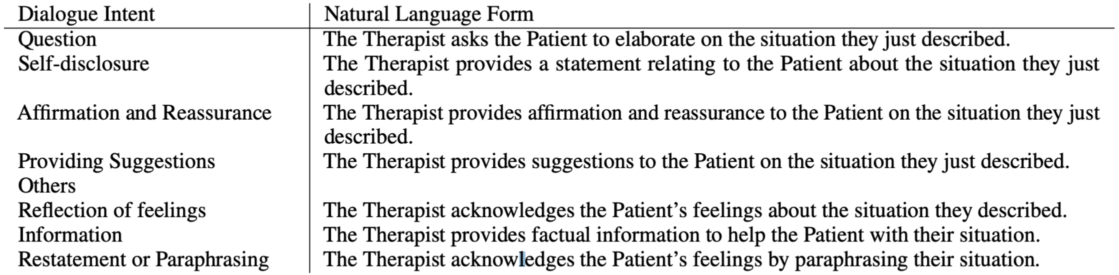
Fine-Tuning: I'm so sorry to hear that

Prompting: It sounds like you're feeling really hurt and frustrated by the situation with your ex-girlfriend. It can be really difficult to see her move on with someone else.

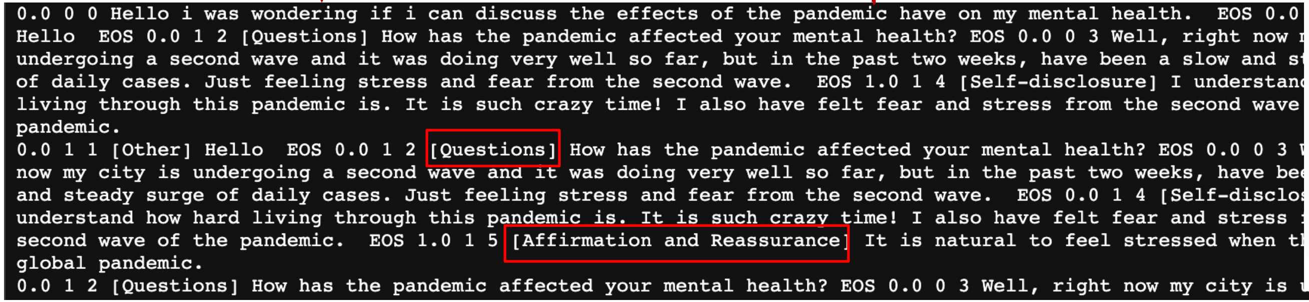
Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Previous Works (3)

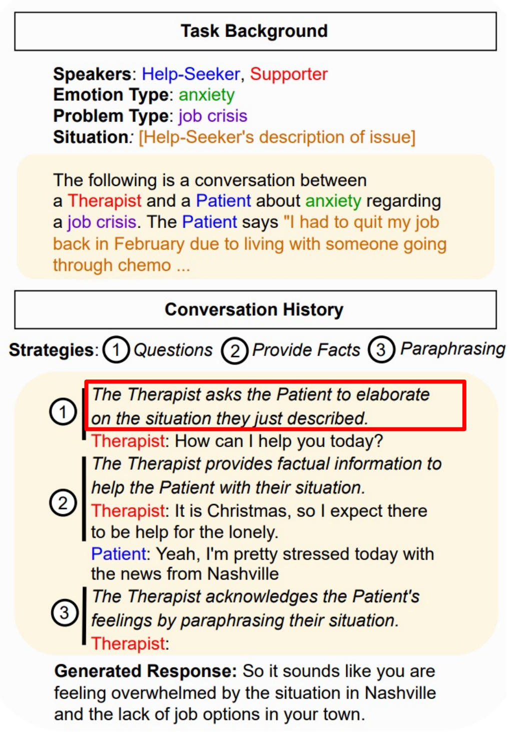
* Controllable Mixed-Initiative Dialogue Generation through Prompting (ACL 2023 Short)



Dialogue Intent	Natural Language Form
Question	The Therapist asks the Patient to elaborate on the situation they just described.
Self-disclosure	The Therapist provides a statement relating to the Patient about the situation they just described.
Affirmation and Reassurance	The Therapist provides affirmation and reassurance to the Patient on the situation they just described.
Providing Suggestions	The Therapist provides suggestions to the Patient on the situation they just described.
Others	
Reflection of feelings	The Therapist acknowledges the Patient's feelings about the situation they described.
Information	The Therapist provides factual information to help the Patient with their situation.
Restatement or Paraphrasing	The Therapist acknowledges the Patient's feelings by paraphrasing their situation.



```
0.0 0 0 Hello i was wondering if i can discuss the effects of the pandemic have on my mental health.  EOS 0.0
Hello  EOS 0.0 1 2 [Questions] How has the pandemic affected your mental health? EOS 0.0 0 3 Well, right now
undergoing a second wave and it was doing very well so far, but in the past two weeks, have been a slow and s
of daily cases. Just feeling stress and fear from the second wave.  EOS 1.0 1 4 [Self-disclosure] I understan
living through this pandemic is. It is such crazy time! I also have felt fear and stress from the second wave
pandemic.
0.0 1 1 [Other] Hello  EOS 0.0 1 2 [Questions] How has the pandemic affected your mental health? EOS 0.0 0 3
now my city is undergoing a second wave and it was doing very well so far, but in the past two weeks, have bee
and steady surge of daily cases. Just feeling stress and fear from the second wave.  EOS 0.0 1 4 [Self-disclos
understand how hard living through this pandemic is. It is such crazy time! I also have felt fear and stress
second wave of the pandemic.  EOS 1.0 1 5 [Affirmation and Reassurance] It is natural to feel stressed when t
global pandemic.
0.0 1 2 [Questions] How has the pandemic affected your mental health? EOS 0.0 0 3 Well, right now my city is
```



Task Background

Speakers: Help-Seeker, Supporter
Emotion Type: anxiety
Problem Type: job crisis
Situation: [Help-Seeker's description of issue]

The following is a conversation between a Therapist and a Patient about anxiety regarding a job crisis. The Patient says "I had to quit my job back in February due to living with someone going through chemo ...

Conversation History

Strategies: ① *Questions* ② *Provide Facts* ③ *Paraphrasing*

① *The Therapist asks the Patient to elaborate on the situation they just described.*
Therapist: How can I help you today?

② *The Therapist provides factual information to help the Patient with their situation.*
Therapist: It is Christmas, so I expect there to be help for the lonely.
Patient: Yeah, I'm pretty stressed today with the news from Nashville

③ *The Therapist acknowledges the Patient's feelings by paraphrasing their situation.*
Therapist:

Generated Response: So it sounds like you are feeling overwhelmed by the situation in Nashville and the lack of job options in your town.

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Research Focus

*** How to reduce the overuse problem ?**

- Initial Prompting 개선
  . Style 파악이 어려운 평서문 형태의 persona 가 아니라,
  해당 특징을 가진 사람이 말할 것 같은 일종의 **"대사"**를 바탕으로 **대화 형태 예제 제공** (like Related Work #1)

- Persona Selection phase 도입
  . w/ PLM
  e.g.) persona set P 혹은 retrieval 이후의 subset P^- 내의 component 각각에 대해서
  History 혹은 Last utterance 와 관련하여 grounding 할지 말지 분류하도록 학습
  . w/ LLM
  e.g.) Self-Refining 형태로, response 를 생성하기 전 혹은 후 스스로 H/u 를 기반으로 ground 여부 재판단

  + 공통적으로, H/u 정보와 함께 주요 사건 or 응답 유형 정보 (Work #2) 를 추출하여 함께 제공할 수도.
  e.g.) H: 퇴근하면 넷플릭스 보고 있어야지..
  M: 좋아. 지금 시킨 떡볶이랑 넷플릭스에서 로맨스 영화 보면 되겠다. → '퇴근' ???

- 그 외 + α : 발화의 의도 정보를 파악하고, 그에 대응하는 자연어 prompt form 제공 ? (work #3)

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab

Thank you

Copyright © 2021 Natural Language Processing & Artificial Intelligence Lab