Akhmadjonov M.K. Kuznetsov D.P.
Nagovitsyn M.S.
oltinsahifa@gmail.com

Moscow Institute of Physics and Technology
DeepPavlov

Controllable Text Generation Using Special Tokens

RELEVANCE AND MOTIVATION

Natural text generation abilities of large neural networks are surprising. But they are not widely adopted in dialogue systems due to the weak predictability of their results in terms of content and different dialogue attributes. Dialogue act, intent, sentiment, and emotion can be examples of dialogue attributes. Despite the problem described above, their coherence in text generation can be leveraged if proper control methods and techniques are used. The proposed method can help control several attributes of the generated dialogue response.

TASK STATEMENT

Given the dialogue context and a dialogue act. There are four types of dialogue acts:

1. Informative
2. Comissive
3. Directive
4. Question

Generate a response corresponding to the dialogue context and the specified dialogue act.

PROPOSED METHOD/ALHORITHM

The main idea of the solution is to transform the input of the model by embedding the knowledge about the specific dialogue attribute, so as to make the model learn the relations between the dialogue attributes and utterances. And use these relations during the response generation. More formally, unlike in traditional language modeling, we make the model approximate this conditional distribution:

$$P(x_t | x_{t-1}, x_{t-2}, \ldots, x_0, ST, CT),$$

where ST means the special token defined for a specific dialogue attribute type. (dialogue act in our case) and CT is the dialogue context.

DESCRIPTION OF THE ALGORITHM DETAILS

We hypothesize that additional conditioning on the dialogue act token makes the model generate a response in accordance with the given dialogue attribute. In order to check this hypothesis, we simply prepend unique tokens to each utterance in the dataset. DailyDialog was chosen as the primary dataset for implementing the method because it has high-quality annotations of dialogue acts and emotions.

After the data was preprocessed, the model was trained on it. (3 epochs on each setting)

DESCRIPTION OF THE ALGORITHM DETAILS

As for the model, we used DialoGPT-small (~130 mln params.) from Microsoft, for its simple architecture and high performance in generating dialogue utterances. However, the method is totally model-agnostic. We decided to use precisely the pretrained version of the model, as the raw DialoGPT's performance was notoriously delicate on such a small dataset.

After initializing the model, we first finetuned it on the preprocessed dataset without any modifications of the architecture. As a result, it demonstrated no controllability at all.

Then we initialized another instance of the model and extended its token space with special dialogue act tokens and resized its embedding space to fit the new vocabulary in size. After finetuning this setup, we got a noticeable improvement in text generation.

EXPERIMENTAL RESULTS

We compared our results with the baseline in this task, which also uses DialoGPT and achieves controllability using projected attention layers (PALs). To measure the controllability we computed the traditional perplexity metric and classification accuracies of generated utterances:

- DialoGPT:
  - Ppl: 4.84
  - Acc: 35.38 %
  - Balanced acc: 42.37 %
- Extended DialoGPT (ours):
  - Ppl: 8.24
  - Acc: 61.94 %
  - Balanced acc: 68.88 %
- DialoGPT with PALs:
  - Ppl: 15.93
  - Acc: 61.58 %
  - Balanced acc: 69.67 %

EXPERIMENTAL RESULTS

The metrics for the raw DialoGPT model finetuned on the raw dataset were almost equal to the ones DialoGPT demonstrated on the preprocessed dataset. Here is an example of the dialogue generated by our model:

```
A: [DA_1] Such a lovely day! Let's go
somewhere?
B: [DA_1] Yes, I'd like to go to the
beach.
A: [DA_3]Let's go to the beach . It's a
beautiful sight.
B: [DA_2]What's the weather like there?
A: [DA_2] It's supposed to be sunny, but
it's supposed to rain . Do you want to
go with me?
B: [DA_4] I'd like to.
```

CONCLUSIONS AND DEVELOPMENT PROSPECTS

As the results show, our approach conduces high controllability while being entirely independent of the architecture of the generative model. Moreover, during calculating the metrics one must also consider the performance of the classifiers as they are not perfect.

However, the method itself is inefficient in the way that makes the whole model train on the transformed data. As for some improvements, we would like to leave the default tokenizer vocabulary and extend only the embedding space of the model, and train only new embeddings by freezing the whole model in the future.

REFERENCES

1. DialoGPT: https://arxiv.org/abs/1911.00536
2. DailyDialog: https://arxiv.org/abs/1710.03957
3. DialoGPT PALs: https://www.dialog-21.ru/media/5761/evseevdaplusetal052.pdf
4. BERT dialogue act classifiers were taken from deeppavlov open-source library: http://docs.deeppavlov.ai/en/master/features/models/bert.html

Implementation of the approach can be found on GitHub:
https://github.com/mumtozee/InnPrac2022/tree/main/baseline

Model weights can be found on HuggingFace:
https://huggingface.co/imumtozee/DA-ctrl-bot