Topic: Modern Military Conflicts

The chosen topic for my chatbot is modern military conflicts. This means that the user will be able to ask the bot high level questions about a range of military conflicts such as world war 2 or the Vietnam war. The user would also be able to ask specific questions and get the chatbot to identify different pieces of military equipment.

Chatbot Modules:

Rule based

For the rule-based component, the bot will use the AIML language to provide an XML file that aids the bot in choosing an appropriate response to the users given input. This has limits however as the bot will only be able to answer preprogramed questions that are already within the XML file. An AIML entry consists of two elements, a pattern and a template. The pattern is the input string that the bot will identify, while the template is the bots predefined response based on the pattern input. An example of this is <category><pattern> HELLO</pattern><template>Hello, how are you</template></category>. This example would mean that the bot would respond to the user entering “Hello” with “hello, how are you”.

Similarity Based

For the similarity-based component, the bot will use a range of techniques to determine the appropriate AIML response to a given input. This has increased range compared to a strictly rule based AIML approach as the similarity-based approach can enable the bot to respond to multiple variations of a question, regardless of phrasing and still be identified as the same core question.

The bot will take a series of question and answer pairs, convert them to a bag of words model which a dictionary of each word within a document is and its frequency turned into a vector once stop words have been filtered out. The user’s input will also be processed in this way. Both bag of word models will then be used to calculate TF-IDF values for each phrase. A TF-IDF value is a representation of the importance of a word within the document. It is calculated by taking the dot product of the TF value (the number of occurrences of a word / the total word count of the document) and IDF value (the log of the number of samples divided by the number of samples that contain the specific word).

Once the TF IDF values have been calculated, in order to determine the similarity of a given input phrase compared to the phrases stored within the question answer pair, a Cosine similarity value is calculated. The cosine similarity value is a representation of the cosine of the angle between two given vectors. This means that the closer a cosine similarity value is to 1, the more similar the two vectors are. A cosine value is calculated by using the formula:

Cos θ = the dot product of the two vectors divided by the magnitude of the two vectors

The vectors used for the calculation of the cosine similarity are the vectors calculated for the TF-IDF values.

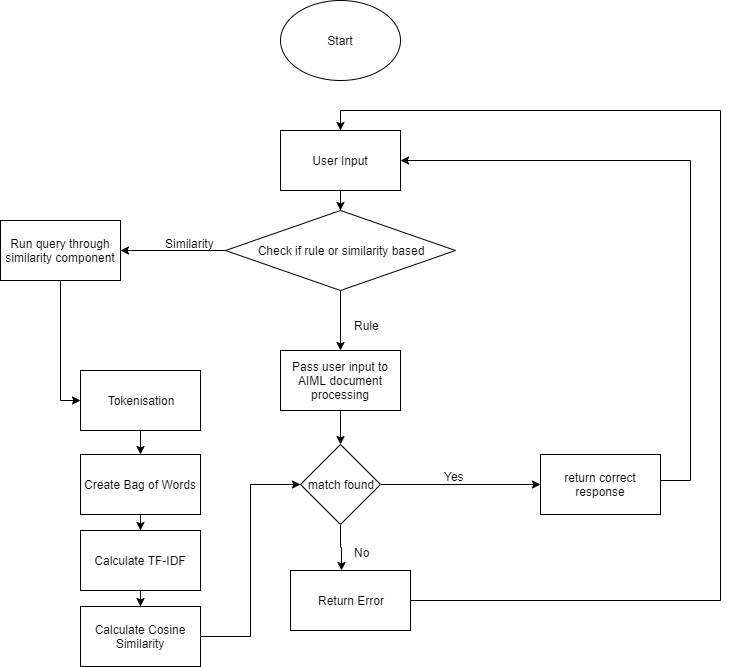
Image Classification

The user feeds into the bot a file path containing the image they wish to classify. The bot then will attempt to classify the image and identify its contents within the image using a convolutional neural network along with a dataset of images to train the CNN.

Toy World Reasoning System

The toy world reasoning system component will build upon the NTLK portions of the chatbot to translate natural language into corresponding first-order logic and then perform reasoning based on the results. This will then be triggered using keywords and key phrases and then add the information from the user input to a database.

Sequence NN and Reinforcement Learning

 The sequence to sequence network extension component will be used to provide a fallback option for when the other components fail to detect any pattern in the input in order to always provide a response. This will be achieved by using a deep neural network. This means that a neural network with at least two hidden layers will be used.