

Read me

NLP(2021/9-2021/12) mini project
Cross Over Conversation Generator
<https://git.arts.ac.uk/21006514/NLP-project>

Yiwen
Li

Part 1 **brief introduction**

Inspired by <Trading Tales> by sophinisba on A03, I came across to this idea of using NLP technology to generate crossover conversations, to deepen my understanding toward crossover by the randomness of language models.

The idea is to train numbers of language models using lines from several different fiction movies, and eventually make each of them generating prefix for on another, so to create a 'back and forth' conversation.

Part 2 **Bilbo Hagrid Gandalf and Dumbledore**

Over 1000 lines were compiled from eleven films, sorted by character. In these four characters, Gandalf and Dumbledore were trained purely on movie dialogues, while Bilbo and Hagrid were trained on dialogues with extra persona, to give them more personality and stronger opinions.

Part 3 **Human and computer**

Constructing 'back and forth' conversation wasn't easy, because using generated text as prefix is risky for those text has it's own randomness. I found a way to make the conversation more consistent by making all the sentence end with comma(,), so as a result I have to improv a little to add punctuation mark at the end of every line.

Apart from training the cross over conversation generator, i also hired some human writers to write some dialogues with the exact same theme and structure and characters. And I extracted some sentence form those human creations I gathered as initial prefix for the generator, all the out comes are in the out come folder

Part 4 **Notebook and checkpoint**

The note book only works on colab, and before generating anything, all the checkpoint files need to be uploaded into a 'checkpoint' folder under the 'content' folder. There are four checkpoint folders for four characters, because those files are too big to be uploaded to github so please get access to those files through the link I put in the txt file.