

Week1-1 과제

Paperswithcode에서 NLP sub task 중에 2개를 선택하여 본인 블로그에 정리해보세요.
요. task 별로 아래 3가지 항목에 대해서 정리하세요.

1. Dialogue Generation

문제정의

시리, 알렉사 같은 가상 비서가 사용자의 요청에 대한 응답으로 적절한 답변 반환해야 하는 텍스트 생성 문제

데이터 소개

Persona-Chat

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.

Table 2: Example dialog from the PERSONA-CHAT dataset. Person 1 is given their own persona (top left) at the beginning of the chat, but does not know the persona of Person 2, and vice-versa. They have to get to know each other during the conversation.

각 대화는 인공 페르소나(가상 인물)를 가정한 두 명의 크라우드 소싱 작업자 사이에서 수행되었다. 가상 인물들의 특징은 3~5개 문장으로 설명됨.

- 페르소나1, 페르소나2의 프로필 문장과 각 페르소나 간 교차 대화로 구성됨
- 훈련 데이터 : 8939개의 완전한 대화
- 검증 데이터 : 1000개 대화
- 테스트 데이터 : 968개 대화

SOTA 모델 소개

P² Bot

주요 키워드 : 인간과 유사한 응답 모방이 아닌 상대 이해 기반 모델링, 상호 페르소나 인식을 통합해 개인화된 대화 생성 품질 향상

2. Text Summarization

문제정의

원본 문장에서 중요한 정보를 추출하고 요약하는 문제

데이터 소개

WikiHow

How to Help Save Rivers

Method 1: Reducing Your Water Usage

- 1 **Take quicker showers to conserve water.** One easy way to conserve water is to cut down on your shower time. Practice cutting your showers down to 10 minutes, then 7, then 5. Challenge yourself to take a shorter shower every day.
- 2 **Wait for a full load of clothing before running a washing machine.** Washing machines take up a lot of water and electricity, so running a cycle for a couple of articles of clothing is inefficient. Hold off on laundry until you can fill the machine.
- 3 **Turn off the water when you're not using it.** Avoid letting the water run while you're brushing your teeth or shaving. Keep your hoses and faucets turned off as much as possible. When you need them, use them sparingly.

Method 2: Using River-Friendly Products

- 1 **Select biodegradable cleaning products.** Any chemicals you use in your home end up back in the water supply. Choose natural soaps or create your own cleaning and disinfecting agents out of vinegar, baking soda, lemon juice, and other natural products. These products have far less of a negative impact if they reach a river.
- 2 **Choose recycled products instead of new ones.** New products take way more water to make than recycled products. Reuse what you already own when possible. If you need to buy something, opt for products made out of recycled paper or other reused material.

Article 1:

One easy way to conserve water is to cut down on your shower time. Practice cutting your showers down to 10 minutes, then 7, then 5. Challenge yourself to take a shorter shower every day. Washing machines take up a lot of water and electricity, so running a cycle for a couple of articles of clothing is inefficient. Hold off on laundry until you can fill the machine. Avoid letting the water run while you're brushing your teeth or shaving. Keep your hoses and faucets turned off as much as possible. When you need them, use them sparingly.

Summary 1:

Take quicker showers to conserve water. Wait for a full load of clothing before running a washing machine. Turn off the water when you're not using it.

Article 2:

Any chemicals you use in your home end up back in the water supply. Choose natural soaps or create your own cleaning and disinfecting agents out of vinegar, baking soda, lemon juice, and other natural products. These products have far less of a negative impact if they reach a river. New products take way more water to make than recycled products. Reuse what you already own when possible. If you need to buy something, opt for products made out of recycled paper or other reused material.

Summary 2:

Select biodegradable cleaning products. Choose recycled products instead of new ones.

다양한 저자가 다양한 분야에 관해 작성한 온라인 지식을 기반으로 하는 데이터셋

- 230,000 개 이상의 문서와 요약 쌍을 제공

SOTA 모델 소개

BertSum

주요 키워드 : 문장 분할 및 구두점 복원, 지능형 가상 비서로 통합될 것을 기대