HoloBOT AI Integration

Author: Iulia Groza
Team members:
Andrei Dragan
Raluca Rusu
Ionut Dutescu
Iulia Groza

Background

The project will be focused around providing at-home therapy through an interactive chatbot that will extract **5W1H (who, did what, when, where, why, how)** information from conversations. The tool will have two main functionalities:

1. **Providing users with therapy** - the chatbot will engage in conversations with them, gather context & background, and identify potential solutions to the problem. **Note:** the chatbot is a conversation tool for improving the mental health of the user; making a diagnosis for problems of psychological nature is out of the scope of this project.
2. **Extracting 5W1H from the obtained context** - the information gathered from the whole chat will be centralised and then 5W1H will be extracted from the context. This concise information will offer a helping hand for psychologists when working with their clients to identify the main problems.

Design

From an implementation point of view, the project can be split in two main tasks:

1. Building the Conversational AI Model
2. 5W1H Extraction from a conversation

Building the Conversational AI Model

There are multiple solutions that can be explored throughout the development of the project, depending on the given requirements and delivery deadlines:

[Discarded] Option A: Building a Therapy Chatbot using NLP

The goal is to create a virtual conversational agent that can offer emotional guidance, provide helpful insights, and alleviate some of the challenges faced by those seeking mental health support.

The **dataset** will encompass various conversation types, including basic exchanges, frequently asked questions about mental health, classical therapy discussions, and general advice given to individuals facing anxiety and depression.

To train the model effectively, the dataset incorporates the concept of **"intents"**. Each intent represents the underlying purpose behind a user's message. For example, if a user expresses sadness, the associated intent would be "sad". Each intent is accompanied by a set of **patterns**, which are example messages aligning with the specific intent, as well as corresponding responses that the chatbot should generate based on that intent. Through defining multiple intents and their respective patterns and responses, the model learns to identify user intents and generate relevant and compassionate replies.

NLP Concepts used: Intent Prediction Model, Pattern & Response Analysis.

Description of the algorithm

1. Extract the intents from the dataset (`intents.json`) in order to create classes.
2. Tokenize and lemmatize extracted words (convert to lowercase and remove duplicates).
3. Create the training set: a bag of words for each prompt, with 1 if the word matches the current pattern. The output is a '0' for each tag and '1' for each tag (current pattern).
4. Vectorize the text data (e.g., using TF-IDF or word embeddings) and train the model using the patterns as input and the corresponding intents as target variables.
5. Create the model with 3 layers: first layer - 128 neurons, second layer - 64 neurons, third layer - number of intents to predict output intent with Softmax.
6. Train the model using stochastic gradient descent with Nesterov accelerated gradient.
7. Fit and save the model in a Hierarchical Data Format 5 file.

Pros:

- Better suited for the niche of the problem.
- Gives the team the chance to explore NLP solutions.

Cons:

- Might be too complex given the workload and deadlines of the assignments.

[Chosen] Option B: Integrating the API of an already existing LLM

Another option would be to integrate the API of an already existing LLM, such as OpenAI's ChatGPT, Google's Bard or Meta's LLaMA-2. The input from the user will serve as prompts that will be injected into the LLM via the associated API.

Pros:

- A few of these APIs are free to use.
- No additional computing resources required.
- Providing a better conversational AI model.

Cons:

- Unsafe to use for conversations on mental health and issues of personal matters.
- NATO HUMINT requires us to make use of free resources. Free versions of such APIs might limit us in terms of accuracy.

5W1H Extraction from a Conversation

Although more research should be conducted in order to identify the concepts needed to solve this problem, some options include making use of **Convolutional Neural Networks (CNN)** and **Bidirectional Long Short-Term Memory (LSTM)**.

We will formalise the 5W1H information extraction as a **multi-class classification problem**. For each token in an article, we want to predict if the current token is an element of what, who, where, when, why, or how.

The **dataset** input will be built by using the conversation consisting of the **input + output** from the model of the previously built therapy chatbot. We will manually classify some conversation tokens to be used as training output.

Concepts used: CNN, LSTM, classification.

Used Tools & Technologies

- https://jupyter.org/
- https://scikit-learn.org/stable/
- https://www.kaggle.com/
- https://openai.com/blog/openai-api
- https://www.docker.com/
- https://docs.conda.io/en/latest/

Reference Works

- https://www.kaggle.com/code/jocelyndumlao/chatbot-for-mental-health-conversations#Pattern-and-Response-Analysis
- https://iopscience.iop.org/article/10.1088/1742-6596/978/1/012078/pdf
- https://ceur-ws.org/Vol-2554/paper_06.pdf