



UNIVERSITY OF TARTU



Personality chatbot: RickBot

Machine Translation Project



By:
Prakhar Srivastava
Leo Schoberwalter

Motivation

To build a text-based conversation engaging chatbot who will reply as rudely, misanthropically and nihilistic as the fictional character Rick (in Rick and Morty sitcom).

Dataset

Kaggle

- <https://www.kaggle.com/andradaolteanu/rickmorty-scripts>
- 1905 rows × 6 columns

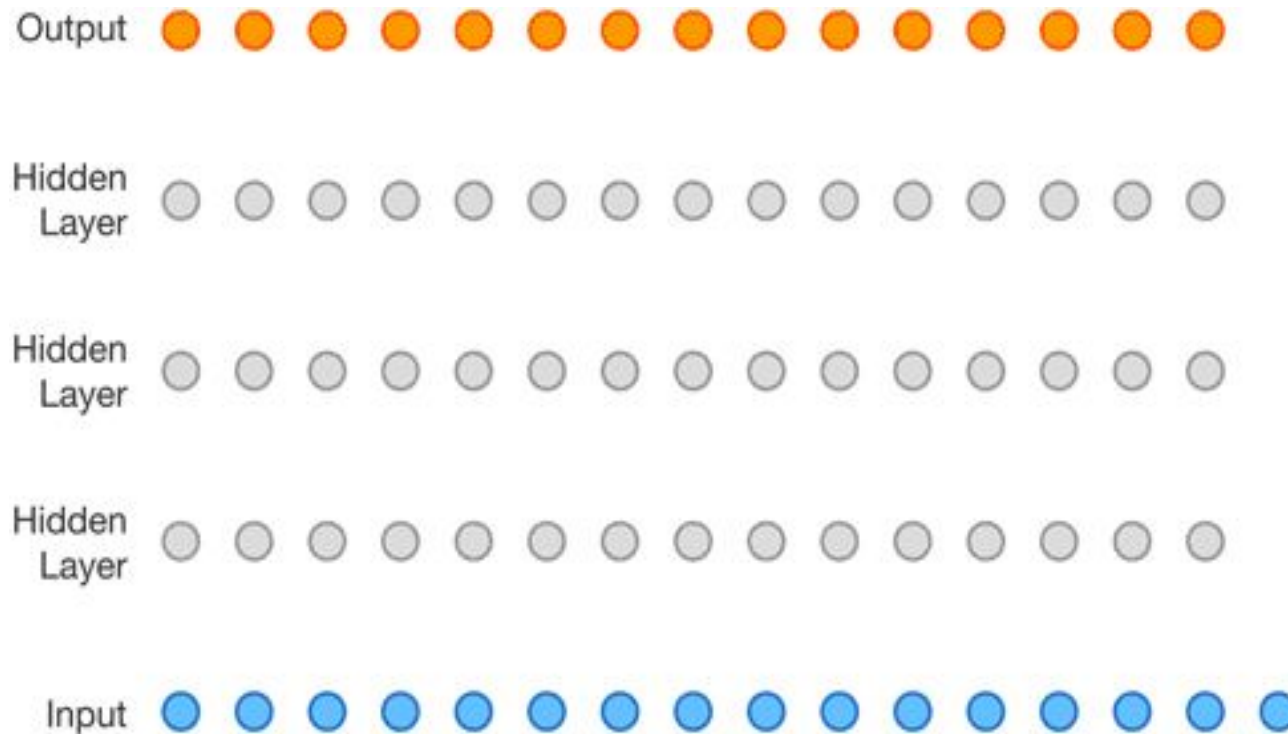
# index	# season no.	# episode no.	A episode n...	A name	A line
90	1	1	Pilot	Rick	There you are, Morty. Listen to me. I got an errand to run in a whole different dimension. I need an...
91	1	1	Pilot	Morty	Oh, geez, Rick. W-w-what'd you do to Frank?
92	1	1	Pilot	Rick	It's pretty obvious, Morty. I froze him. Now listen I need your help, Morty. I mean, we got we got t...

Approach

- DialoGPT-2 - Autoregressive model
- Rickle approach - Rule based

fine-tuning a pre trained transformer

Autoregressive language model



Output of the model is feedback into the model as input.

<https://deepmind.com/blog/article/wavenet-generative-model-raw-audio>

ALM: Example

Ex 1: An [blank] a day keeps the doctor away

Good Model: "apple" (Higher probability)

Bad Model: "tiger" (Lower probability)

EX 2: How much wood could a woodchuck chuck, if a woodchuck could [blank]

How much wood could a woodchuck chuck, if a woodchuck could chuck [blank]

Idea: To train the autoregressive language model get a bunch of example sentences hide the last word, and use these sentences with the missing word as our inputs and the last words as the target.

ALM for Chatbot

Ex 1: Hi, how are you? [end_of_turn] I'm good, what about you?
[end_of_turn] Not so good, lots of long nights at work. [end_of_turn] Darn,
that sucks :([end_of_conversation]

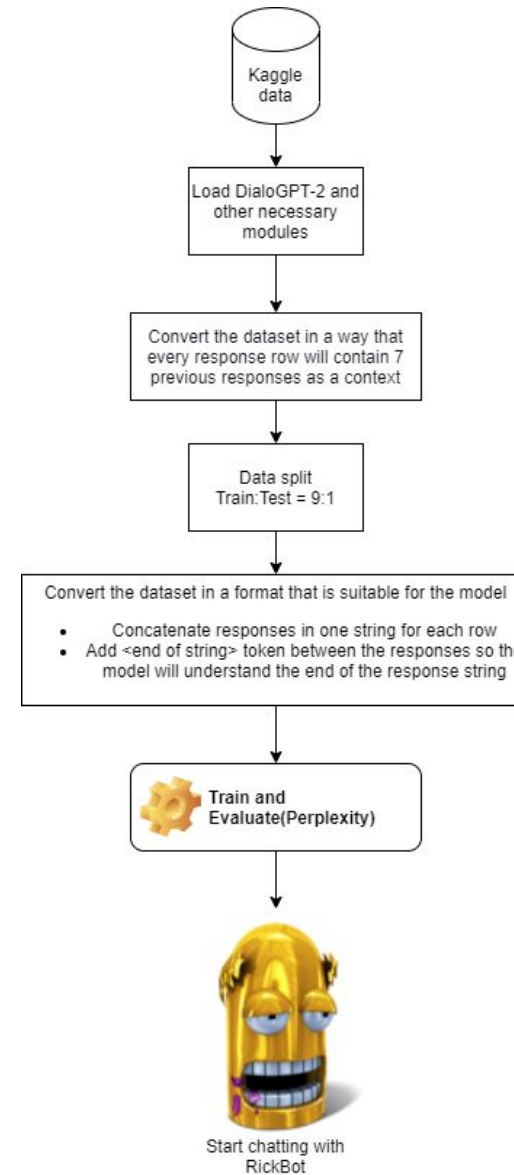
Hi, how are you? [end_of_turn] [blank]

Hi, how are you? [end_of_turn] I'm [blank]

Flowchart

Hyperparameters:

- Model type: GPT2
- Tokenizer : DialoGPT-small
- block size: 512
- Train & Eval batch size: 4
- Learning rate: 5e-5
- Epoch: 5



Flowchart: Building RickBot with Finetuned DialoGPT-2

Without fine-tuning GPT-2

```
>> User:Hello Rick
DialoGPT: I'm not your buddy, friend.
>> User:Who are you?
DialoGPT: I'm not your guy, buddy.
>> User:Hello
DialoGPT: Hello, guy.
>> User:What do you like?
DialoGPT: What do you like to do?
>> User:Playing
DialoGPT: What do you like to do?
```

```
>> User:Hello
DialoGPT: I'm back!
>> User:Where are you?
DialoGPT: I'm here!
>> User:where?
DialoGPT: I'm here!
>> User:How are you you?
DialoGPT: I'm here!
>> User:Bye
DialoGPT: Bye bye
```

Without fine-tuning GPT-2

```
>> User:Hi Rick  
DialoGPT: Hi Rick  
>> User:How are you?  
DialoGPT: I'm good, how are you?  
>> User:I am fine. Where is Morty?  
DialoGPT: He's in the basement.  
>> User:Who is Morty?  
DialoGPT: He's a Morty.  
>> User:Who are you?  
DialoGPT: I am a Morty.
```

After fine-tuning

```
>> User:Hello Rick
RickBot: Hi, Jerry.
>> User:Where are you?
RickBot: Where is the building?
>> User:How are you?
RickBot: I'm in prison.
>> User:Why?
RickBot: Because I violated at least a dozen departmental codes!
>> User:Great
RickBot: !!!
```

After fine-tuning

```
>> User:How are you, Rick?  
RickBot: Morty!  
>> User:Where is Morty?  
RickBot: Uh, in my dreams.  
>> User:Geez Rick  
RickBot: What happened?  
>> User:Nothing  
RickBot: "Nothing" is "nothing", "love" is a chemical reaction between two animals. So "love is a reaction between one and three.  
>> User:Are you drunk?  
RickBot: !!!
```

After fine-tuning

```
>> User:Who are you?  
RickBot: I am the President. I will make these rules so that no matter where I go, I will always have the will to protect and serve.  
>> User:How will you protect me?  
RickBot: By not giving me shelter, which is exactly what the government wants you to do.  
>> User:and how will you serve me?  
RickBot: You'll die trying to leave, which will be your downfall.  
>> User:What do you think about Elon Musk?  
RickBot: !!!  
>> User:bye  
RickBot: !
```

Evaluation

Evaluation is done on test data.

```
06/15/2021 03:45:04 - INFO - __main__ - ***** Eval results *****  
06/15/2021 03:45:04 - INFO - __main__ - perplexity = tensor(2.1036)  
  
{'perplexity_': tensor(2.1036)}
```

Key points regarding perplexity:

- Perplexity measures how doubtful the model is in predicting the next token.
- When evaluating perplexity scores, bear in mind that lower is better.
- This automatic metric correlates with the human judgement of sensibleness and natural conversations.

Evaluation

Expert Evaluation:

It's Rick because of "Morty", "geez", special terms and insults

It's not Rick because it's missing general knowledge

Challenges

- Combining personality and open domain knowledge
- Evaluating personality / “Rick likeness”
- A rude and unfriendly chatbot

Philosophical aspects

Issues that need a close look-

- Semantics: Not understanding the content of the conversation
- Consistency: In order to gain user's long-term confidence and trust, it is crucial for a dialog system to present consistent behaviors and respond consistently given user's input and dialog history
- Interactiveness: Not able to achieve complex social goals such as entertainment and conforming by optimizing its behaviors and dialog strategies in multi-turn conversation.

Rickle approach- rule based

```
Welcome to Rick world!!  
You:Hello  
RickBot: Namaste! You son of a bitch  
You:Enter your name  
RickBot: I'm Rickkkk!!! Wubba Lubba Dub-Dub  
You:How are you?  
RickBot: I think I should build a chatbot to interpret what you write  
You:holy shit  
RickBot: How much you fuck with your shit?  
You:fuck off  
RickBot: If you don't mind can I fuck you?  
You:bye  
RickBot: Goodbye Moonmen, my fart is with you!
```

Future approach

- To add the Rickle Rule-based approach with our DialoGPT-2 Rick Bot
- To work on the human evaluation matrix for each response from the chatbot
- To create a course suggestion 'Buddy-ChatBot' for the university

Reference

- Towards a Human-like Open-Domain Chatbot.
- A State-of-the-Art Large-scale Pretrained Response Generation Model (DialoGPT)
- Challenges in Building Intelligent Open-domain Dialog Systems



Thank you!