Reverse Jeopardy!

Shaemus Melvin, Valerie Kim

Motivation

- 2011, Watson defeats GOAT Ken Jennings
  - Reads clue in the form of an answer
  - Responds with the response in the form of a question
- Watson is a landmark achievement in Question Answering. However, our task is more related to Question Generation
  - *“Generating questions that can be answered by the input content*.” [1]

Jeopardy Questions are Clues

Clue: “Objects that pass closer to the sun than Mercury have been named for this mythological figure.”

Category: "THE SOLAR SYSTEM"

Contestant Response: "Who is Icarus?"

—

Clue: “These parts of a peach tree are glossy green, pointed & lance shaped.”

Category: “PARTS OF PEACH”

Contestant Response: “What are leaves?”

Methodology

Data processing

- Data originated on J!Archive, which has every question since 1985 [2]
- Data retrieved from Reddit in CSV form [3]
- Cleaning performed by removing audio, video, clues
  - Denotated through “(audio clue)”, “(video clue)”, and link to photo (“a href=”)
- Size of dataset reduced.
  - QA Pairs: ~220,000 → ~205,000

Model Mismatch

IMPORTANT: Encoder, Decoder, Encoder-Decoder Transformers are different!

Encoder: BERT

- Requires ‘Context’ feature to precisely identify answer of a question
- Pretraining objective: identify the answer inside a context

Decoder: DialoGPT

- Generate Text
- pretraining objective: predict the next token

Encoder-Decoder: T5

- Task based text-to-text
- pretraining objective: transfer learning

Implementation

Chatbot/DialoGPT

- Training done in chatbot fashion in clue/response pairs - no other context given
  - Specifically, conversation history is not given as it's not relevant and can actually muddy the responses
- Prompt given is response - input to chatbot is "Ireland", desired response "This European island nation has a green, white, and orange flag" (or something similar)

T5 Question Generation (in progress)

Training done with task/response pairs.

- Task: "jeopardy: {example['answer']} in {example['category']} for {example['value']}"

Evaluation [1] to be done with:

- Well formed Jeopardy Clue structure
- Semantic Correctness
- Answer Existence

Results

- Getting kicked off of Google Colab :(
- Questions grammatical and phrased correctly, but incorrect and sometimes nonsensical
- Topics seem to come up repeatedly (false) etymology is a common clue topic, with the model specifically saying many things are from the Old English for “small”

```
Response: Canada
Question: The first U.S. flag was made by the U., not by this country
Response: Blue
Question: The name of this color is from the Latin for "colorful"
Response: Shrek
Question: This animated film about a princess & a prince is based on a book by James Woods
Response: Mozart
Question: In 1871 this composer's "Die Another Day" was performed at the Berlin concert hall
Response: Silence of the Lambs
Question: In this film, the title character is the daughter of a famous man
```

Discussion/Future Work

- Continue training the model(s)
  - Move training to local devices rather than Colab to get past GPU quota limits
  - Incorporate categories in addition to question/answer pairs
- Compare results of chatbot and QA models
- Test question generation
  - Humans, potentially trivia solving models like Watson

References

[1] https://devopedia.org/question-generation

[2] J!Archive: https://j-archive.com/

[3] Reddit source: https://www.reddit.com/r/datasets/comments/1uyd0t/200000_jeopardy_questions_in_a_json_file/

[4] https://github.com/patil-suraj/question_generation