Uppsala University

PROJECT TITLE: **BARPT**

PROJECT GROUP: **KORARIMA**

DATE : 22/11/2023

1. Project Title

- Socially Adaptive Furhat Bartender Assistant

2. Authors

- Zhenbang Dai

- Yixin Huang

- Emre Kıran

- Faheem Magdoom

- Hsiao Yuan Hung

3. Abstract

This project developed a Furhat bartender assistant robot focused on emotion recognition to optimize human-robot interaction. The project is divided into two subsystems: the user perception subsystem, managed by Yixin Huang, Emre Kıran, and Faheem Magdoom, employing CNN models for emotion recognition; and the interaction subsystem, overseen by Zhenbang Dai and Hsiao Yuan Hung, focusing on interaction design based on emotional feedback. The project enhanced Furhat's accuracy in emotion recognition, achieving 88% training accuracy and 62% test accuracy, and demonstrated natural and adaptive interactive capabilities in user interactions.

4. Introduction

The project aimed to develop a Furhat bartender assistant capable of recognizing and responding to human emotional states, enhancing the human-robot interaction experience in the service industry. Yixin Huang, Emre Kıran, and Faheem Magdoom were responsible for developing and implementing emotion recognition technology using CNN models. Zhenbang Dai and Hsiao Yuan Hung were dedicated to designing and implementing Furhat's interactive responses based on these emotional insights. The project's success reflects the collaborative efforts of a multidisciplinary team and a deep understanding and application in machine learning, artificial intelligence, and human-robot interaction.

5. Methodology

5.1 Overall System Design

- Scene Selection: The project focused on deploying the Furhat robot in a bar environment.

- High-Level Goals: The primary goal was to enable Furhat to recognize and adaptively respond to human emotions, thereby enhancing the naturalness and efficiency of human-robot interaction and to get an accuracy of over 65 % in our model and a chatbot system.

- Implementation Strategy:

  - Integrated advanced emotion recognition models and technologies, such as CNN models built with Keras and OpenCV's Haar Cascade for face detection.

  - Utilized custom gestures and expressions to simulate Furhat's emotional responses.

  - Introduced personalized settings in Furhat's interaction design, like facial expressions and gestures corresponding to different emotions.

5.2 User Perception Subsystem

- Design: The subsystem's design focused on using CNN models to analyze facial images and identify users' emotional states.

- Implementation Details:

  - Conducted facial detection and emotion recognition using 48x48 pixel grayscale images.

  - Implemented feature extraction and normalization from images.

  - Utilized pretrained models for emotion prediction, including emotions like anger, disgust, fear, happiness, neutrality, sadness, and surprise.

- Results Analysis:

  - Significantly improved the accuracy and stability of emotion recognition, with training accuracy reaching 88% and test accuracy at 62%, indicating the model's strong performance in learning and generalizing emotion recognition.

  - The system effectively annotated predicted emotional labels in real-time video processing, enhancing its practicality and reliability.

5.3 Interaction Subsystem

- Design: Based on the user perception subsystem's analysis, this subsystem was responsible for generating Furhat's corresponding social responses.

- Implementation Details:

  - Implemented emotion-driven responses through predefined gestures and expression libraries, such as happiness, anger, and surprise.

  - Designed various emotional response modes to adapt to different interaction scenarios.

  - Implemented personalized settings, adjusting Furhat's facial expressions and voice according to different users' emotional states.

- Results Analysis:

  - Demonstrated more natural and adaptive interactive abilities in user tests.

  - Successfully simulated responses for different emotional states, enhancing the user interaction experience.

6. General Discussion

6.1 Overall Process Discussion

The core of this project was to develop a Furhat robot capable of recognizing and responding to human emotions. The process covered multiple steps from emotion recognition to social interaction. Notably, we integrated advanced emotion recognition technologies, like CNN models and Haar Cascade face detection, and custom gestures and expressions to simulate Furhat's emotional responses.

6.2 Challenges Faced

The initial design pattern was centered around the MultiEMOVA or DiffusionFER dataset,with ensemble modal training(CNN and SVM), PY - FEAT, and an Interactive sub System.

However, challenges arose when employing a Single CNN use, revealing a test accuracy of 64% and a train accuracy of 48%. Subsequent attempts with a Single CNN with SVM yielded even lower results, with a test accuracy of 28% and a train accuracy of 14%.

Notably, the accuracy levels remained consistent across both databases, and the implementation of ensemble models proved to be resource-intensive without significant improvements.

Identifying areas where the model and plan were falling short, it became evident that emotion detection was not robust, and the system faced hurdles with webcam input.

The devised solution involved transitioning to a single model use, coupled with hyperparameter tuning. To address issues in detecting specific emotions like Anger, Surprise, and Fear, images were systematically switched between Left and Right. Additionally, a shift to a black and white dataset, standardized at 48 x 48 pixels, was implemented to enhance model performance. Increasing the number of epochs further contributed to refining the system. Leveraging Haar Cascade, known for its effectiveness in detecting edges and corners, proved advantageous for black and white images and real-time object recognition.

The refined model yielded promising results, showcasing a significant improvement in accuracy. The Test Accuracy rose to 89%, while the Train Accuracy reached 63%, both rounded up. This comprehensive approach, encompassing adjustments in image processing, dataset characteristics, and model architecture, successfully addressed the initial challenges and led to a substantial enhancement in the overall system performance.

The second problem was subsystem development - including monotonous 'if statements,' non-humanly and robotic interactions with noticeable lag, a lack of dialogue order, and the repetition of emotion dialogues due to a rule-based system. To address these issues and instill more natural and engaging chatbot-like features, several modifications were implemented.

The monotonous 'if statements' were replaced with a more dynamic system that leveraged Furhat functions for face movements based on emotion. This not only introduced variety and expressiveness to the interactions but also contributed to a more human-like and engaging user experience. By associating specific facial expressions with corresponding emotions, the chatbot was able to convey a broader range of responses.

To tackle the non-humanly and robotic nature of interactions, a single user angle was set to provide a consistent and relatable perspective. This adjustment aimed to create a more personalized and user-friendly experience, enhancing the overall perception of the system.

In order to address the issue of dialogue order and avoid repetitive emotion dialogues, a structured approach was adopted. Breaks and conditions were introduced in the dialogue flow, ensuring a more coherent and contextually relevant conversation. By incorporating dynamic elements into the conversation flow, the system achieved a smoother and more natural interaction pattern.

The result of these modifications was a transformed subsystem with chatbot-like features. The integration of Furhat functions for facial expressions, a fixed user angle, and a refined dialogue structure led to a more fluid and engaging user experience. The system no longer felt rule-based, and the interactions were characterized by a natural flow, reducing the perceived robotic nature of the dialogues.

Overall, these enhancements contributed to the creation of a chatbot system that exhibited more lifelike and responsive features, offering users a more enjoyable and interactive experience.

Use of ChatGPT or Similar Tools

- We used tools like ChatGPT for initial project design

- We used tools like ChatGPT for getting a pseudo code to mix CNN with SVM (to creat ensemble models)

- We used Bing AI to give a pseudo code with break statements and conditions for rule based system

- These tools helped us quickly understand complex concepts and algorithms, providing innovative ideas and solutions for the project.

Ethical Issues

- Privacy Protection: We prioritized user privacy and data security in real-time emotion recognition.

- Ethics in Human-Robot Interaction: Ensured that Furhat's responses did not cause misunderstandings or discomfort, respecting users' emotional states.

Conclusion

The project successfully developed a Furhat robot capable of recognizing and responding to human emotions, enhancing the naturalness and efficiency of human-robot interaction. Despite some challenges

References

1. Spezialetti, M., Placidi, G., & Rossi, S. (2020). Emotion Recognition for Human-Robot Interaction: Recent Advances and Future Perspectives. Frontiers in Robotics and AI, 7, 120. Link.

2. Filippini, C., Perpetuini, D., Cardone, D., & Merla, A. (2021). Improving Human-Robot Interaction by Enhancing NAO Robot Awareness of Human Facial Expression. Sensors (Basel), 21(19), 6438. Link.

3. Ramis, S., Buades, J. M., & Perales, F. J. (2020). Using a Social Robot to Evaluate Facial Expressions in the Wild. Sensors (Basel), 20(23), 6716. Link.

4. Tan, C., Ceballos, G., Kasabov, N., & Subramaniyam, N. P. (2020). FusionSense: Emotion Classification Using Feature Fusion of Multimodal Data and Deep Learning in a Brain-Inspired Spiking Neural Network. Sensors (Basel), 20(18), 5328. Link.

5. Grabowski, K., Rynkiewicz, A., Lassalle, A., Baron-Cohen, S., Schuller, B., Cummins, N., Baird, A., Podgórska-Bednarz, J., Pieniążek, A., Łucka, I. (2019). Emotional expression in psychiatric conditions: New technology for clinicians. Psychiatry Clin Neurosci, 73(2), 50-62. Link.

Appendix

Appendix A: Furhat Functions for Face Movements Based on Emotion

To enhance the expressiveness and naturalness of the Furhat bartender assistant, a set of functions for face movements based on emotion was implemented. These functions allowed Furhat to dynamically adjust its facial expressions in response to recognized user emotions. The mapping of emotions to facial movements was carefully designed to create a more engaging and human-like interaction. Specific facial expressions, such as smiles for happiness or furrowed brows for anger, were associated with corresponding emotions to convey a more responsive communication.

Appendix B: Single User Angle Setting

In addressing the non-humanly and robotic nature of interactions, a single user angle setting was established. This configuration ensured a consistent and relatable perspective for Furhat during interactions, creating a more personalized and user-friendly experience. By maintaining a fixed angle, the system provided users with a sense of stability and continuity in the interaction, contributing to a more natural and engaging conversation.

Appendix C: Dialogue Structure Refinement

To mitigate issues related to monotonous 'if statements,' lack of dialogue order, and repetitive emotion dialogues, a structured approach was adopted. Breaks and conditions were introduced in the dialogue flow to ensure coherence and contextually relevant conversations. This refined dialogue structure contributed to a smoother and more natural interaction pattern, enhancing the overall quality of the chatbot-like features.

Appendix D: Use of ChatGPT and Similar Tools

Throughout the project, tools like ChatGPT were employed for various purposes, including initial project design, generating pseudo code for combining CNN with SVM for ensemble models, and obtaining pseudo code with break statements and conditions for rule-based systems. These tools facilitated a deeper understanding of complex concepts and algorithms, providing innovative ideas and solutions for the project.

Appendix E: Ethical Considerations

In the development of the Furhat bartender assistant, ethical considerations played a crucial role. Privacy protection was prioritized in real-time emotion recognition to safeguard user data. Additionally, ethical considerations in human-robot interaction were upheld, ensuring that Furhat's responses did not cause misunderstandings or discomfort, thereby respecting users' emotional states.

Appendix F: References

1. Spezialetti, M., Placidi, G., & Rossi, S. (2020). Emotion Recognition for Human-Robot Interaction: Recent Advances and Future Perspectives. Frontiers in Robotics and AI, 7, 120.

2. Filippini, C., Perpetuini, D., Cardone, D., & Merla, A. (2021). Improving Human-Robot Interaction by Enhancing NAO Robot Awareness of Human Facial Expression. Sensors (Basel), 21(19), 6438.

3. Ramis, S., Buades, J. M., & Perales, F. J. (2020). Using a Social Robot to Evaluate Facial Expressions in the Wild. Sensors (Basel), 20(23), 6716.

4. Tan, C., Ceballos, G., Kasabov, N., & Subramaniyam, N. P. (2020). FusionSense: Emotion Classification Using Feature Fusion of Multimodal Data and Deep Learning in a Brain-Inspired Spiking Neural Network. Sensors (Basel), 20(18), 5328.

5. Grabowski, K., Rynkiewicz, A., Lassalle, A., Baron-Cohen, S., Schuller, B., Cummins, N., Baird, A., Podgórska-Bednarz, J., Pieniążek, A., Łucka, I. (2019). Emotional expression in psychiatric conditions: New technology for clinicians. Psychiatry Clin Neurosci, 73(2), 50-62.