Improved Multi-Round Chatbot using Data Augmentations

Meiwei Zhang

April 2020, Athlone Institute of Technology

Content

- Introduction
- Knowledge Manager Model.
- Loss design
- Data Augmentations
- Tricks Sharing
- Results

Introduction

- Improve the smooth between different topics
- Use knowledge graph in Chatbot
- Use Data Augmentations, including text matching, NER genration, conversation selection

Knowledge Manager Model

- prior and posterior distributions over knowledge are used to facilitate knowledge selection. prior distribution only includes utterances, and posterior distribution contains response, knowledge manager try to approximate the two model during training.

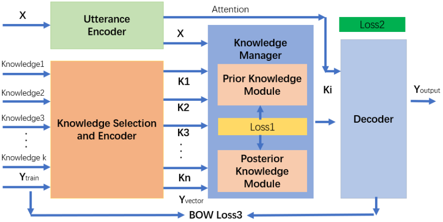

Figure: Knowledge Manager

Loss design

Generation Loss

Use the cross entropy of the real dialogue and the generated text dialogue as the first loss function

Selection Loss

There are two loss functions, the first is the loss of similarity between the prior and a posteriori knowledge modules, and the second is the loss of the correlation between the selected knowledge and the real response

Training Data


```
{
    "goal":
    [
        ["START", "our music", "Ann-Marie"]
        ...
    ],
    "knowledge":
    [
        ["our music", "type", "plot"],
        ["our music", "domain", "movie"]
        ...
    ],
    "conversation":
    [
        "what's the movie name?",
        "our music."
        ...
    ]
}
```

Data Augmentations knowledge selection

- Improve the accuracy of the knowledge selection module by calculating the correlation between actual replies and numerous knowledge, filtering knowledge that is not relevant to the current output

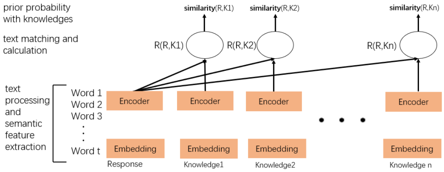

Figure: Knowledge selection

Data Augmentations

NER Generation

Entities, such as person names, industry specific terminologies, appear frequently in knowledge driven dialogues. These discrete entities, even in the same category, have very various forms, resulting in data sparseness that is believed to be a potential reason for overfitting. Our models are supposed to learn the abstract representations of the entity in the specific context, rather than remembering the entity itself. Therefore, we generalized the entities of same category with a unique string. For example, we replaced person names "Ma Long" with the string of "person name"

Data Augmentations conversation selection

- Try to control the number of conversation rounds to change the training data

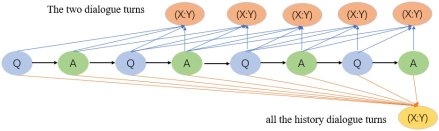

Figure: conversation selection

Tricks Sharing

- In the training process, we found that the difference between the three loss functions is relatively large, so we adopted two tips. The first is to train in stages, that is, to first train the loss function with large magnitude, and the second is to add regularization.

```
epoch 0 step 300 | bow loss 43.136368 kl loss 0.033205 nll loss 86.616600 total loss 43.136368
epoch 0 step 600 | bow loss 38.504749 kl loss 0.037576 nll loss 86.384499 total loss 38.504749
epoch 0 step 900 | bow loss 37.805897 kl loss 0.034011 nll loss 86.941132 total loss 37.805897
valid dataset: bow loss 36.984787 kl loss 0.040510 nll loss 86.980247 total loss 124.005539
epoch 0 step 1200 | bow loss 36.126198 kl loss 0.037957 nll loss 85.520470 total loss 36.126198
epoch 0 step 1500 | bow loss 35.788834 kl loss 0.036625 nll loss 85.708076 total loss 35.788834
epoch 0 step 1800 | bow loss 35.995262 kl loss 0.035108 nll loss 86.896278 total loss 35.995262
valid dataset: bow loss 35.618786 kl loss 0.036790 nll loss 86.970612 total loss 122.626221
epoch 0 step 2100 | bow loss 35.625717 kl loss 0.034435 nll loss 86.690086 total loss 35.625717
epoch 0 step 2400 | bow loss 35.405842 kl loss 0.033317 nll loss 86.952438 total loss 35.405842
epoch 0 step 2700 | bow loss 35.085587 kl loss 0.031792 nll loss 86.765343 total loss 35.085587
valid dataset: bow loss 34.918949 kl loss 0.035117 nll loss 86.970116 total loss 121.924187
```

Figure: training

Tricks Sharing

- we defineLossMSEis mean square error,LossABSis ab-solute errorm based on liner regression. As k varies from 50 to 1000, degree of convergence, radioof initial loss value to end loss value, decreases as shownin Fig.

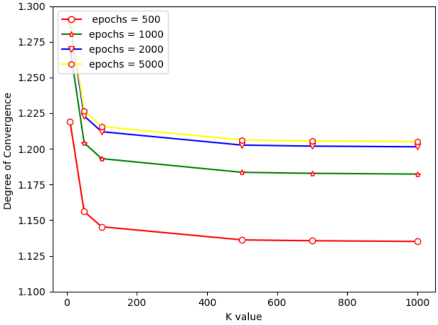

Figure: loss regression

Experiment

- According to the previous introduction, we will divide the following experimental groups

Data Augmentation & Loss Regularization	Data Set					
	D-1	D-2	D-3	D-4	D-5	D-6
Entity Generation	√	√		√	√	√
Knowledge Matching				√	√	
Dialogue Extraction					√	√
Loss Regularization		√	√	√	√	√

Figure: dataset

Results

	F1	BLEU1	BLEU2	DISTINCT1	DISTINCT2
D-1	34.97%	0.34%	0.19%	0.03%	0.10%
D-2	36.28%	0.33%	0.18%	0.04%	0.11%
D-3	30.89%	0.30%	0.15%	0.03%	0.11%
D-4	**41.42%**	**0.37%**	**0.22%**	**0.04%**	**0.11%**
D-5	32.89%	0.33%	0.17%	0.01%	0.04%
D-6	27.45%	0.29%	0.12%	0.01%	0.03%

Figure: Results

Results

- Experiment with D-4, which is processed with Entity Generation, Knowledge Matching and Loss Regularization, scores highest among all experiments, compared with baseline D-1. The five criteria improve 7%, 0.03%, 0.03%, 0.01%, 0.1% respectively.
- Comparing D-2 with D-1, Loss Regularization is proved effective, due to avoiding over-fitting.
- The comparison of D-2 and D-3 reveals that Entity Generation increases the model's performance. Generalization ability of our model is enhanced.
- From D-2 and D-4, There is obvious performance increase result from Knowledge Matching which raises the accuracy of the encoder and decoder.
- The performance of D-5 is worse than D-4, so Dialogue Extraction does not work as we expect. Although increasing the number of samples, it decreases the information of each sample. It is a symbol of under-fitting.

Conclusion

- Multi-round chatbots fall short in changing the conversation subject smoothly. Knowledge selection mechanism are adopted to address this problem
- Our experiments witness that Entity Generation, Knowledge Matching enhance the model's performance.
- NER generation enhances generalization ability of our model, Knowledge selection increases the accuracy of the encoder and decoder
- Loss Regularization is proved effective in different loss value scenario.

Thanks, The End