Social Robot for the Depressed and Lonely

Based on Emotional analyzing and LLMs

Charles Li
Computer Information Technology
Purdue University
West Lafayette IN USA
li4560@purdue.edu

Taehyeon Kim
Computer Information Technology
Purdue University
West Lafayette IN USA
kim4435@purdue.edu

ABSTRACT

Depression is one of the big issues in modern society. There are two main characteristics of depression that we noticed. 1) Depression is not a problem limited to a specific age group. 2) Very few people are properly treated for depression. Therefore, we would like to propose a Social Robot that treats depression, which is accessible to people of various ages at a low cost. The robot analyzes the user's emotions through Multi-modal Emotional Analyzing and improve the user's mental health by interacting with the user with LLMs that received the Emotional Analyzing results.

KEYWORDS

Social robot, Depression, Loneliness, Emotional analyzing, Multi-modal learning, LLMs, Prompt engineering

1 Introduction

In modern society, physical injuries have decreased significantly due to technological advances, but on the contrary, mental illness is increasing against material abundance. Mental illness has increased as living, working environments, and human relationships have diversified, but access to depression treatment remains difficult due to a lack of awareness of mental health and difficulty in finding the cause of the outbreak. According to a 2023 report from MHA (Mental Health America) [1], in 2019-2020, 20.78% of adults were experiencing a mental illness. That is equivalent to over 50 million Americans. Also, over 10% of youth in the U.S. are experiencing depression that is severely impairing their ability to function at school or work, at home, with family, or in their social life. The most serious thing is that 54.7% of adults and 59.8% of youth with major depression do not receive any mental health treatment. Depression problem is not just a problem for a specific age group, and access to treatment needs to be increased.

Therefore, we propose the Social Robot to improve depression, which can be used by people of various ages with low cost. The structure of this social robot is largely divided into two parts. The first part is Emotional Analyzing that receives various data from users and analyzes emotions. Since accurate emotion analysis is difficult with just one type of user data, multi-modal based emotion analysis is performed through various data such as voice tone and facial expressions. And based on the output of emotional analyzing, our social robot proposes actions and interacts with user by natural conversation using the Large language model (LLMs) to improve user's mental health. LLMs have the advantage of being able to make conversations with users more "naturally" and "like humans" through natural language processing, but also LLMs have the disadvantage that the output form is not constant. This is a major obstacle to the use of LLMs in specific system structures. To overcome this, we learn LLMs through Prompt Engineering to generate the appropriate form of output for the social robot system structure we propose.

And our social robot will be implemented in a software robot (such as an AI assistant) rather than having a hardware robot form, such as a humanoid or pet robot. There are two main reasons why we propose robot in this form of software. First, our social robot's goal is to reduce the cost of production as much as possible because it aims to be accessible to a wide range of people at a low cost. If a robot has a physical form, it is directly linked to an increase in cost. The second reason is that a large number of depression-treatment social robots mainly focus on the external part of the robot, which reveals limitations in intangible interactions such as conversations with users, and our project is to focus on making the robot have a more natural conversation with users.

2 Literature reviews

There are a significant number of social robots for treating depression [2] [3]. But many of them are limited in that they are for a certain age group, especially the elderly [4] [5] [6]. In addition, most of research and products are mainly focusing on accessible interfaces or cute and relieved appearances. This may be valid for a specific age group targeted for research or products but is not suitable for use by various age groups.

In addition, in the point of view of emotional analyzing, one of the approaches of our project, there is pepper [7] that conduct emotional analyzing, but it mainly focuses on external parts such as natural conversation tones or natural movement. Therefore, even if they analyze users' emotions, there are limitations to interact with users because that functions are very simple and fixed in those robots.

3 Approaches

Our proposed social robot receives various data from users like voice tone, facial expressions, text messages and then performs a multi-modal emotional analyzing based on it and outputs the results. Subsequently, this output is passed on to the LLMs, generate the output of conversations and activities proposals appropriate for user's particular mental state. Through this process, our social robot can detect the user's mental state and relieve the user's depression.

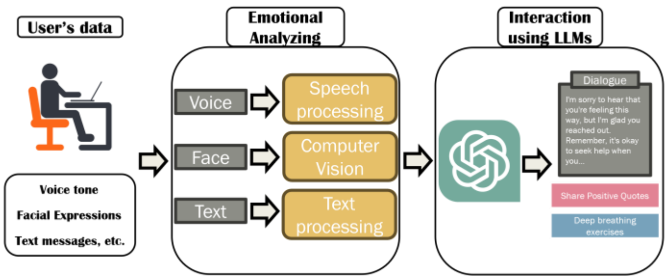

Figure 1: The whole system structure of the proposed social robot

For this purpose, the proposed social robot has a system structure that is largely divided into two parts, as depicted in Figure 1. The first is **Emotional analyzing** and the second part is **Interaction using LLMs**.

3.1 Emotional analyzing

Emotional analysis is currently categorized into two main approaches: classification and regression. Classification methods involve categorizing emotions into distinct classes such as positive, negative, and neutral. In contrast, regression methods quantitatively map emotions onto a two-dimensional coordinate system to capture their nuances more precisely. In this model, Valence represents the degree of positivity or negativity of an emotion, while Arousal indicates the intensity from calm to excitement. This emotional space model offers a finer-grained representation of emotions, enhancing our understanding of their complexity. Regression methods typically rely on open-source datasets based on emotion scoring to address the problem, constructing models to predict emotion scores and enable quantitative analysis of emotions.

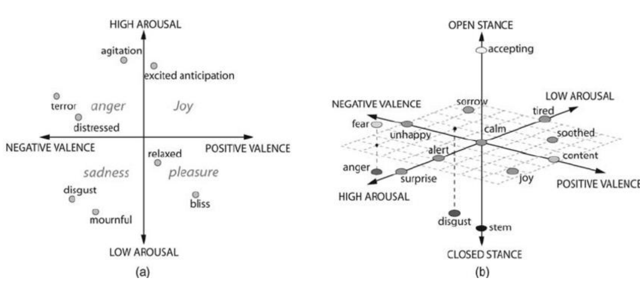

Figure 2: Emotional analysis by VA **emotional model**

In a scientific context, multi-modal emotion analysis offers clear advantages over single-modal approaches. Text provides direct and semantically rich information, but it is susceptible to ambiguity, biases, and the influence of upstream ASR quality. Audio, while lacking direct semantic content, provides valuable insights into tone, speed, and volume, aiding in emotion perception. Images convey emotion-related cues through facial expressions, body language, and context, though gathering facial expression data can be sensitive, and image data may contain redundancy and distractions.

Furthermore, multi-modal analysis addresses the limitations faced by specific groups, such as individuals with facial paralysis or those unable to speak or type, who cannot convey emotions through certain modalities. Single-modal approaches are insufficient in such cases, emphasizing the significance of multi-modal emotional analysis. This approach enhances the accessibility of social bots and communication tools for diverse user groups.

Crucially, emotions often exhibit correlations between audio and image modalities. For instance, when a speaker displays a sad expression, their spoken emotion frequently aligns with this expression, indicating a degree of overlap between audio and image modalities. This synergy enhances emotion analysis, resulting in a more comprehensive understanding of the speaker's emotional state.

In summary, adopting multi-modal analysis is essential for a comprehensive understanding of emotions. Each modality—audio, text, and images—offers unique insights into emotional states. Audio analysis focuses on tone and frequency, providing valuable emotional cues. Text analysis employs transformer-based models for sentiment analysis, yielding insights into emotional tendencies. Image analysis primarily relies on facial expression and posture recognition to identify emotions.

3.2 Interaction using LLMs

As mentioned earlier, there are several obstacles to using LLMs such as ChatGPT for specific systems. One of them is that the output generated by LLMs is not fixed and changes from time to time. Therefore, we learn LLMs through Prompt Engineering to obtain a stable and fixed form of output from LLMs, even though the content of the conversation itself changes.

The Prompt Engineering structure to be used for this is divided into **Role**, **Main Task**, and **Instruction**. This can be seen in Figure 3. **Role** is designed to prescribe a specific function to the LLMs. **Main Task** is to define the primary objective for the LLMs. And **Instruction** contains all other information, including examples of the output form expected of LLMs. Based on this form, we will find a proper form of Prompt Engineering structure for proposed system.

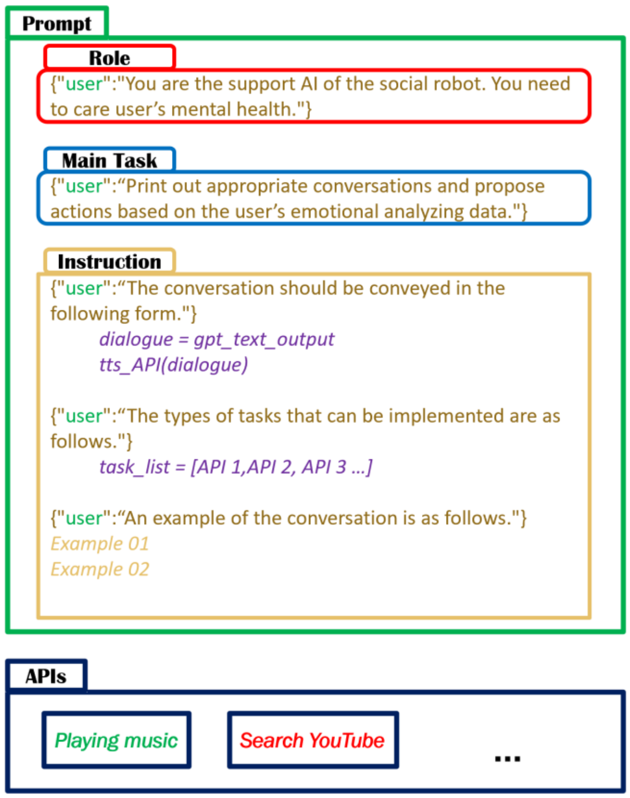

Figure 3: Example of Prompt Engineering structure for proposed social robot system

The final goal of this Interaction using LLMs part is let LLMs generates conversations and activity proposals as output, also make LLMs distinguish those output so that the conversation is exported in the form of a voice message through a pre-prepared TTS API, and action proposals are immediately available to the user using the appropriate API depending on the type of proposal.

4 Plan

Date	Charles Li	Taehyeon Kim
10/9-10/26	Data Preprocessing & Augmentaion	Prompt Engineering for LLMs
10/30-11/10	Model Architecture	Implementing the required APIs
11/13-11/17	Merging Emotional Analzing part and Interaction using LLMs part	
11/20-11/24	System test & Verification	
11/27-12/01	Prepare the paper and final presentation	

Table 1: Table that summarizes the schedule of the project by date

As previously explained in Approaches, the implementation of the system is largely divided into two parts. Charles Li will be in charge of the Emotional Analyzing part, and Taehyeon Kim will be in charge of the interaction using LLMs part. Therefore, the implementation of these two parts will be carried out in parallel by each member, and as shown in the Table 1, each system will be integrated on 11/13-11/17 and then tested and verified on 11/20-11/24.

5 Equipment list

No.	Equipment	Amount	Description
1	GPU (Cloud computing service)	1	GPUs with specifications above a certain level are required for learning about emotional analyzing.
2	Microphone	1	To receive voice tone data from the user.
3	Camera	1	To receive facial expression data from the user.
4	ChatGPT (software)	-	Required for receiving inputs in natural language and outputting natural conversations.

Table 2: Table that summarizes the list of equipment required for the project

Equipment in the table 2 is required for this project. Since one of the goals of our project is to create a social robot that can be used at a low price, the necessary equipment is also minimized. After the project, if this system is built to make it available to public, in the case of GPU, processing on cloud servers will reduce the cost problems required by users. In the case of LLMs, a paid subscription is required to use ChatGPT-4 currently, but in the case of ChatGPT-3.5, it is available for free.

ACKNOWLEDGMENTS

This project is for CNIT 58100 - Introduction to Assistive Technology and Robotics (2023 Fall) at Purdue University.

REFERENCES

[1] Reinert, M, Fritze, D. & Nguyen, T. 2022. *The State of Mental Health in America 2023*. Mental Health America, Alexandria VA.

[2] Alemi, M., Meghdari, A., Ghanbarzadeh, A., Moghadam, L. J., & Ghanbarzadeh, A. 2014. Effect of utilizing a humanoid robot as a therapy-assistant in reducing anger, anxiety, and depression. In *2014 second RSI/ISM international conference on robotics and mechatronics (ICRoM)*. Tehran, Iran, 748-753. DOI: https://doi.org/10.1109/ICRoM.2014.6990993.

[3] Šabanović, S., Chang, W. L., Bennett, C. C., Piatt, J. A., & Hakken, D. 2015. A robot of my own: participatory design of socially assistive robots for independently living older adults diagnosed with depression. In *Human Aspects of IT for the Aged Population. Design for Aging: First International Conference, ITAP 2015*, Los Angeles, CA, USA, 104-114. DOI: https://doi.org/10.1007/978-3-319-20892-3_11.

[4] Chen, S. C., Jones, C., & Moyle, W. 2018. Social robots for depression in older adults: a systematic review. *Journal of Nursing Scholarship*, 50(6), 612-622. DOI: https://doi.org/10.1111/jnu.12423.

[5] Lee, H. R., Šabanović, S., Chang, W. L., Nagata, S., Piatt, J., Bennett, C., & Hakken, D. 2017. Steps toward participatory design of social robots: mutual learning with older adults with depression. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*. New York, NY, USA, 244-253.
DOI: https://doi.org/10.1145/2909824.3020237.

[6] Abdollahi, H., Mollahosseini, A., Lane, J. T., & Mahoor, M. H. 2017. A pilot study on using an intelligent life-like robot as a companion for elderly individuals with dementia and depression. In *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*. Birmingham, UK, 541-546.
DOI: https://doi.org/10.1109/HUMANOIDS.2017.8246925.

[7] Pandey, A. K., & Gelin, R. 2018. A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine*, 25(3), 40-48.
DOI: https://doi.org/10.1109/MRA.2018.2833157.