

An Artificial Intelligence Conversational Chatbot Developed for Non-Native English Speakers

Pengcheng Li (John Li)

Henry M. Gunn High School, California, USA

Abstract. As common knowledge would dictate, the best way to learn a language is to talk with native speakers of said language. However, it may not be the easiest for non-fluent, non-native speakers of languages such as English to find native English speakers to converse with. This paper proposes a conversational chatbot that would help non-native speakers to converse with artificial intelligence with equal, if not greater, fluency compared with a native English speaker.

Keywords: Conversational artificial intelligence, Educational chatbot, Automatic speech recognition, Educational data mining.

1. Introduction

With the increase in internationalization in the past century, traditional languages are playing a growing role in business. English, being the current lingua franca, is the top language used for worldwide communications, earning it great importance and acceptance. This acceptance has resulted in a sort of “boom” in learning English.

Although many pupils around the world are taught how to read and write English fluently, many pupils have few opportunities to practice their oral English, specifically for their pronunciation. In order to solve the above problems, this paper proposes a conversational chatbot that can converse with people who may not have access to talk with a native English speaker, which can effectively improve their English skills.

The chatbot will start when the user initiates a conversation with a greeting. Then the chatbot and the non-native speaker will take turns discussing. If the speaker is not able to understand, the speaker may ask the chatbot to repeat the sentence, and if the speaker still cannot comprehend the sentence, the speaker can ask for the sentence in their own native tongue. If the speaker pronounces a word wrong, the chatbot will take note of the error and report it at the end of the session. If the chatbot is not able to comprehend what the speaker means, it may ask for the speaker to rephrase or repeat the sentence. If after multiple rephrasing's, the meaning of the sentence is still not understood by the chatbot, the conversation ends, and the artificial intelligence gives feedback about the improvements the user made and some minor mistakes the user may want to focus on in the next session.

The data set used to power this model will consist of conversations in daily life taken from a variety of sources such as television show scripts, and the normal day-to-day conversations et al.

2. Related Word

Note that automatic speech recognition is a way of converting a stream of audio and converting it into text and doing something with the text. In this case, transcribing the text and analyzing how to respond should be a rational solution. However, it will be challenging for computers to handle these analyses because computers just only own the basic ability such as doing math equations. In order to solve this problem, computers must first take the stream of audio and then transcribe it into a series of vectors. Next, the computer need essentially converts these series into tokens by looking at Phonemes, Graphemes, etc.—through a process called “tokenization”.

ASR can be used to transcribe what the user has said into text. This text can then be fed into a transformer such as BERT or GPT-2. The transformer would then generate an appropriate response and put it through speech synthesis. Although ASR is not strictly required to “converse” with a transformer it defeats the purpose of practicing oral English. Although there are no specific utterances

the ASR must listen for, due to the nature of the application, the ASR must be able to listen to and transcribe almost all everyday terms.

3. Models

3.1 Hidden Markov Model (Automatic Speech Recognition)

In this paper, Witt and Young put out a method to use Hidden Markov Models and detect mispronunciations in non-native speech. The pair discovered a formula to approximate a “score of correctness or confidence for each phone of a desired transcription” [2]. By using the formula, the chatbot could measure how familiar the speaker is with the pronunciation of their phones. By stringing the results of multiple phones, the chatbot can deduce how accurate a speaker was with a certain word. If the speaker made enough pronunciation mistakes greater than a certain

Threshold, then the system should silently take a note of the mistake and bring it up at the end of the session. Furthermore, Witt and Young also introduce an extended version of their formula. The pair suggest that the extended version of the equation could be used to determine an overall assessment of the speaker’s strengths and weaknesses [2]. We will discuss utilizing this second equation later on in the paper.

3.2 Text Transformer (Conversational Chat AI)

To implement this part of the communication chatbot, we propose using a modern text transformer such as GPT-2 motivated by research finished in [3]. In their paper, Thomas Wolf et al. puts forward a new approach in the field. By exploiting Transfer Learning and merging it with a high capacity transformer, the team was able to create a conversational agent which boasts improvements in multiple categories ranging from 20% absolute improvement to 46% [3]. By utilizing their implementation of a conversational agent in the application, the app can use the transcribed text obtained from the model above and generate an appropriate response to continue the conversation.

4. Data and Criteria

4.1 Educational Data Mining

In this application, many things can be tracked. However, this paper will focus on two specific cases. In the first case, we can track the time it takes for the speaker to respond to the chatbot. The gap in time between how long it takes between the chatbot’s last sentence and the speaker’s response can be used as a metric to determine their oral comprehension with a long time suggesting that the response was challenging for the reader to comprehend. Another reasonable case would be the gap in time between each word the user utters. These gaps could be used to track the user’s memory access speed. For example, if the gaps shrink over a set period of learning with the application, the application can assume that the reader has gotten faster at pulling words out of their memory to string a sentence.

4.2 Criteria for Evaluating Applications

4.2.1 Student Progress

For the evaluation aspect of this app, the application can use the second formula described in the Hidden Markov Model section, which can calculate an approximate overall assessment of the speaker’s abilities. Using this, all the chatbot needs to do is just store the evaluations over time to see the speaker’s progress over a period of time. The application can also take advantage of the data mined from the section above regarding response and spacer times as an indicator of fluency.

4.2.2 Validity of Responses

Another evaluation is to look at how grammatically correct and sensical the response is from the transformer. This could be done by allowing users to use their judgment to determine if a response is nonsensical and report it as needed. To improve upon this, the application could assign weights to each report based on the score discussed above. By assigning weighting each report based on said score, the application can reduce the total amount of false positives from things such as a poor English speaker falsely marking conversations they do not understand even though they likely are valid in a fluent manner.

5. Experiments

There are many things that the application can be embedded into the experiment. However, this paper will only cover two of the most impactful experiments, misheard input (audio input incorrectly transcribed) and inappropriate responses (responses from the chatbot that are nonsensical).

5.1 Misheard Input

Because automatic speech recognition is not perfect, every time the application does not fully understand the user's input, the application can prompt the speaker to repeat what they have said. If what the user repeated matches the previous transcription, the application then marks their mistake as due to their oral English abilities and not hardware issues. Likewise, if the user's input was different the second time and the application can comprehend their audio, the application will just assume it was an input error and continue.

5.2 Inappropriate Responses

Due to the nature of the application, there will always be cases where the chatbot may generate a response that is essentially nonsense. These inappropriate responses may hinder the development of the speaker's oral English and give false negatives. For example, if the application outputs a genuine incomprehensible response, the application may falsely assume the user is challenged when comprehending the response whereas the last sentence and response should not even be taken into account. The application can test for this by adding a button or some input that the user may activate upon suspecting a genuine incomprehensible response.

Once the user has pressed the button, the application will discard any reports about the last sentence and response. Furthermore, as stated in the section above about the validity of a response from the transformer, the user's flagged sentence and context should be sent to a web server for further processing and transformer improvement.

6. Conclusion

This paper proposes a conversational chatbot to help non-native speakers communicate more fluently and skillfully. The proposed system is composed of hidden Markov model and the text transformer model, which have been proven to be effective from our experiment results. We have evaluated our proposed model from two aspects, the student progress and the validity of responses are all obtained promising performance. In the future, we will consider embedding our model into the human-like robot and try to reduce the computational cost of our model, which can provide more real and comfortable using experiments.

References

- [1] Budzianowski, Paweł, and Ivan Vulić. "Hello, It's GPT-2 -- How Can I Help You? Towards the Use of Pretrained Language Models for Task-Oriented Dialogue Systems." ArXiv:1907.05774 [Cs], Aug. 2019. arXiv.org, <http://arxiv.org/abs/1907.05774>. 1-6.
- [2] Witt, S., and Young, S. (1997). Language learning based on non-native speech recognition.
- [3] In Proceedings of Eurospeech. 1-4.
- [4] Wolf, T., Sanh, V., Chaumond, J., and Delangue, C. (2019). Transfertransfo: A transfer learning approach for neural network based conversational agents. ArXiv:1901.08149 [Cs]. <http://arxiv.org/abs/1901.08149>. 1-4.