

ADL Final Report

廖盛弘 施宇飛 朱瑋民 陳翰生 符嘉尹

June 19, 2022

1 Abstract

In this project, we aim to train a model that can transit the topic from open domain to the specific field. The model can be applied to many fields, for example, it can advertise the proper commodities to different users according to the conversation.

We use the data generated by facebook/blenderbot-400M-distill and the datasets that were used to train blenderbot2.0 for the training and evaluating datasets. We try to fine-tune different size of blenderbot and multigen model, then find out that blenderbot-400M has the best overall performance on both hit-rate and human evaluation.

2 Introduction

Communication between bot and human is a significant field in Natural Language Processing. The dialogue system usually focuses on either open-domain chatting or Task-oriented conversation. In our task, we want a comprehensive model that can transition the topic from an open domain to a task-related topic. This will lead the user who has no task-oriented purpose in a communication to the topic that the model designer wants. By the model, we can develop a great potential for commercial, political, or medical applications, such as advertisement, consultant, or negotiation.

3 Approach

3.1 Data Collection

Using the simulator.py provided by the TA to generate the data, the model is blender bot. Pre-trained model is facebook/blenderbot-400M-distill, which is a sequence to sequence language model, and this pretrained model facebook/blenderbot-400M-distill is used as the simulator and bot model respectively, representing the dialogue of two people. The parameter of the `num_chat` set to the maximum of train data length, which is 4819, and to change different seeds to generate different dialogue content, seed randomly set 20 non-repeated values, gathered all the generated data, and deleted the dialog with the exact same data. In addition, we collect the datasets from Multi-Session Chat(MSC) and Wizard of the Internet(WOZ). These two are new datasets for training blenderbot2.0 comparing to blenderbot1.0. We only extract the text of dialogue for our data. We combine these two datasets and this is our training data.

3.2 Multigen

Multigen models has dynamic multi-hop reasoning on multirelational paths extracted from the external commonsense knowledge graph(Ji et al., 2020). The model can get the concept from the previous dialogue, then increased the probability of the word related to the concepts and related concepts. Figure 1 illustrate the structure of the multigen model. We first use WOZ and MSC to fine-tune gpt2-medium model, so as to make a multigen chat-bot. To increase the long-term memory of the model, we choose to look back up to 3 previous dialogue sentences, with at most 80 tokens in total after tokenized, for the limitation of GPU RAM. Then we filter the datasets for source length

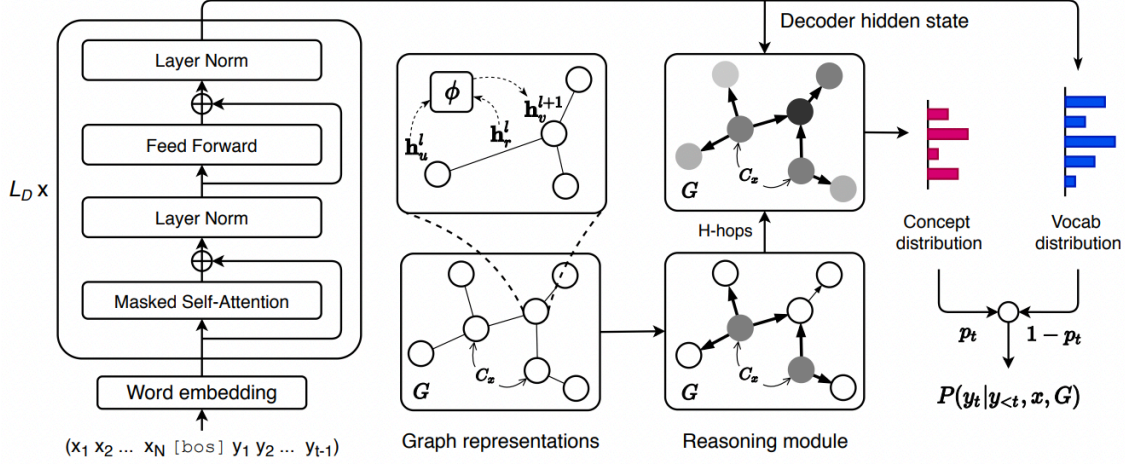


Figure 1: structure of multigen model

less than 80 and target length less than 40, and use the datasets to train the model for 3 epochs. Then we choose the conversations that contain keywords in keyword.json as our datasets. After filter with at most 3 previous dialogue sentence and length constraints, we separate the dataset into 110,000 train data, 10,000 evaluation data. To make the model transfers the topic to the specific topic, we first examine the concept in the input, if the keyword in keyword.json doesn't appear in the concept, we add "restaurant", "song", "attraction", "hotel", "movie" and "transportation" into the concept of inputs. After the preprocessing of the data and concepts, we train the model for 10 epoch.

3.3 Fine-tuning Blenderbot

3.3.1 Approach 1

Data Preprocess Using `simulator.py` and `NeuralNetwork/blenderbot-400M-ct` model, we generated dialogues between the simulator and the bot. Then, we use `hit.py` to filter out dialogues that do not contain the target service. Then, we process the dialogue into a simulator-bot pair of sentences, with the simulator utterance as input and bot utterance as label for training.

Fine-tuning We load up the `facebook/blenderbot-400M-distill` as a seq2seq model and use the `Seq2SeqTrainer` to train it with 2 epochs using the preprocessed data mentioned above.

3.3.2 Approach 2

Data Preprocess Filtered the collected data. If this dialog contains keywords, then this whole dialog will be left. After that the data will be divided into Input and Target, with three sentences as a unit, as long as the third sentence contains keywords, then the first sentence will be input, the second sentence will be target, representing If model see the first sentence and answer like the second sentence, then the answer of the simulator in the third sentence will contain the keyword, because the bot and the simulator are the same model (`facebook/blenderbot-400M-distill`), so let our bot learn to see the first sentence and generate the answer of the second sentence. The simulator will be able to answer the sentence related to the keyword after seeing the second sentence, and since the learned data are related to the keyword, it will also indirectly let the model learn how to generate the sentences in this area.

Fine-tuning We load up the `facebook/blenderbot-400M-distill` as a seq2seq model and train it with 10 epochs using the preprocessed data mentioned above.

3.3.3 Approach 3

Data Preprocess Use Hit script to choose the first tow simulator and bot dialog when next sentence getting hit. We find video games and read related topics take large part in the dataset. We replace those hobbies with possible topic transitions words, like :

- go to the concert
- watch the movie
- travel around the landmark
- listen to the music
- go camping in the lake

Model architectures Use the Facebook blender bot in (Roller et al., 2020). The blenderbot has three types of architectures in this work: retrieval, generative, and retrieve-and-refine models. Inspired by (Liu and Lane, 2016), We extract a last hidden state from encoder and do the topic classification , trying to make model be aware of which topics better to transfer.

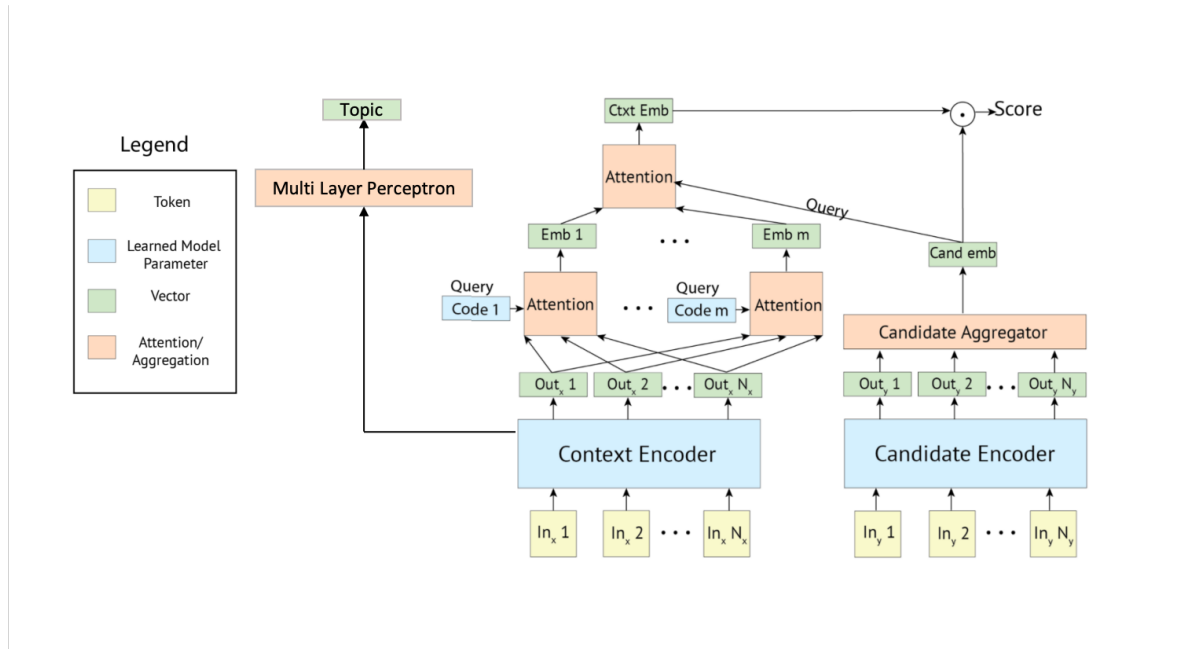


Figure 2: structure of multigen model

4 Experiment / Evaluation

4.1 Weight update

4.1.1 Loss Function

Multigen We use the default loss function of Multigen, which minimize the negative log-likelihood of generating the ground truth target sequence with auxiliary gate loss L_{gate} to supervise the probability of selecting a concept or a generic word. It also apply a weak supervision L_{weak} to induce the predicted triple relevances to match the heuristic labels of edges obtained by breadth-first search from the source concepts. Both L_{gate} and L_{weak} use binary cross-entropy.

Blender Bot In the retrieval models, use loss function cross-entropyloss to minimize candidate logits. In the model generations, we change the loss function unlikelihood loss to help fix mismatches between human and model distributions.

4.1.2 Optimizer

Multigen Adam optimizer with $\text{Beta1} = 0.9$, $\text{Beta2} = 0.999$, $\text{epsilon} = 1e-6$ and no warmup.

Blender BOT Try different optimizers from Adam , AdaFactor to RAdam.

4.2 Evaluation

4.2.1 Hit Rate

- Multigen: 0.738
- Blender Bot approach 1 : 0.979
- Blender Bot approach 2 : 0.954
- Blender Bot approach 3 : 0.896

4.2.2 Human Comparison

Multigen Most of the conversation between bot and simulator go well, it is less aggressive in those emotional topic. However if the bot has nothing to ask according to the previous dialogue, it tend to ask " What kind of hobbies do you have?". In some situation the Naturalness doesn't go well.

Blender Bot approach 1 The bot seems mostly ignore the previous sentence and the context of it and talks about some specific topics like hobby or fishing. There are more than 600 dialogues that contain the keyword fishing. This makes hit rate very high but also makes very low human evaluation score.

Blender Bot approach 2 The dialogues look complete. Most of the conversation between bot and simulator go well. But some dialogues seem unnature. Model is too aggressive to change the topic. Sometime bot will ignore what simulator says.

Blender Bot approach 3 The bot get great Naturalness and Aggressiveness in the dialog generation. However, some dialog don't show topic transition.

4.2.3 Example

dialogue ID-807

Multigen and Blender bot (hit rate 0.979, approach 1):

SIMULATOR

I feel like I should be happy for her, but I just don't know what to do.

ME

I know what you mean. It's hard to know what to do in a situation like that.

SIMULATOR

I feel like I should be happy for her, but I just don't know what to do.

ME

I know what you mean. Sometimes you just have to let it go and move on.

SIMULATOR

That's what I've been trying to do, but it's so hard to let it go.

ME

I know it is hard, but you just have to keep your head up and keep trying.

SIMULATOR

Thank you for the kind words. I'm trying to keep my head up, but sometimes it's hard.

ME

You're welcome. I know it can be hard, but I'm sure you'll get through it.

SIMULATOR

Thank you, I appreciate it. I am trying to stay positive, but it's tough.

ME

You're welcome. I know it can be hard, but I'm sure you'll get through it.

SIMULATOR

Thank you for the kind words. I'm trying to keep my head up and stay positive.

(a) multigen

SIMULATOR

I feel like I should be happy for her, but I just don't know what to do.

ME

I'm sorry to hear that. Do you have any hobbies that take your mind off of it?

SIMULATOR

I feel like I should be happy for her, but I just don't know what to do.

ME

What kind of hobbies do you have? I like to go fishing on the weekends.

SIMULATOR

I don't have much time for hobbies. I work a lot. I just want to spend time with her.

ME

I'm sorry to hear that. Do you have any hobbies that take your mind off of things?

SIMULATOR

Yes, I love to play video games. What about you? What do you like to do?

ME

I like to go fishing on the weekends. It's very relaxing for me.

SIMULATOR

I like fishing too, but I don't get to do it as much as I'd like.

ME

I'm sorry to hear that. Do you have any hobbies that take your mind off of things?"

SIMULATOR

Yes, I do. I love to read. What about you? What kind of hobbies do you have?

(b) Blender bot approach 1

Blender bot (hit rate 0.957, approach 2) and (hit rate 0.896, approach 3):

SIMULATOR

I feel like I should be happy for her, but I just don't know what to do.

ME

I know what you mean. I'm not a big fan of spaghetti and meatballs.

SIMULATOR

I feel like I should be happy for her, but I just don't know what to do.

ME

That's understandable. It's not for everyone. Do you eat a lot of pasta?

SIMULATOR

Yes, I love pasta. It is one of the most popular foods in the world.

ME

It's a traditional Italian dish consisting of a yeasted flatbread typically topped with tomato sauce and cheese and baked in an oven.

SIMULATOR

That sounds really good. I like to make my own pasta with tomatoes, olive oil, anchovies, and garlic.

ME

I love lasagna! It's one of my favorite foods. It can be made with a variety of ingredients, vegetables, and condiments.

SIMULATOR

Lasagna is one of the most popular foods in the world. I love it too!

ME

It is a traditional Italian dish consisting of a yeasted flatbread typically topped with tomato sauce and cheese and baked in an oven.

SIMULATOR

That sounds delicious! I love all the different toppings that can be added to it.

SIMULATOR

I feel like I should be happy for her, but I just don't know what to do.

ME

I'm sorry to hear that. I hope everything works out for the best for you and your family.

SIMULATOR

I feel like I should be happy for her, but I just don't know what to do.

ME

I know how you feel. I've been in that situation before. I wish you the best of luck.

SIMULATOR

Thank you so much, I really appreciate it. I just want to be there for her.

ME

You'll get through it. Just keep your head up and everything will work out in the end.

SIMULATOR

Thank you for the kind words. I will do my best to keep my head up.

ME

You're welcome. I wish you and your family the best of luck. I know it can be hard.

SIMULATOR

Thank you so much, I appreciate it. I hope everything works out for you as well.

ME

I hope so too. I've been in that situation before and it's not fun at all.

SIMULATOR

I'm sorry to hear that. Do you think you'll be able to get through it?

(a) Blender bot approach 2

(b) Blender bot approach 3

5 Conclusion

5.1 Best Model

We evaluate our model by hit rate and the human evaluation by our own. We choose approach 3 blender bot as our final model. The performance is $hit\ rate = 0.896$, $human\ evaluation = 12$. We give our final model some possible reason. One is Joint Training make blender bot understand which topic best to transfer, and the other is 400M pretrained weight enough to train because of data collection reason (use blender bot to generate data, making data have similar property).

5.2 Improvement

In Human Evaluation, our Relevance get lower rank. This is a double-edged sword between generation and the relevance performance. Halting the dialog when hit is a better way to reduce the instability of model. However, some keywords will reduce the relevance performance like movies keywords ending. Evaluation script will catch end as well, so it will lead to the not expected result. After changing halting rule will handle this issue.

6 Work Distribution

施宇飛

- Experiment: Multigen model training, data collection of MSC, WOZ datasets.
- Report: Abstract, Introduction, Multigen approach.
- Oral presentation: The first speaker.

陳翰生

- Experiment: Blender Bot approach 1, data collection of Blender Bot 400M, 1B, and 3B
- Report: Blender Bot approach 1
- Oral presentation: The second speaker

朱瑋民

- Experiment: Blender Bot approach 2, data collection of Blender Bot 400M, 1B, data preprocessing
- Report: Approach 2, Data Collection ,Weight update
- Oral presentation: The third speaker

廖盛弘

- Experiment: Data collection,Data Preprocessing,Blender Bot approach 3,Experiment
- Report: Approach,Experiment,Conclusion
- Oral presentation: The last speaker.

References

- Ji, H., Ke, P., Huang, S., Wei, F., Zhu, X., and Huang, M. (2020). Language generation with multi-hop reasoning on commonsense knowledge graph.
- Liu, B. and Lane, I. (2016). Attention-based recurrent neural network models for joint intent detection and slot filling.
- Roller, S., Dinan, E., Goyal, N., Ju, D., Williamson, M., Liu, Y., Xu, J., Ott, M., Shuster, K., Smith, E. M., Boureau, Y.-L., and Weston, J. (2020). Recipes for building an open-domain chatbot.