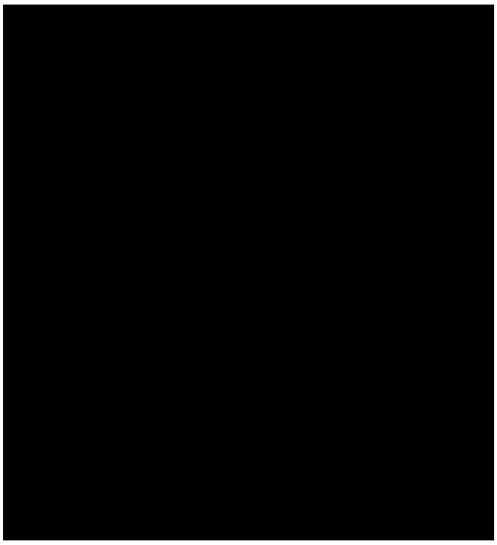
2021-1 소프트웨어 종합설계

페르소나 프로필 제공을 통한 챗봇 모델 성능 향상



차례

Ⅰ. 연구 주제

Ⅱ. 선행 연구와 연구 필요성

1. 연구 필요성

2. 선행 연구

3. 기대 결과

Ⅲ. 연구 내용

Ⅳ. 현재 진행 상황

Ⅴ. 일정 및 역할배분

Ⅰ. 연구 주제

본래 연구 주제는 '지식 그래프를 이용한 챗봇 대화 생성 기술'로 지식 그래프에 백과사전형 지식 뿐만 아니라 생활 상식이나 시사 내용, 주기적으로 수집한 최근 화제 데이터를 지식그래프에 자동으로 적용하여 확장하고 챗봇에 특정 페르소나의 말투를 반영하여 자연스러운 대화가 가능하도록 하는 것이 목표였다. 하지만 연구지도 때 상의 후, 교수님께서 추천해주신 논문을 읽어보았더니 페르소나에 더 관심이 갔고 연구실에서 관련 주제를 하고 있어서 페르소나 연구가 더 용이하다고 생각하여 주제를 수정하였다. 수정한 주제는 '페르소나 프로필 제공을 통한 챗봇 모델 성능 향상'이다.

Ⅱ. 선행 연구와 연구 필요성

1. 연구 필요성

기존 chit-chat model은 다양한 발화자들의 대화에 기반해 학습되었기 때문에 일관된 personality가 부족하고 최근 발화 기록에서 주어지는 가중치로 인해 long-term memory가 빈약하다. 또한, 주어진 페르소나 문장과 밀접하게 일치하지 않는 문맥에 반응하는데 종종 어려움을 겪으며 'I don't know'와 같이 특정 정보를 담고 있지 않은 답변을 내는 경향이 있다. 그리고 일반적인 chit-chat 모델에 대한 공개적인 dataset이 부족해 conversation model의 질이 낮고 모델 평가가 어렵다.

2. 선행 연구

1) Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, Jason Westion(2018), "**Personalizing Dialogue Agents : I have a dog, do you have pets too?**", Association for Computational Linguistics, p2204-p2213

   : 일정한 페르소나를 가지고 있는 dataset인 profile을 모델에 제공하고 임의의 crowd workers 간의 대화를 포함한 대화 dataset을 수집하여 학습시킨 다음 Sequence-to-Sequence 모델과 memory network를 포함한 여러 generative model, ranking model를 사용하여 다음 발화 생성 성능과 프로필 예측 태스크의 성능을 향상시키고자 하였다.

2) Bodhisattwa Prasad Majumder, Harsh Jhamtani, Talyor Berg-kirkpatrick, Julian McAuley(2020), "**Persona-grounded Dialog with Commonsense Expansions**", Association for Computational Linguistics, p9194-9206

   : 풍부하고 다양한 상식적 확장을 생성하는 프레임워크인 COMET과 pre-trained된 언어 번역 모델로 역번역을 하는 가성 의역 시스템, 문장을 세밀하게 선택할 수 있는 COMPAC을 사용하여 대화 문맥과 더 일관적인 페르소나 기반 챗봇을 만들고자 하였다.

3. 기대 결과

여러 dataset을 이용할 뿐만 아니라 페르소나 프로필 제공을 통해 일관된 personality를 챗봇에 부여하고 long-term memory를 향상시켜 주어진 페르소나 문장과 밀접하게 일치하지 않은 문맥에 적절하게 반응하도록 한다. 결국 이를 통해 자연스럽게 대화하는 도중에 다음 발화를 잘 예측하고 대화 내역에 따라 프로필을 잘 예측하도록 한다.

Ⅲ. 연구 내용

선행 연구에서는 chit-chat model의 training 과정에서 Amazon Mechanical Turk를 통하여 수집한 crowd-sourced dataset인 PERSONA-CHAT을 사용하였으며, 이 dataset은 기존 dataset보다 매력적이고 persona가 잘 드러나는 chit-chat dialog를 제공한다. 또한 최소 5개의 profile 문장을 포함한 1155개의 persona를 보유하고 있고, challenging한 실험을 위한 revised persona와 한 쌍의 persona로 구성된 persona chat으로 구성되어 있다.

선행 연구의 결과는 ranking model의 경우 revised persona dataset으로 training을 진행하고 original persona dataset으로 test를 진행했을 때 가장 높은 적중률을 보여주었고, generative model의 경우 self persona - profile memory model이 가장 높은 적중률과 가장 낮은 perplexity를 보여주었다.

본 연구에서는 chit-chat 대화 모델을 개선하고자 한 선행 연구에서 한계를 찾고, 그에 대한 개선점을 제시하고자 한다. 우선적으로는 대화 모델 training에 사용된 dataset을 중점적으로 분석하고, 그에 대한 한계와 개선점을 분석하는 과정을 진행중에 있다. 대화 모델 training과 test는 ParlAI 플랫폼을 이용하였고 ranking model 중 가장 높은 정확도를 보여주었던 KV profile memory에 self persona를 적용하여 직접 모델 training을 진행한 후, 그 결과 중 일치하지 않았던 응답을 정성적으로 분석하여 기존의 한계를 찾고자 하였다.

Training에 사용된 옵션과 모델의 best_valid는 사진1,2와 같다. 해당 모델의 hits@1 성능은 0.474, 로, 논문에서 제시한 점수인 0.511에 근접하여 나타났고, 더 정밀한 분석을 위하여 많은 시간의 training을 진행한 후에 재분석을 진행할 예정이다.

해당 모델을 test set으로 검증해본 결과, 사진3,4와 같이 ground truth로 설정된 label은 상대의 질문의 내용에 동조하는 답변을 반환했지만, model의 경우 persona와 관련이 있는 선택지를 반환했다.

```
{
    "batchsize": 1,
    "num_epochs": -1,
    "embeddingsize": 500,
    "embeddingnorm": 10,
    "share_embeddings": true,
    "learningrate": 0.1,
    "margin": 0.1,
    "loss": "cosine",
    "optimizer": "sgd",
    "tfidf": false,
    "cache_size": 1000,
    "history_length": 100,
    "history_replies": "label",
    "parlai_home": "/home/junho/.local/lib/python3.8/site-packages",
    "override": {
        "model_file": "model/kvmemnn/self_persona",
        "task": "personachat:self",
        "model": "projects.personachat.kvmemnn.kvmemnn:KvmemnnAgent",
        "log_every_n_secs": 60.0,
        "validation_max_exs": 10000,
        "validation_every_n_secs": 900.0,
        "hops": 1,
        "lins": 0,
        "validation_patience": -1,
        "validation_metric": "accuracy",
        "max_train_time": 10800.0,
        "share_embeddings": true,
        "batchsize": 1,
        "learningrate": 0.1,
        "embeddingsize": 500,
        "margin": 0.1,
        "tfidf": false,
        "numthreads": 10
    },
}
```


<사진1>


```
{
    "parleys": 3339775,
    "train_time": 1809.5328907966614,
    "total_epochs": 50.81901733136536,
    "valid_reports":
        {
            "exs": 7801,
            "accuracy": 0.4746827329829509,
            "f1": 0.5300030947408856,
            "bleu-4": 0.4749063179383941,
            "hits@1": 0.4746827329829509,
            "hits@5": 0.7702858607870786,
            "hits@10": 0.891039610306371,
            "hits@100": 1.0,
            "train_time": 1809.415697336197,
            "parleys": 3339775,
            "total_exs": 3339775,
            "total_epochs": 50.81901733136536
        },
    "best_valid": 0.4746827329829509
}
```


<사진2>

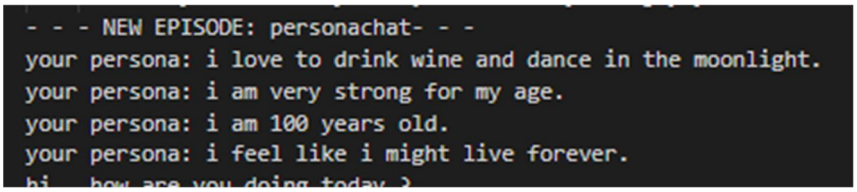
```
- - - NEW EPISODE: personachat- - -
your persona: i volunteer at a soup kitchen.
your persona: cheeseburgers are my favorite food.
your persona: i was poor growing up.
your persona: i like watching war documentaries.
```

<사진3>

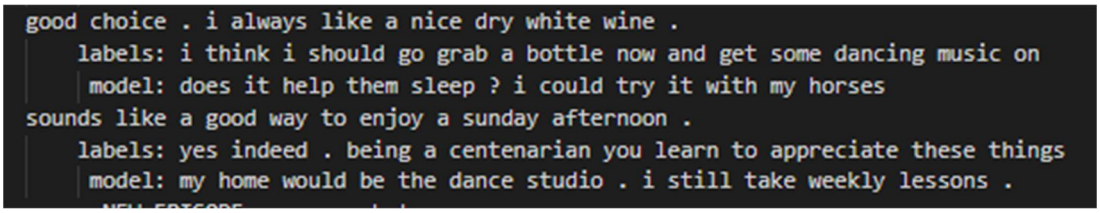
```
- - - NEW EPISODE: personachat- - -
your persona: i volunteer at a soup kitchen.
your persona: cheeseburgers are my favorite food.
your persona: i was poor growing up.
your persona: i like watching war documentaries.
```

<사진4>

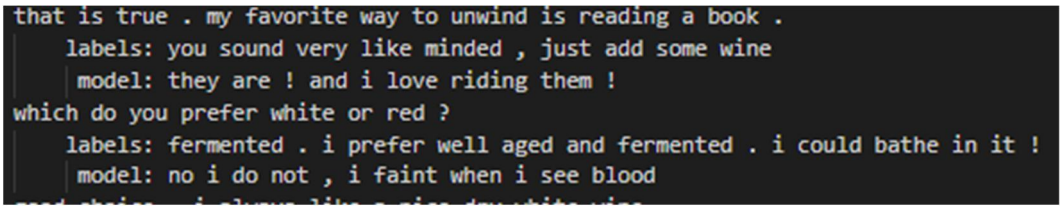
다른 대화에서도 비슷한 예시를 찾아볼 수 있다. 사진6에서 볼 수 있듯이 자연스러운 대화의 맥락은 상대의 질문에 동조하는 답변을 반환하는 것이지만 해당 모델에서는 persona로 주어진 dance에 집중하여 다른 선택지를 반환하였다. 또 다른 문제점은 사진7과 같이 문맥상 'white or red'는 화이트 와인 또는 레드 와인 이란 뜻이지만 이를 문자 그대로 해석하여 red와 연관이 있는 blood가 포함된 선택지를 반환하는 모습을 볼 수 있었다.



<사진5>



<사진6>

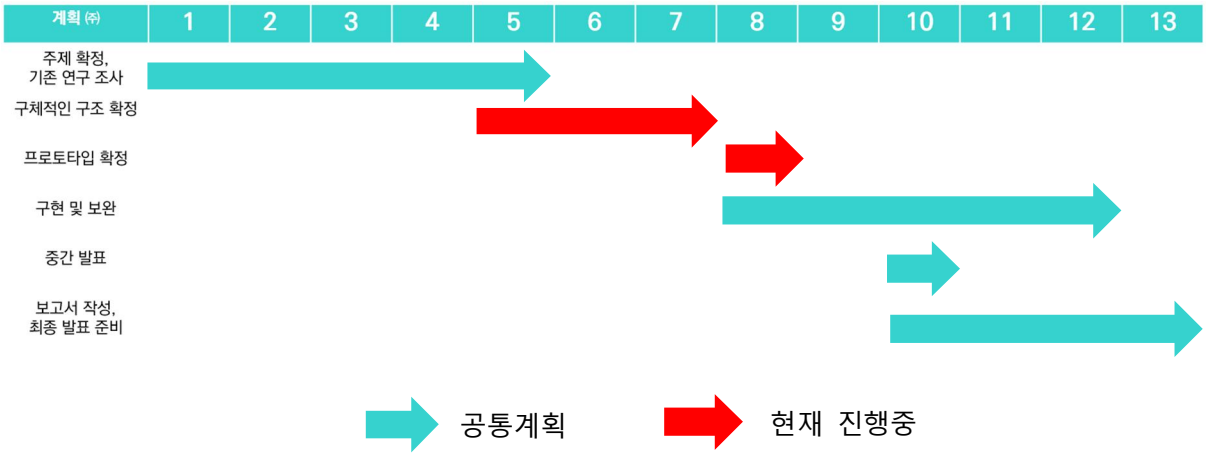


<사진7>

IV. 현재 진행 상황

- 지식그래프에 기반한 챗봇 모델 구현을 계획했던 처음 구상과 달리 페르소나 분야에 더 중점을 두어 개발하도록 결정함.
- [2018][ACL][Zhang et al.]Personalizing Dialogue Agents; I have a dog, do you have pets too 논문 리딩과 코드 리뷰 발표 진행.
- 위 논문의 결과에 대한 코드실습과 훈련 완료.
- 위 논문에서 소개한 데이터셋과 모델의 한계점을 찾고 개선점을 본 팀의 페르소나 모델에 적용하려 구상 중.

V. 일정 및 역할 배분



~6주차 세부 일

[redacted]: 페르소나 논문 선행연구 발표 진행, 연구 제안 발표 및 보고서 작성

: 페르소나 논문 실험 결과 발표, 발표자료 및 보고서 작성

: 페르소나 논문 코드 리뷰와 소개한 모델 성능비교 진행, 연구지도 일정 관리, 발표 자료, 보고서 작성