BABEŞ BOLYAI UNIVERSITY, CLUJ NAPOCA, ROMÂNIA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Mindful: A Therapy Chatbot Integrating a 5W1H Extraction Tool

– MIRPR report –

Team members

- Iulia-Diana Groza, Computer Science in English, Group 933, iulia.groza@stud.ubbcluj.ro

2023-2024

Abstract

This paper addresses the main challenges of building Mindful, a web-based **mental health interactive chatbot** that performs a comprehensive **5W1H (What, Who, When, Where, Why, How)** extraction on 1:1 conversations with its users. In therapy, the 5W1H framework is fundamental for unlocking deeper insights into therapeutic dialogues, aiding in the development of more personalized and effective mental health treatments. Extensive research has been conducted on 5W1H extraction from news articles (Hamborg et al. 2019) [1], but the challenge of extracting 5W1H information from therapeutic conversations remains unaddressed at the time of writing. An additional aspect that highlights the innovation that comes with developing Mindful is the lack of publicly available & universally applicable tools for 5W1H extraction, as most applications are developed for specific projects or particular use cases.

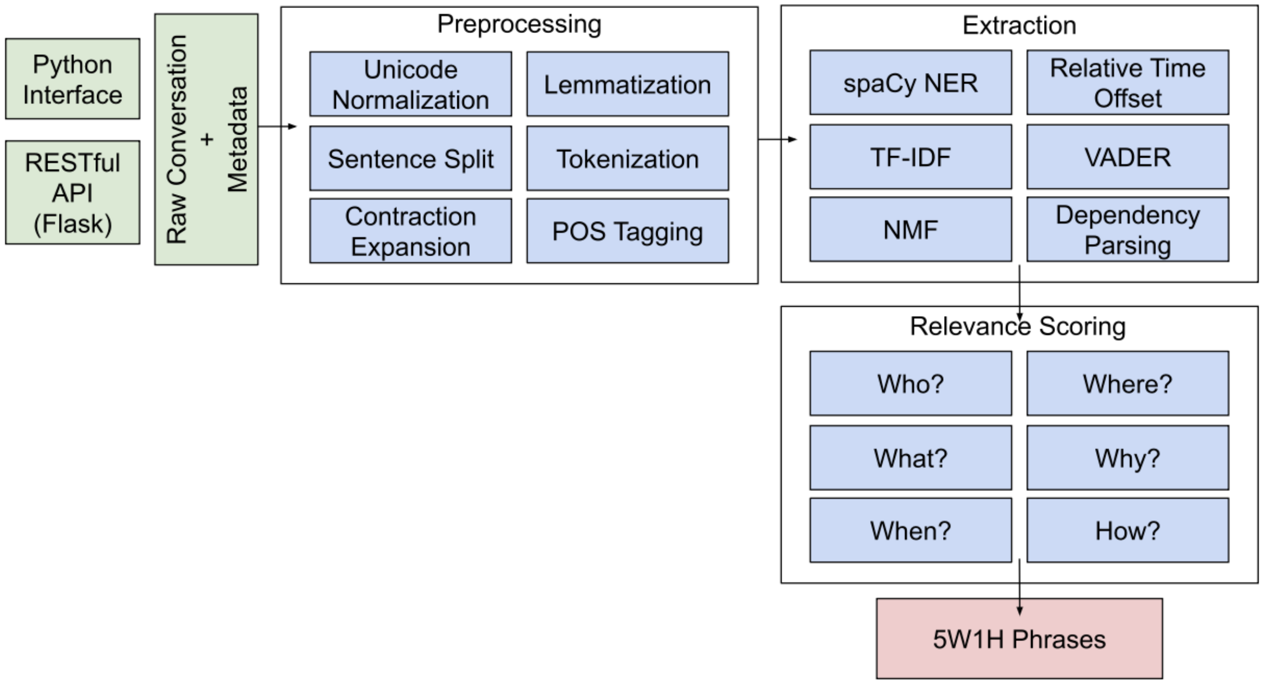
To achieve this, we employed advanced natural language processing techniques, including custom **Named Entity Recognition (NER)**, **sentiment analysis**, **topic modeling (TF-IDF) & association**, **Non-negative Matrix Factorization** and **dependency parsing**. These intelligent methods were crucial for accurately dissecting complex, nuanced conversations and extracting relevant information.

In our project, we utilized advanced Natural Language Processing (NLP) tools that are underpinned by publicly available datasets, ensuring a robust foundation for text analysis. The core NLP model, **spaCy's en_core_web_lg**, is primarily trained on the **OntoNotes 5.0** corpus, a comprehensive dataset encompassing diverse text genres, which supports a wide range of language understanding tasks including named entity recognition and POS tagging (Pradhan et al., 2013) [7]. Additionally, for sentiment analysis, we employed **NLTK's VADER** tool, optimized for social media text and known for its effectiveness in detecting sentiment nuances in less formal language (Hutto & Gilbert, 2014) [3]. It's important to note that these tools, while proficient in general language processing, have not been fine-tuned with domain-specific data related to mental health, indicating potential areas for further model refinement.

Our results demonstrate a significant advancement in the automated understanding of therapeutic conversations. The models showed promising accuracy in extracting the 5W1H elements, with notable **precision and recall metrics** across various conversation types. These outcomes not only validate the effectiveness of our methods but also pave the way for novel applications in mental health analytics and support systems.

In conclusion, this study not only contributes a novel approach to analyzing therapist-patient dialogues but also sets a precedent for future research in applying NLP techniques to sensitive and complex domains like mental health.

Figure 1: The three-phases analysis pipeline preprocesses a conversation's text, finds candidate phrases for each of the 5W1H questions, and scores these. Mindful can be accessed via Flask, a RESTful API.



Contents

1 Introduction **1**
1.1 What? Why? How? 1
1.1.1 What is the Scientific Problem? 1
1.1.2 Why is it Important? 1
1.1.3 What is the Basic Approach? 1
1.1.4 Related Work and Positioning 2
1.1.5 Objectives, Results, and Conclusions 2

2 Scientific Problem **3**
2.1 Building the Conversational AI Model 3
2.1.1 Problem definition 3
2.2 Performing 5W1H Extraction 3
2.2.1 Extraction for "Who?" & "What?" 4
2.2.1.1 Process and Components 4
2.2.1.2 Accuracy Evaluation and Error Analysis 4
2.2.1.3 Methodology and Workflow 5

3 State of the art/Related work **7**
3.1 Building the Conversational AI Model 7
3.1.1 Description of the Algorithm 7

List of Figures

1 The three-phases analysis pipeline preprocesses a conversation's text, finds candidate phrases for each of the 5W1H questions, and scores these. Mindful can be accessed via Flask, a RESTful API. . . . .

Chapter 1

Introduction

1.1 What? Why? How?

1.1.1 What is the Scientific Problem?

The crux of our research revolves around the intricate task of **extracting meaningful insights from mental health dialogues** – a domain where nuances in language bear significant importance. Specifically, our aim is to develop a robust Natural Language Processing (NLP) system capable of comprehensively extracting the '5Ws and 1H' – What, Who, When, Where, Why, and How – from conversational text (Gupta & Manning, 2019) [2]. This endeavor addresses the scientific challenge of understanding complex human language in a sensitive domain and transforming this understanding into structured, actionable insights.

1.1.2 Why is it Important?

The importance of this research lies in its potential impact on mental health services, where understanding the context and content of dialogues is crucial. By effectively extracting key elements from conversations, mental health professionals can gain deeper insights into patient narratives, aiding in **accurate diagnosis** and **personalized treatment plans**. Furthermore, this technology holds the promise of enhancing AI-driven mental health chatbots, making them more empathetic and contextually aware, thereby bridging gaps in mental health accessibility.

1.1.3 What is the Basic Approach?

The approach used in this research project involved the development of an NLP framework utilizing rule-based methods (Manning et al., 2014) [5], traditional linguistic processing techniques, and sentiment analysis (Hutto & Gilbert, 2014) [3]. We leveraged tools like spaCy for **entity recognition** and NLTK for **sentiment evaluation**, focusing on accurately parsing and interpreting the textual data. By systematically breaking down dialogues into the fundamental components of 5W1H, it is aimed to construct a clear, comprehensive picture of each conversation's thematic structure.

1.1.4 Related Work and Positioning

This research aligns with and extends upon existing work in **computational linguistics and conversational AI** (Jurafsky & Martin, 2021) [4], particularly in the realm of mental health (Miner et al., 2020) [6]. Unlike general-purpose conversational models (Vinyals et al., 2015) [8], our project is distinct in its **focus on the nuanced domain of mental health**, requiring a deeper level of empathy and understanding. Previous studies have laid the groundwork in text analysis and chatbot development; however, our work specifically hones in on the precise extraction of 5W1H components, an area less explored in mental health dialogues.

1.1.5 Objectives, Results, and Conclusions

The primary objective of this thesis is to demonstrate the efficacy of NLP in extracting crucial elements from mental health dialogues and to illustrate how these insights can enhance understanding in therapeutic contexts. Our results indicate a promising capability of the developed system to accurately identify and categorize **key information** within a conversation. Conclusively, this research not only underscores the **viability of NLP in mental health** applications but also opens avenues for more empathetic, context-aware AI solutions in this sensitive field.

Chapter 2

Scientific Problem

2.1 Building the Conversational AI Model

2.1.1 Problem definition

Building chatbots capable of providing emotional support to individuals experiencing anxiety and depression has become a key focus in the field of artificial intelligence. The main challenge that we face is building a complex, well-defined dataset that would offer a great helping hand in providing empathetic responses to the user.

There are multiple solutions that can be explored throughout the development of the project, depending on the available resources and the required time for development & implementation:

1. **[Chosen] Building a Therapy Chatbot using Intent Prediction**

2. **[Discarded] Integrating the API of an already existing Large Language Model**

2.2 Performing 5W1H Extraction

The problem can be split in 4 main tasks:

1. Extraction for **"Who?" & "What?"**

2. Extraction for **"Where?" & "When?"**

3. Extraction for **"Why"**

4. Extraction for **"How?"**

2.2.1 Extraction for "Who?" & "What?"

The primary objective of this project was to develop a reliable method for extracting the most relevant entity, usually the patient's name, from therapy session transcripts. This task was approached by employing Natural Language Processing (NLP) techniques, specifically Named Entity Recognition (NER), to process and analyze the text data from therapy conversations.

2.2.1.1 Process and Components

Named Entity Recognition (NER) Implementation

The core of the solution is *ner_who.py*, a script designed to apply NER to therapy session transcripts. This script utilizes spaCy, a powerful NLP library, to identify and extract named entitiesâparticularly people's namesâfrom the text. The focus is on extracting patient names as they are the primary subjects of therapy sessions.

Key Functions in ***ner_who.py*****:**

1. **Entity Extraction:** The function *extract_person_entities* applies NER to extract potential names from the session text.
2. **Post-Processing:** Extracted entities are standardized and deduplicated in *post_process_entities* to ensure consistency and relevance.
3. **Relevance Determination:** *find_most_relevant_entity* pinpoints the most frequently mentioned name in a conversation, assuming it to be the patient's name.
4. **Dataset Application:** The script can process an entire dataset (CSV format) of session transcripts, applying these functions to each entry.

2.2.1.2 Accuracy Evaluation and Error Analysis

To assess the effectiveness of the NER process, *evaluate_ner_who.py* was developed. This script evaluates the accuracy of the extracted names against a manually annotated test dataset.

Accuracy Measurement: The script calculates how often the NER process correctly identifies the patient's name. **Error Analysis:** A detailed analysis of discrepancies between the NER output and manual annotations is conducted, providing insights into common error patterns and areas for improvement.

2.2.1.3 Methodology and Workflow

The workflow begins with preparing the therapy session transcripts in a structured format (CSV file). *ner_who.py* processes this data, extracting and refining named entities. The output is then evaluated by *evaluate_ner_who.py* for accuracy and reliability.

This evaluation is not just a one-time process; it feeds into an iterative cycle of improvement. Insights from error analysis are used to refine the NER process, adjust parameters, and improve overall accuracy.

Extracting the "What?" from therapy sessions, which typically refers to the main topics or issues discussed, involves a different approach compared to extracting "Who?". While "Who?" extraction is centered around identifying named entities, "What?" extraction is about understanding the themes and subjects of the conversation. Here's a structured approach to extract "What?" from therapy sessions:

1. Keyword Extraction

   Purpose: To identify key phrases or words that frequently appear in the conversations, indicating the main topics.

   Methods:

   - **TF-IDF (Term Frequency-Inverse Document Frequency)**: Identifies important words in each document (session) in the context of a larger corpus (all sessions).
   - **TextRank**: An algorithm similar to PageRank, used for extracting keywords based on the importance of words within the text.

2. Topic Modeling

   Purpose: To discover abstract topics within the therapy sessions.

   Methods:

   - **Latent Dirichlet Allocation (LDA)**: A generative statistical model that assumes each document is a mixture of a small number of topics and that each word's presence is attributable to one of the document's topics.
   - **Non-negative Matrix Factorization (NMF)**: Decomposes multivariate data to identify patterns and is useful in identifying topics in text.

3. Sentiment Analysis (Optional)

Purpose: To understand the emotional undertones of the topics discussed, which can be relevant in therapy sessions.

Methods: Utilize sentiment analysis tools to gauge the sentiment (positive, negative, neutral) of the text associated with the extracted topics or keywords.

Chapter 3

State of the art/Related work

3.1 Building the Conversational AI Model

3.1.1 Description of the Algorithm

1. Extract the intents from the dataset ('intents.json') in order to create classes. Tokenize and lemmatize extracted words (convert to lowercase and remove duplicates).

2. 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).

3. 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.

4. 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.

5. Train the model using stochastic gradient descent with Nesterov accelerated gradient.

6. Fit and save the model in a Hierarchical Data Format 5 file.

Bibliography

[1] Bela Gipp Felix Hamborg, Corinna Breitinger. Giveme5w1h: A universal system for extracting main events from news articles. 2019.

[2] Manning C. D. Gupta, S. A neural network approach to 5w1h event extraction. 2019.

[3] C.J. Hutto and E. Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. 2014.

[4] Martin J. H. Jurafsky, D. Speech and language processing. 2021.

[5] Surdeanu M. Bauer J. Finkel J. Bethard S. J. McClosky D. Manning, C. D. The stanford corenlp natural language processing toolkit. 2014.

[6] Milstein A. Hancock J. T. Miner, A. S. Talking to machines about personal mental health problems. 2020.

[7] et al. Pradhan, Sameer. Towards robust linguistic analysis using ontonotes. 2013.

[8] Le Q. Vinyals, O. A neural conversational model. 2015.