

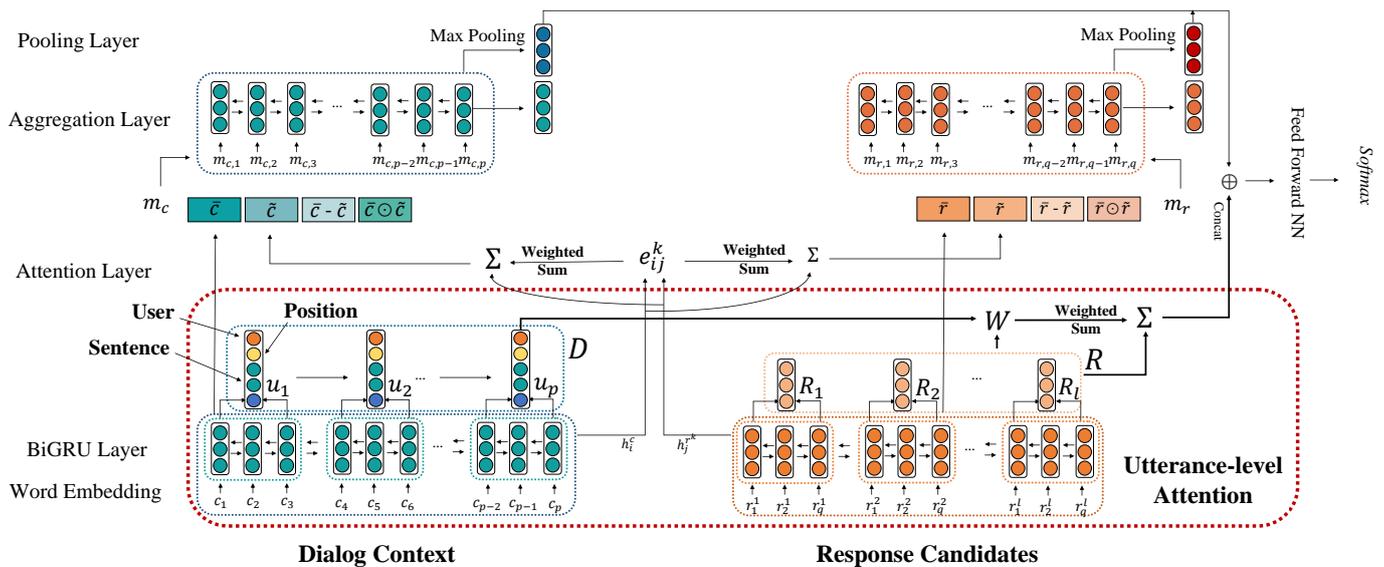
Multi-turn Response Selection

Enhanced Sequential Representation Augmented with Utterance-level Attention for Response Selection

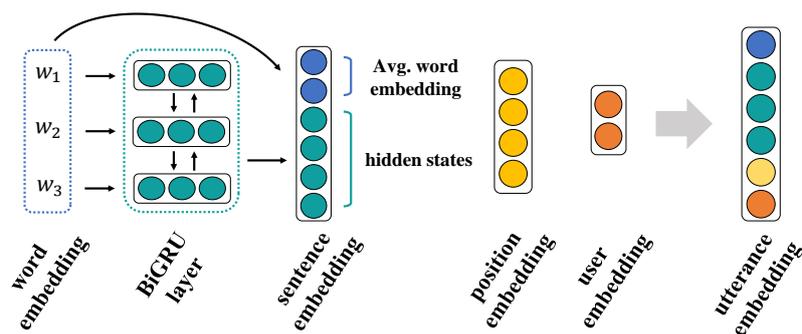
- Whang, T., Lee, D., Lee, C., & Lim, H. (2020). Enhanced Sequential Representation Augmented with Utterance-level Attention for Response Selection. AAI 2019 Workshop on Dialogue System Technology Challenge 7.
- Github: <https://github.com/taesunwhang/dstc7>

01. 연구 내용

- 기존의 검색 기반 대화 시스템 중 일부는 대화 전체를 하나의 문장으로 간주하여 모델의 입력으로 넣기 때문에, 발화 단위의 자질 정보를 충분히 담지 못하는 한계가 존재합니다. 기존 많은 연구들이 RNN 기반의 모델로 설계되었기 때문에 대화의 길이가 길어질 경우 대화 초반과 중반에 있는 발화 정보를 손실할 수 있습니다.
- 아래 그림은 본 연구에서 제안한 모델의 전체 구조도입니다. 자연어 추론 (Natural Language Inference)에서 좋은 성능을 보였던 ESIM (Enhanced Sequential Inference Model)을 기반으로 하여 대화 내 단어 단위 정보를 학습하였고, 발화 단위 임베딩을 추가해 Attention 방식으로 중요한 정보들을 추가해 줌으로써 향상된 성능을 얻을 수 있었습니다.



- 본 연구에서는 1) Sentence 2) Utterance Position 3) User 와 같이 다양한 발화 정보를 모델링하여, 발화 단위 정보를 학습할 수 있도록 설계하였습니다. 각 임베딩은 모두 결합 (concatenation)되어 모델의 입력으로 들어가게 되며, 발화 임베딩을 구성하는 방식은 아래의 그림과 같습니다.



- **(Sentence Embedding)** 문장 내 단어 임베딩의 평균과 양방향 GRU 문장 자질들을 결합하여 발화의 문맥적 자질 정보를 추출하였습니다.
- **(Utterance Position Embedding)** 대화 내 각 단어들이 어떤 발화에 포함되는지를 학습시키기 위해 발화 위치 정보를 모델링하였습니다. 위치 정보를 모델링하는 과정에는 Vaswani et al. (2017)¹에서 제안한 Sinusoid 방법을 사용하였습니다.
- **(User Embedding)** 일대일 또는 여러 명의 화자 간의 대화에서는 각 화자들의 역할 등이 정해져 있습니다. 본 연구에서 사용한 Advising Corpus를 예로 들면, 한 명은 수업 수강을 위한 조언을 구하는 학생의 입장, 다른 한명은 수강 신청을 위해 상담을 해주는 상담사의 역할을 합니다. 이처럼 각 화자는 대화 내에서 각자의 역할을 지니고 있으며, 이러한 정보를 모델이 포착할 수 있도록 모델링하였습니다.

02. 실험 및 결과

- 아래 표는 DSTC 7 Track 1의 task 1과 task 4에 대한 validation set 성능 비교를 진행한 결과입니다. 두 개의 task 모두 주어진 응답 후보 100개에 대해 하나의 정답을 고르는 문제이며, task 4의 경우 정답이 없는 경우가 존재합니다.
- ESIM을 baseline으로 두어 SE (Sentence Embedding), PE (Position Embedding), UE (User Embedding)를 추가한 결과 향상된 성능을 얻었으며, 단어 임베딩으로 ELMo를 사용하였을 때 GloVe를 사용하였을 때보다 효과적이라는 사실 또한 확인하였습니다.

Task 1	Ubuntu						Advising					
	R@1	R@2	R@5	R@10	R@50	MRR	R@1	R@2	R@5	R@10	R@50	MRR
(Lowe et al. 2015)	0.211	0.307	0.446	0.569	0.921	0.320	0.074	0.108	0.210	0.342	0.802	0.162
(Dong and Huang 2018)	0.367	0.452	0.558	0.651	0.917	0.465	0.086	0.156	0.256	0.376	0.834	0.187
ESIM + SE (GloVe)	0.377	0.460	0.568	0.657	0.929	0.473	0.098	0.160	0.294	0.430	0.834	0.204
ESIM + SE + PE + UE (GloVe)	0.384	0.464	0.575	0.662	0.921	0.480	0.112	0.166	0.298	0.438	0.859	0.214
ESIM + SE + PE + UE (ELMo)	0.406	0.493	0.606	0.691	0.928	0.505	0.106	0.160	0.306	0.460	0.858	0.213

Task 4	Ubuntu						Advising					
	R@1	R@2	R@5	R@10	R@50	MRR	R@1	R@2	R@5	R@10	R@50	MRR
(Lowe et al. 2015)	0.215	0.328	0.500	0.622	0.925	0.351	0.162	0.200	0.262	0.346	0.736	0.228
(Dong and Huang 2018)	0.378	0.507	0.634	0.717	0.931	0.500	0.156	0.242	0.400	0.546	0.888	0.278
ESIM + SE (GloVe)	0.386	0.512	0.636	0.722	0.939	0.507	0.170	0.278	0.406	0.536	0.894	0.293
ESIM + SE + PE + UE (GloVe)	0.394	0.515	0.651	0.728	0.941	0.512	0.190	0.272	0.424	0.578	0.908	0.311
ESIM + SE + PE + UE (ELMo)	0.417	0.542	0.662	0.735	0.938	0.533	0.186	0.266	0.386	0.498	0.854	0.291

¹Vaswani et al. (2017) Attention is all you need. In NIPS 2017.