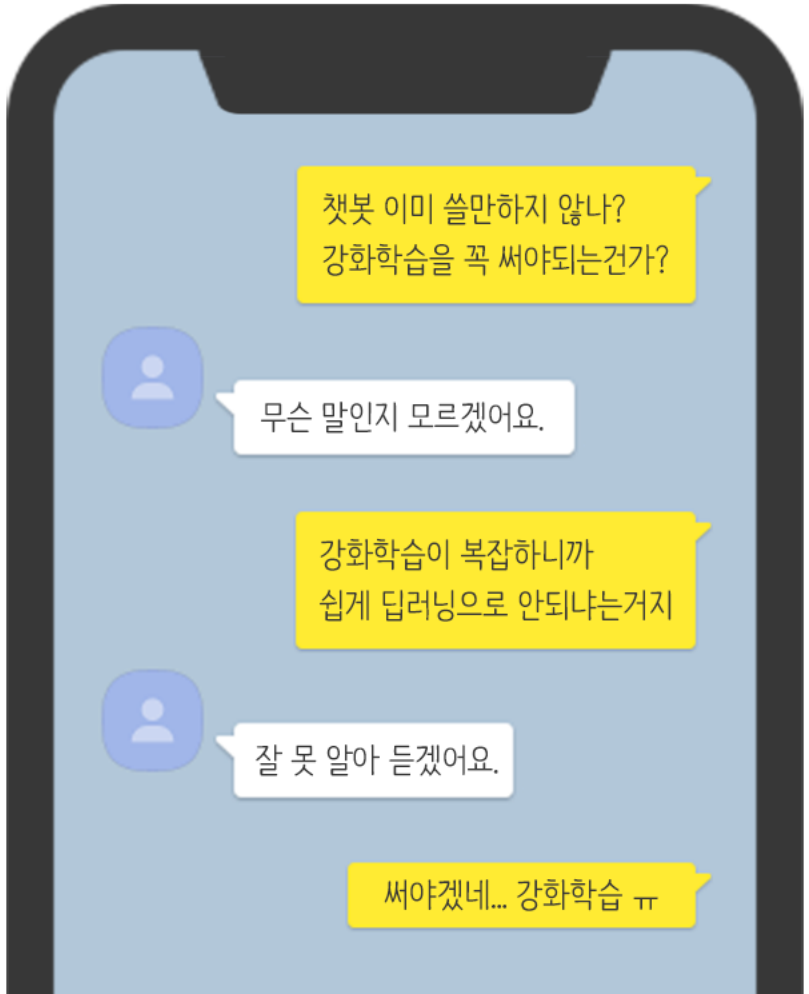
DEEP REINFORCEMENT LEARNING
FOR DIALOGUE GENERATION



2020. 05. 13.
Moon Sungwon

기존의 방법이 대체 어땠길래?

기계 번역

Machine Translation

규칙 기반

Ex) *I'm looking for someone.*

→ i/PRP be/VBP look/VBG
for/IN someone/NN

→ 나/NP 는/FX 찾/VB
고있/IN 누군가/NN

→ *나는 누군가를 찾고 있습니다.*

통계 기반

Ex) *I'm looking for someone.*

→ 'I' 가 나오면 **대체로** '나'로 시작

→ '나' 다음엔 '는'이 **많이** 나오고_

→ 끝 단어가 **주로** 지금 나오니
'나는 누군가'_

→ (중략)

→ *나는 누군가를 찾고 있습니다.*

신경망 기반

I'm looking for someone.

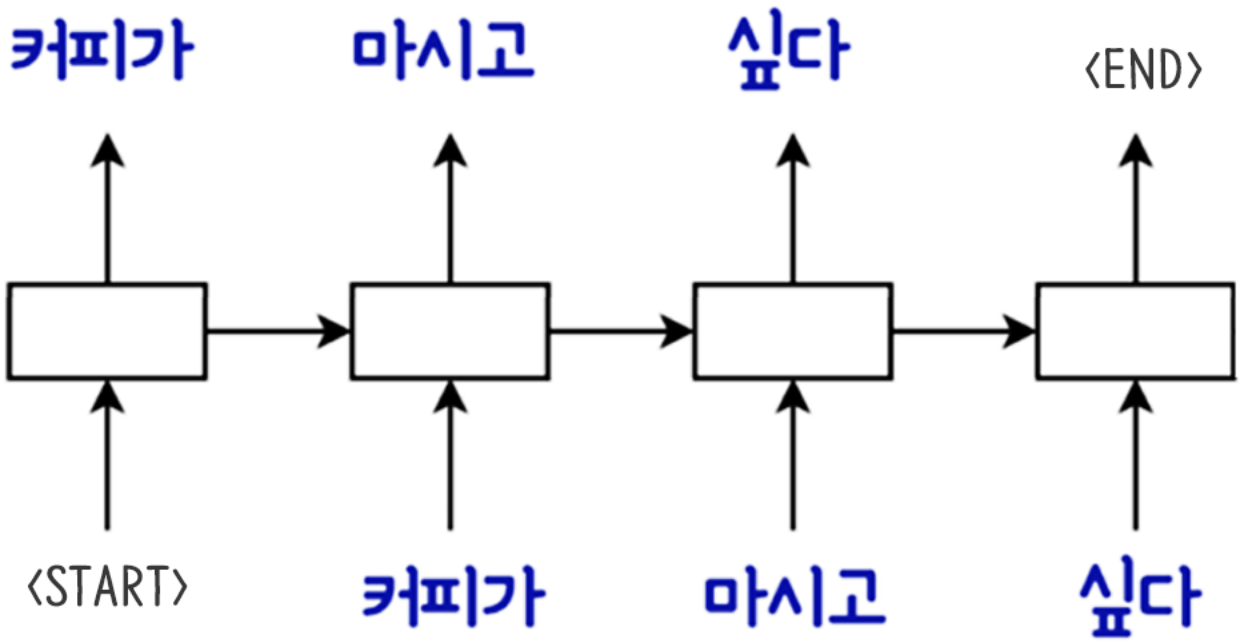
↓

(엄청난 번역 모델)

↓

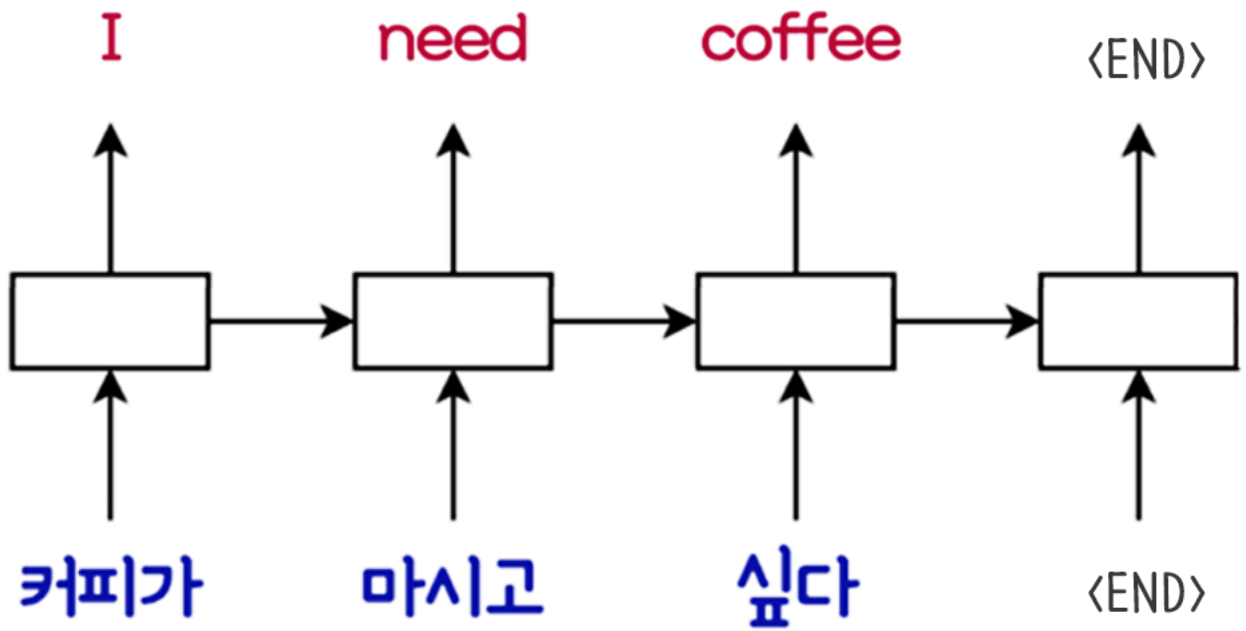
나는 누군가를 찾고 있습니다.

기존의 방법이 대체 어땠길래?



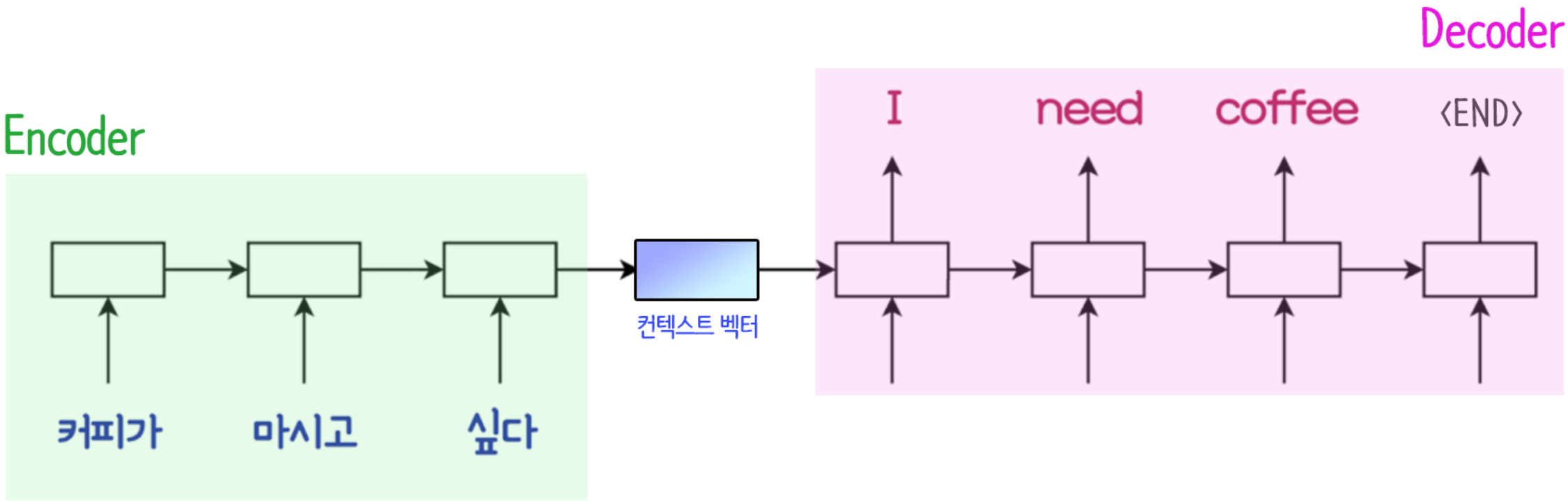
▲ Vanilla RNN

기존의 방법이 대체 어땠길래?



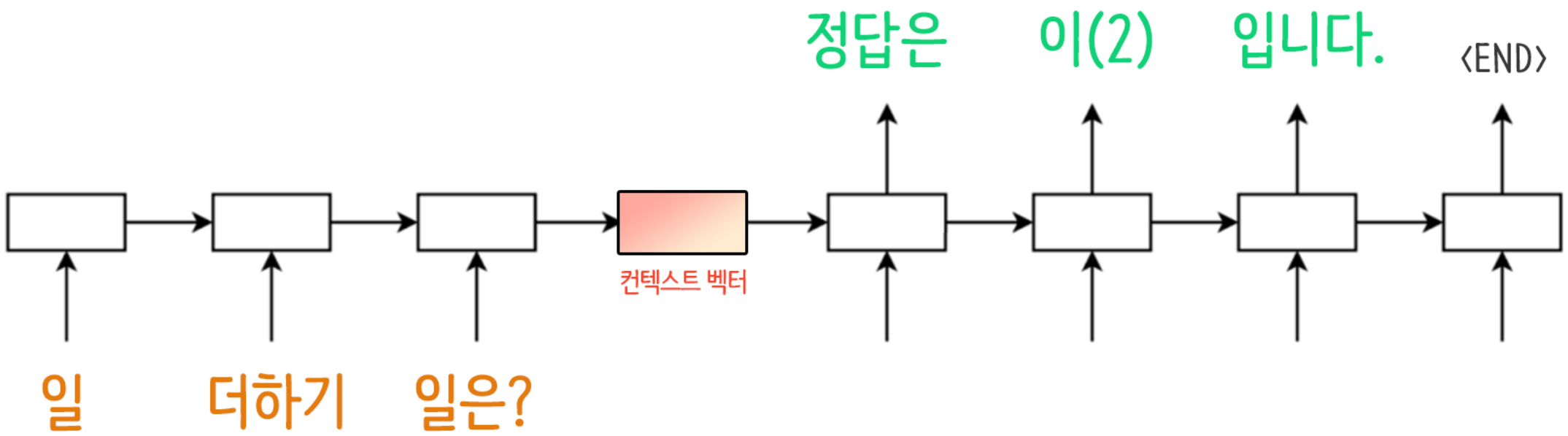
▲ Vanilla RNN

기존의 방법이 대체 어땠길래?



▲ Sequence-to-Sequence (Seq2seq)

기존의 방법이 대체 어땠길래?

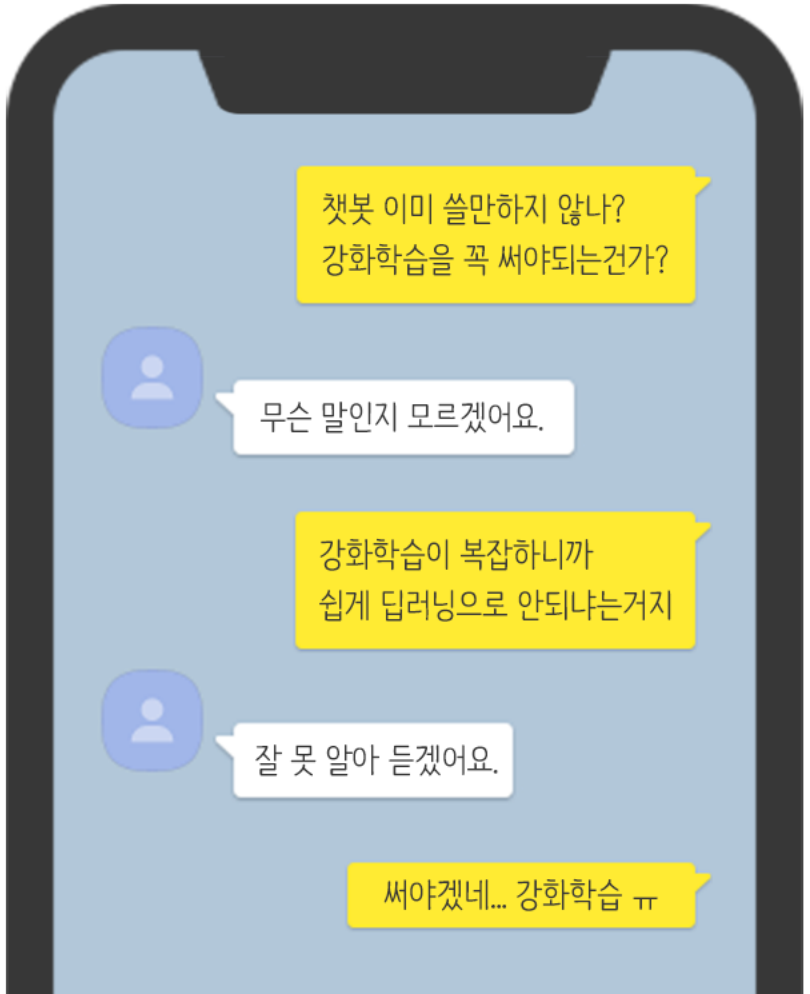


▲Seq2seq을 챗봇으로?

CrossEntropy 기반의 문제점

실제 정답	학습 데이터	모델 출력
“		
Yeah 다시 돌아왔지	Yeah 다시 돌아왔지 ‹PAD› ‹PAD› ‹PAD›	‹PAD› ‹PAD› ‹PAD› ‹PAD› ‹PAD› ‹PAD›
내 이름 레인	내 이름 레인 ‹PAD› ‹PAD› ‹PAD›	‹PAD› ‹PAD› ‹PAD› ‹PAD› ‹PAD› ‹PAD›
스웩을 뽐내 WHOO!	스웩을 뽐내 WHOO! ‹PAD› ‹PAD› ‹PAD›	‹PAD› ‹PAD› ‹PAD› ‹PAD› ‹PAD› ‹PAD›
They call it! 왕의 귀환	They call it! 왕의 귀환 ‹PAD›	‹PAD› ‹PAD› ‹PAD› ‹PAD› ‹PAD› ‹PAD›
후배들 바빠지는 중!	후배들 바빠지는 중! ‹PAD› ‹PAD› ‹PAD›	‹PAD› ‹PAD› ‹PAD› ‹PAD› ‹PAD› ‹PAD›
”		

Dull 떨어진 문장들



Baseline mutual information model (Li et al. 2015)
A: Where are you going? (1)
B: I'm going to the restroom. (2)
A: See you later. (3)
B: See you later. (4)
A: See you later. (5)
B: See you later. (6)
...
...
A: how old are you? (1)
B: I'm 16. (2)
A: 16? (3)
B: I don't know what you are talking about. (4)
A: You don't know what you are saying. (5)
B: I don't know what you are talking about . (6)
A: You don't know what you are saying. (7)
...

그래서 핵심은...

“

CrossEntropy 기반 Seq2seq의 고질병을

강화학습으로 극복!

= Loss에 집중! #Policy #Reward

”

좋은 대화의 3요소 *(Reward)*

정보성

일관성

응답 용이성

좋은 대화의 3요소 *(Reward)*

문장 1, 문장 2

$$State = [\, p_i,\, q_i \,]$$

$$Action = p_{i+1}$$ 문장 3

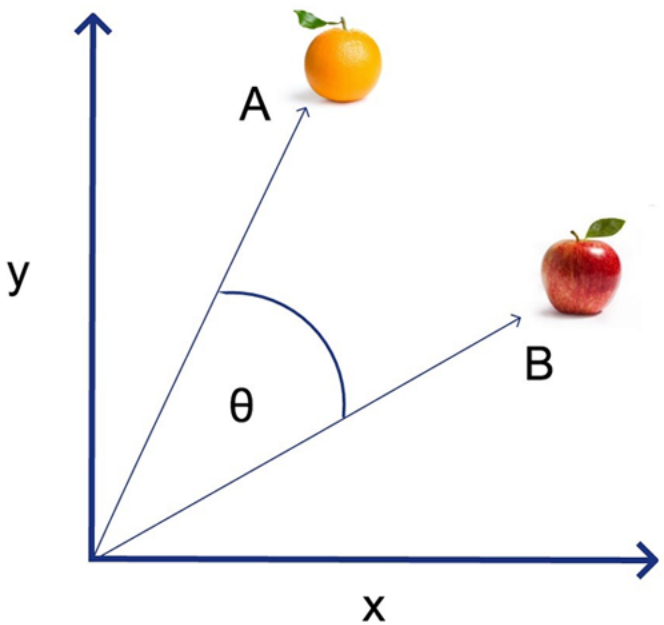
$$Policy = p_{RL}\left(p_{i+1} \middle| p_i,\, q_i \right)$$

정보

용이성

정보성 Reward

Cosine Similarity



$$r_2 = -\log\cos\frac{h_{p_i} \cdot h_{p_{i+1}}}{\|h_{p_i}\|\|h_{p_{i+1}}\|}$$

유사

A: 오늘 뭐할거야?
B: 피씨방 갔다가 술 마실거야.
A: 내일 뭐할거야?
B: 피씨방 갔다가 술 마실거야.

유사

▲ Bad Case

A: 오늘 뭐할거야?
B: 피씨방 갔다가 술 마실거야.
A: 그럼 내일은?
B: 영화보고 볼링을 칠까?

▲ Good Case

일관성 Reward

A: 오늘 뭐할거야?
B: 피씨방 갔다가 술 마실거야.
A: 그럼 내일은?
B: 영화보고 볼링을 칠까?
A: 너무 좋지!

Predict!

A: 오늘 뭐할거야?
B: 피씨방 갔다가 술 마실거야.
A: 그럼 내일은?
B: 영화보고 볼링을 칠까?
A: 너무 좋지!

Predict!

Reward!

$$r_3 = \frac{1}{N_a} \log p_{\text{seq2seq}}(a|q_i, p_i) + \frac{1}{N_{q_i}} \log p^{\text{backward}}_{\text{seq2seq}}(q_i|a)$$

응답용이성 Reward

Dull 떨어진 답변 리스트

잘 모르겠어요.
이해가 안돼요.
당신이 뭐라고 하는지 모르겠어요.
왜요?
무슨 뜻이에요?
…

A: 오늘 뭐할거야?
B: 피씨방 갔다가 술 마실거야.
A: 왜? Loss++

A: 이거 논문 제목이 뭐야?
B: 잘 모르겠어. Loss++

$$r_1 = -\frac{1}{N_{\mathbb{S}}} \sum_{s \in \mathbb{S}} \frac{1}{N_s} \log p_{\text{seq2seq}}(s|a) \qquad (1)$$

최종 Reward

$$r(a, [p_i, q_i]) = \lambda_1 r_1 + \lambda_2 r_2 + \lambda_3 r_3 \quad (4)$$

where $\lambda_1 + \lambda_2 + \lambda_3 = 1$. We set $\lambda_1 = 0.25$, $\lambda_2 = 0.25$ and $\lambda_3 = 0.5$. A reward is observed after the agent reaches the end of each sentence.

실험

Agent A

Agent B

Dataset

OpenSubtitles Dataset

```
5
00:01:15,200 --> 00:01:20,764
Nehmt die Halme, schlagt sie oben ab,
entfernt die Blätter

6
00:01:21,120 --> 00:01:24,090
und werft alles auf einen Haufen
für den Pflanztrupp.

7
00:01:24,880 --> 00:01:30,489
Das Zuckerrohr beißt euch nicht.
Nicht so zaghaft! Na los, Burschen, los!
```

실험

Agent A

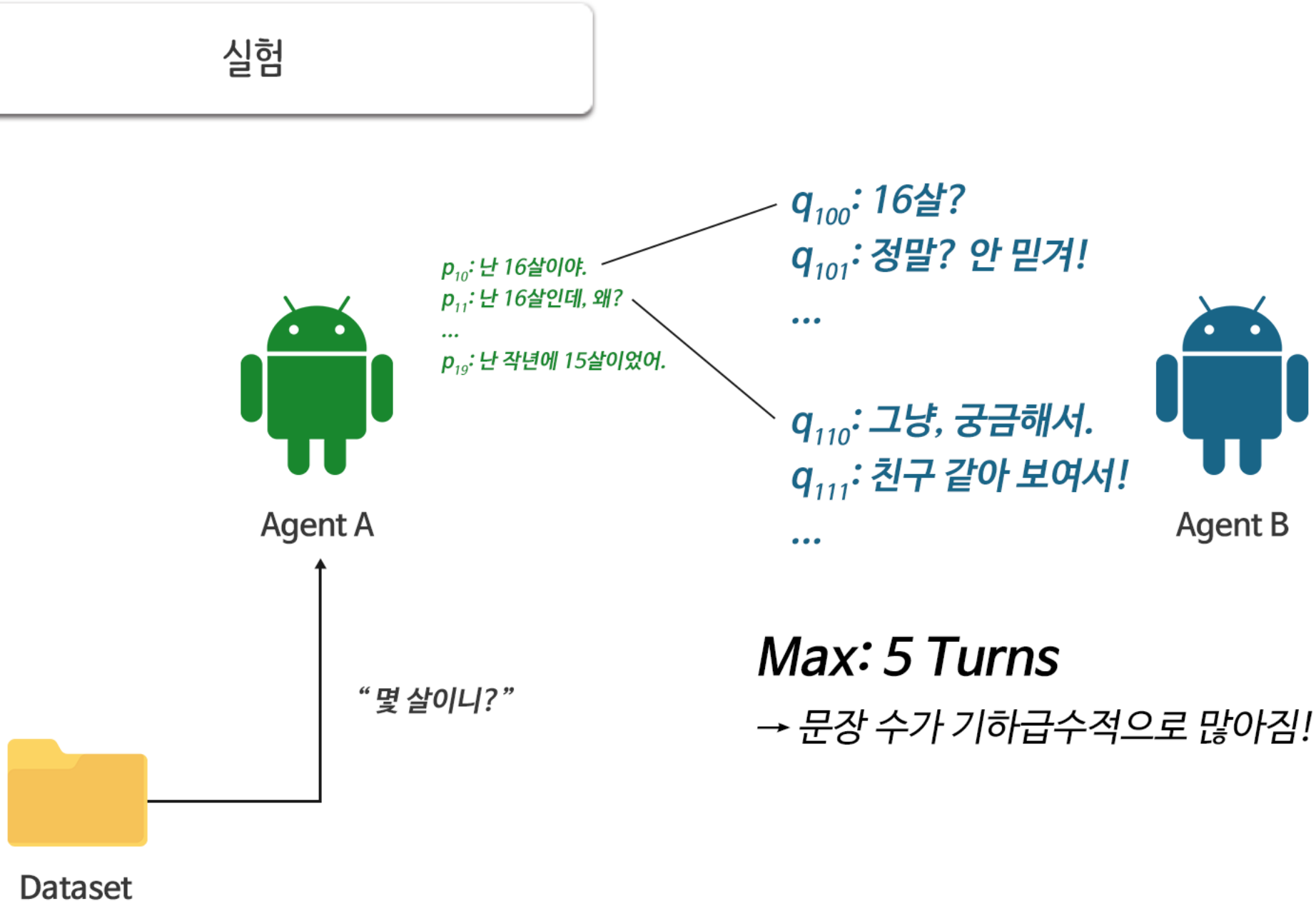
p_{10}: 난 16살이야.
p_{11}: 난 16살인데, 왜?
...
p_{19}: 난 작년에 15살이었어.

Agent B

"몇 살이니?"

Dataset

실험



Max: 5 Turns

→ 문장 수가 기하급수적으로 많아짐!

실험 결과

Model	# of simulated turns
SEQ2SEQ	2.68
mutual information	3.40
RL	4.48

Table 2: The average number of simulated turns from standard SEQ2SEQ models, mutual information model and the proposed RL model.

```
if sentence[idx] in dull_sentences:
    break

elif mean(sentence[idx-1] == sentence[idx]) > 0.8:
    break
```

실험 결과

Setting	RL-win	RL-lose	Tie
● single-turn general quality	0.40	0.36	0.24
● single-turn ease to answer	0.52	0.23	0.25
● multi-turn general quality	0.72	0.12	0.16

Table 5: RL gains over the mutual information system based on pairwise human judgments.

Q.
"오늘 무슨 요일이야?"의
답으로 가장 적합한 것은?
(복수 선택 가능)

1) 오늘은 금요일이야.
2) 어제는 휴일이었어.
3) 배고프다.

Q.
"어제는 많이 피곤했어."의
대답 중 가장 답변하기 좋은 것은?
(복수 선택 가능)

1) 힘들었겠다.
2) 왜? 무슨 일있었어?
3) 배고파 정말로.

Q.
다음 중 가장 자연스러운 대화는?

A: 오늘 점심은 뭘 먹지?
B: 숙이네로 가자.
A: 진짜 지겹다. 다른건 없어?
B: 그럼 숙이네 갈까?
A: 좋아, 바로 가자!

A: 공부 많이 했어?
B: 아니 하나도... 너는?
A: 나도 하나도.. 참 어려워.
B: 그러게 말이야... 피곤하지?
A: 맞아. 너는 어때?

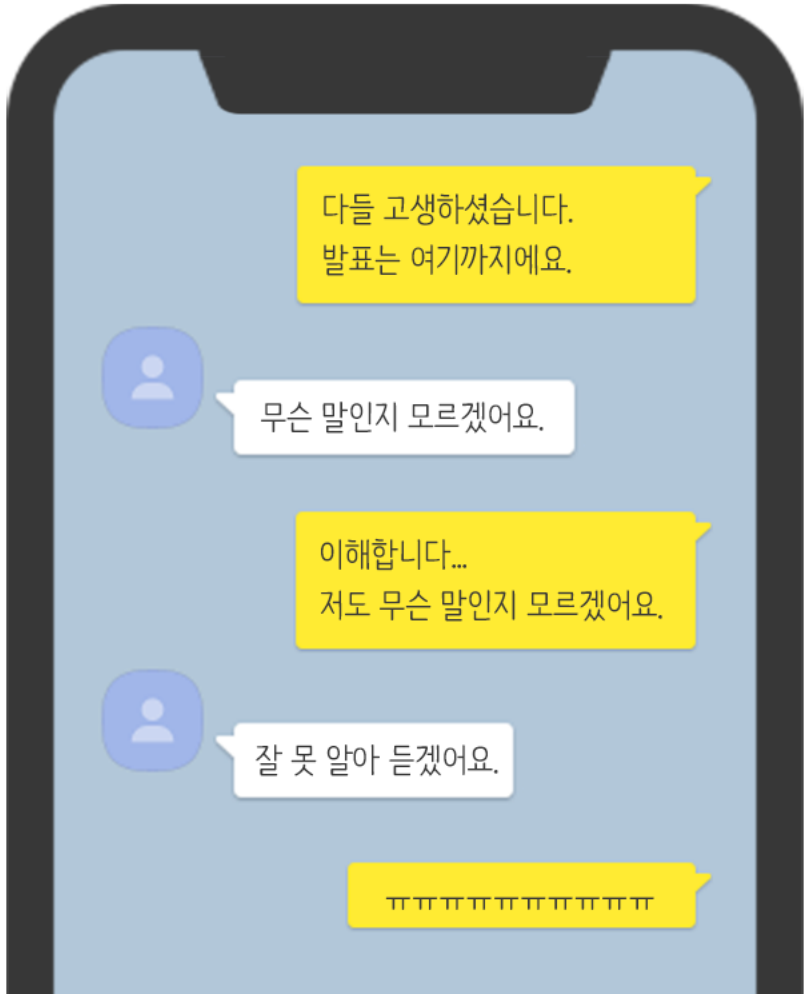
...

실험 결과

Input Message	Mutual Information Model	RL
How old are you?	I'm 16.	I'm 16, why are you asking?
What's your full name?	I have no idea.	What's yours?
I don't want to go home tonight.	Really?	Why?
Do you have any feelings for me?	I don't know what you are talking about.	I don't want to hurt your feelings.
How much time do you have here?	Not long enough. Sorry, sir.	Ten seconds.
Shall we get started?	Of course!	Yes. We've got a lot of work to do here.
Do you play football?	No, i don't	Yes. I love football!
We'd have to talk to him.	I mean, he's a good guy	About what ?
How come you never say it?	Because I don't want to hurt you.	I don't think it's a good idea to say it.

Table 3: Sampled responses generated from the mutual information models and the proposed RL model.

THANK YOU !


다들 고생하셨습니다.
발표는 여기까지에요.
무슨 말인지 모르겠어요.
이해합니다...
저도 무슨 말인지 모르겠어요.
잘 못 알아 듣겠어요.
ㅠㅠㅠㅠㅠㅠㅠㅠㅠ