Veritas AI - AMES (Always My Emotional Support) Chatbot

Avery Ainsworth

Received September 02, 2023
Accepted November 05, 2023
Electronic access November 15, 2023

This project is designed to help catch the warning signs of teens and young adults who are battling depression or other mental health issues. The AMES (Always My Emotional Support) chatbot engages in conversations with high school and college students. Through the use of algorithms, the chatbot seeks to identify individuals struggling with mental health issues and provide targeted responses to attempt to get needed help directly to the individual. In a conversation, the chatbot asks questions and solicits responses to engage with the individual. With the information that the chatbot receives, the responses can be used to identify different emotions and the potential for those emotions being acted upon. After analysis and identifying a person's mental state, the chatbot determines if the individual should be referred to a mental health specialist. This research process centers on the construction of a deep learning model designed for sentence-level sentiment prediction. The models use a three layer process, which includes a bidirectional LSTM layer and a dense layer. These layers are compatible, and the training accuracy for the model after ten epochs is 75%. Although the models work well together and the results are promising, there is room for improvement based on the validation accuracy currently proving results around 30%. The second part of the chatbot that was built incorporated follow-up questions or words of advice based on the responses from the user. If the users' responses had a negative sentiment the chatbot reacted with additional questions and attempted to further engage the user with some advice. The goal of the chatbot is to identify people who are struggling with mental health issues before they act on their negative views either through self-harm or externally toward other individuals. As the chatbot's models improve, you are likely to see better advice and more positive outcomes. The findings from this project should inspire others to build models that can help combat the mental health crisis in this country

Introduction

Today in the United States, the leading cause of death for adolescents and young adults is firearms[1]. The U.S. is at a breaking point where immediate action and prevention needs to take place to prevent school shootings and self-harm. The recent developments in AI can be used to help find the warning signs that are commonly present with individuals prior to self-harm and mass casualty events. Harnessing AI in the right way can help reduce the frequency of these terrible events. Young adults in the U.S., age range 15-24, are twenty three times more likely to be killed by a firearm compared to other high-income countries (Katsiyannis, et al)1 School shootings and acts of self-harm are on the rise in teens due to a multitude of reasons; including but not limited to, more accessibility of firearms, lack of available mental health professions across the U.S., the proliferation of body shaming and bullying on social media and continued fall out from the impacts of COVID-19. The reduction in gun laws, which has resulted in easier access to guns, and a steep rise in mental health issues play a key role in these types of events. Additionally, the isolation and lack of social interaction during the Covid-19 pandemic had a severely negative impact on the mental health of young adults. This isolation and lack of support has led to an increased number of high school and college students developing mental health issues like anxiety and depression. (Jones, et al)[2] Mental health issues, when left untreated, are a key contributor to school shootings and self-harm. In many school shootings, the students that attacked the school were either attending students or past students of the school. Additionally, these attacks were the result of a combination of the feeling of isolation, bullying, paranoia and depression. In some cases, the individual experienced great losses in their lives and childhood traumas such as, physical or emotional abuse or unstable families with absent parental figures. In the past year, 37.1% of students have suffered from poor mental health (Jones, et al)[3] and based on research from the American Foundation for Suicide Prevention most young adults who are suicidal show some type of warning signs to their peers or family members before attempting to end their lives. These warning signs are present in conversations with adults, peers, but more commonly they are present online in chats or posts. These communications are distinct because they commonly express feelings of hopelessness, having no reason to live, feeling trapped, or unbearable pain (American Foundation for Suicide Prevention)[4]. The proliferation of these online communications means that with the right technology, many lives can be saved by identifying the warning signs and attempting

to surround the individual with resources to prevent a negative outcome such as self-harm or an attempted mass casualty event. By using new AI language learning technology, it is now possible to detect warning signs from individuals' online posts and chats. The right use of generative AI and language learning will be able to save lives and get support for individuals that are in a crisis state. The AMES (Always My Emotional Support) project is designed to help catch warning signs of teens and young adults who are battling depression or other mental health issues and attempt to help get them the right resources in a time of crisis. If effective, this will help prevent acts of self-harm or mass casualty events. The "AMES" chatbot will engage in conversations with high school and college aged students, by identifying key words and phrases to intervene, attempt to use language that can mitigate and de-escalate a potential crisis moment and guide them to resources for additional support as they struggle with mental health issues. In the current beta version of the chatbot, it used a large set of almost 40,000 tweets to test the viability of the solution. As the chatbot continues to evolve, it can easily be adapted to plug directly into various social media platforms, such as Twitter, through API connections. Through the API connections it will identify specific phrases and words to engage in a conversation with individuals and ask them questions. As the chatbot identifies specific emotions and potential risk factors based on the conversation, the chatbot will identify the specific risk, such as suicidal thoughts or anger towards others and provide the appropriate resources to contact. Additionally, as the chatbot and AI technology continues to develop, it will be possible very quickly to automatically connect the individual to the resources. Although the AMES chatbot won't prevent all events related to self-harm and mass casualty events, the continued development and implementation of AMES or a similar product into social media can have an impact in the short run and help develop a culture of support, instead of isolation. This paper is focused on how a beta test of the chatbot idea was created and the results of the beta test. The project goal is to help detect students having an acute mental health crisis.

Methodology

Dataset and Data Processing

The dataset used was based on a sample of 39,827 random tweets from various individuals during 2017. Of the tweets used, there were no duplicate values to ensure a broader set of individuals and material to include in the analysis. Tweets are a good source of data for beta testing since they are publicly available, have large volumes of digestible data and are conversational in nature. All of these components make Twitter, or any other social media that takes peoples' comments, ideal for integration with the AMES chatbot. Given all the public data, it is easy to adapt the sentiment from the public tweets into conversations between the users and the chatbot. The chatbot is only using the data from the tweets to identify the patterns in speech, so then when conversation between the user and chatbot has concluded, the chatbot can compare the response to the trends shown in the dataset. In no cases are any personal identifiable information used or stored in the chatbot. Twitter datasets are suitable for helping to identify mental health issues because of the proliferation in sharing via the social media platform and the short form content. Additionally, the broad demographic base of Twitter will allow the language learning model to learn new context, slang and communication among various backgrounds, age groups and geographies. The chatbot AI and language learning is able to classify the dataset into thirteen different specific emotions. This level of specificity is enough to help identify and detect at risk teens and young adults. The emotions that the dataset classifies are: neutral, sad, enthusiastic, worrisome, loving, funny, hateful, happy, bored, relieved and surprised. If you were to only focus on positive and negative classifications, it would not be able to determine simple differences such as sadness or anger. By using a broader set of emotions the AMES chatbot can determine more specific risks associated with an individual and attempt to guide a more positive outcome with the correct resources. When the chatbot receives the tweets, it breaks the information into three columns which are the tweet ID, sentiment and content columns. However, given privacy concerns, the chatbot will only be using the sentiment and content columns at this time.

AI Model

The AMES neural network is made up of three main layers. The first layer is an embedding layer. The embedding layer's function is to help process the data more efficiently. The layer takes the integer-encoded vocabulary and looks up the embedding vector for each word-index. The next layer in the neural network is a bidirectional long short-term memory networks ("LSTM") layer. LSTM layers allow the model to remember important context and are an important part of neural networks. LSTM is particularly useful for AMES sentiment analysis since the model has to be able to recall prior points in the conversation in an attempt to predict the sentiment of the individual[5,6]. This layer takes the information from both directions; from right to left and left to right allowing for a more meaningful output since the LSTM layers take information from both directions. The last layer in the neural network is the dense layer. The dense layer plays a key role since it receives the input from all the prior two layers. This model has done a good job at having a high training accuracy, but there is still room for improvement with the validation accuracy. The next part of the project was to incorporate the model into a

chatbot. chatbots are automated conversation systems that use natural language processing (NLP) to respond appropriately to users[7,8]. After completion of the model, it was incorporated into a chatbot. The chatbot accessed the model through a txt file of questions and chose three at random. With each subsequent chatbot interaction the chatbot would remove the prior question asked so there was not one question that was asked twice, creating the opportunity for better language learning as the chatbot evolves. After the three questions were picked from the file, the chatbot began asking the user the first question. With each response the chatbot would save the user's input and use the model.predict () function on the user response. The function returned a list of percentages that represented the likelihood of the user's response being one of the thirteen emotions. With each response the highest percentage was taken from the list and converted into the sentiment it represented and depending on what the sentiment was the chatbot would respond to the user in a sentiment that represented an openness for communication and dialogue. If the user typed "bye" the chatbot would not ask another question, but if the user typed anything outside of the phrase "bye", the chatbot would ask the next question. This process was repeated for three different questions. All of the user's responses were saved and the sentiments were predicted by the model.

Results

This relatively simple process was able to use three different layers to predict the sentiment of a sentence from various individuals. The three layers used are an Embedding layer, a Bidirectional LSTM layer and a Dense layer. These layers have proven to work well together, and the training accuracy for the model after only ten epochs is 75%. Although the training accuracy shows promise, there needs to be continued work around the validation accuracy since it is only around 30% at this time. The low validation accuracy was caused by the number of sentiments to classify (compared to the dataset) being relatively high. This caused a lower accuracy score than most models that only classify responses as positive and negative. The advantage of this chatbot will be, as the language learning models evolve, you should expect to see accuracy scores improve and get higher results in identifying at-risk individuals. The chatbot component of the project has been completed. The development team began with the development of questions, that when asked, can provide insight on how the person providing the response is feeling. This database of questions was then put into a txt file and three of them were selected at random and asked to the user. The model.predict () function is then applied to the response to determine the sentiment of the response. This function returns a list of percentages, and the models take the highest percentage from the list given the generative AI and large language learning models ability to learn from the data set. This learning will result in a higher than normal probability of being able to predict sentiment. The models take the largest value and convert that into the sentiment it represents. After the chatbot knows what the sentiment of the response was, the chatbot then replies with an appropriate response to the user. The chatbot has been tested in beta form, by taking almost 40,000 tweets from randomly selected individuals in an attempt to get a broader, robust set of tweets to ensure proper testing of the chatbot. The chatbot has been successful at identifying the sentiment behind the response, except in responses where there is an adverb such as "not" prior to a positive adjective. As the generated AI and large language model learn and interpret words together, instead of independently, the ability to correctly read the use of the adverb will be resolved.

Discussion

The AMES model has been completed, but the results are not as good as expected given the early stage beta development. As the development process evolved, the model went from a regular LSTM layer to a Bidirectional LSTM. This change was expected to result in a significant improvement in the training accuracy and validation accuracy. With the change to bidirectional LSTM the training accuracy went to 98-99%. Unfortunately, the validation accuracy only went from 25% to 33%. Further testing and refinement will be able to improve the validation accuracy. Although the validation accuracy is low under the current configuration, the AMES chatbot does well at the intended goal of predicting the sentiment of an individual. There have been multiple studies that have explored using models to predict sentiments and chatbots have proven to be good at supporting this process. The majority of work that has been done on sentiment has solely focused on the binary positive or negative sentiment[9,10]. This rudimentary work does not support the real need in the market to help identify the risk associated with an individual in crisis. This project has been focused on broadening the scope of sentiment analysis and seeking a more comprehensive result to address the growing challenges with teens and young adults. Based on the results of the data set, the emotions with the positive attributes received the highest precision and were most likely to receive accuracy rates near 100% continuously. However, the negative sentiment that was driven by a negative adverb, with a positive adjective created the lowest accuracy rates under 20%, driving the overall accuracy rates down in the dataset testing. Interestingly, a negative sentiment scored well in the dataset with accuracy rates over 80%. The ability to accurately predict the negative sentiments supports the viability of this chatbot and solution given the prevalence of negative sentiments from people in an acute crisis state. The overall F1 score from the current work come in at 69.45% further sup-

porting that the preliminary work, if expanded on and further developed is a potential opportunity to mitigate individual crisis events.

Conclusion

The findings incorporated in this report are very important because they represent a potential scalable resource to address a growing crisis in the U.S. and help mitigate the risk of this type of crisis growing around the world. The AMES chatbot could be very useful at helping to identify people who are struggling with an acute mental health issue in real time and attempt to provide resources to reduce the risk of a negative outcome. The findings from this project should also inspire other people who build sentiment models to consider picking up where AMES is and furthering the development of a fully deployable chatbot that supports mental health. The AMES chatbot was able to classify emotions into thirteen different categories such as: happiness, sadness, worry, emptiness, fun, anger, hate, neutral, enthusiasm, love, boredom, relief and surprise. This list is already more specific than most sentiment analysis done today, but this is only the starting point. This should be used to build more data and classify more responses as even more specific emotions to achieve better results.

Acknowledgements

Thank you for the guidance of Ihita Mandal, mentor from Carnegie Mellon University in the development of this research paper.

References

1 A. Katsiyannis, L. Rapa, D. Whitford and S. Scott, *An examination of US school mass shootings, 2017-2022: Findings and implications*, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9388351/, Advances in neurodevelopmental disorders.

2 R. Chatterjee, *NPR*.

3 S. E. Jones, K. A. PhD1, M. H. PhD1, S. D. MS1, V. D. L. PhD2, J. T. PhD2, C. L. MPA1 and P. J. MPA1, *January–June*.

4 *American Foundation for Suicide Prevention*.

5 X., *Medium*.

6 F. Pascual, *Getting started with sentiment analysis using Python. Hugging Face – The AI community building the future*, https://huggingface.co/blog/sentiment-analysis-python, n.d.).

7 Author, *Pykit*.

8 S. Subramanian, S. Breviu, C. Soshnikov, D. Bornstein and A., *Learn the Basics - PyTorch Tutorials 2.0.1+cu117 documentation*.

9 *Pipelines - hugging face*, https://huggingface.co/docs/transformers/main_classes/pipelines.

10 D. Kuria, *How to create a sentiment analysis model from scratch*, https://www.makeuseof.com/create-sentiment-analysis-model/.

Extra Info on Figures, Tables and Equations

This is a link to my GitHub repository which holds the source code for this project: https://github.com/averyainsworth/ames-chatbot/blob/69209dd88a8f5b28e3953f9f5b1fc152c6f46c60/sentiment$_a$*nalysis*.*ipynb*