

**SITUATION-AWARE PROACTIVE SERVICE BY
OBSERVING THE USER BEHAVIORS FOR
SERVICE ROBOTS**

Rajendran Hoashalarajh

228087F

Master of Science Major Component of Research

Department of Electrical Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

April 2024

**SITUATION-AWARE PROACTIVE SERVICE BY
OBSERVING THE USER BEHAVIORS FOR
SERVICE ROBOTS**

Rajendran Hoashalarajh

228087F

Dissertation submitted in partial fulfillment of the requirements for the
degree
Master of Science Major Component of Research

Department of Electrical Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

April 2024

DECLARATION

I declare that this is my own work and this Dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date: 25/06/2024

The supervisor should certify the Dissertation with the following declaration.

The above candidate has carried out research for the Master of Science Major Component of Research Dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Prof A.G.B.P. Jayasekara

Signature of the Supervisor:

25/06/2024
Date:

DEDICATION

To my parents for their love, continuous support and encouragement

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my supervisor, Professor A. G. B. P. Jayasekara, whose continuous guidance was instrumental in the successful completion of this thesis.

I am also deeply grateful to my review panelists, Professor Ruwan Gopura and Dr. V. Logeeshan, for their invaluable comments and suggestions, which significantly enhanced the quality of my work. Additionally, I want to thank my colleagues in the Intelligent Service Robotics Group for their insightful input and support throughout this research journey.

Finally, I extend my sincere gratitude to all the participants who volunteered for the experiments, as their contributions were crucial in validating the proposed system.

This work was supported by the University of Moratuwa Senate Research Grant SRC/LT/2020/12.

ABSTRACT

Robots are being used in industry for along time for the tasks that are hazardous and repetitive for human. Especially, robots are preferred in the industries for handling operations which require precision. Service robotics is a special kind of robotics technology that brings the smart robotic systems into our home. Human-robot interaction is considered to be a crucial aspect in designing Domestic service robots since they are regularly interacting with the human. In addition, the interaction between service robot and human should be smooth and natural.

User behaviors are important aspect that should be considered in domestic service robot design to deliver an enhanced human-robot interaction. User behaviors include expression of emotions, preference, context and engagement of the user. User behavior observation is a valuable in providing appropriate services and responses to the user.

This work presents the a proactive interaction manager used to enhance the human-robot interaction. The proposed proactive interaction manager has the capability of perceiving the emotion of the user from vision and voice. It can assess the user engagement and context of the user based on the vocal responses of the user. The goal of the proactive interaction manager is to provide intelligent suggestions, context-aware responses and user support and caring based on the user behavior.

Several experiments were conducted to validate that the user acceptability of the services provided by the proposed proactive interaction manager. Results of the experiments were analysed statistically to validate the claim of the hypothesis. The results of the experiments strongly agree that the proposed system is capable of achieving user satisfaction considering multiple attributes that represent the quality of interaction. Moreover, the results from the experiments were interpretable thereby the human can understand why the system came to a specific decision while there are other options for selection. The experiments were conducted with the use of Moratuwa Intelligent Robot platform which consists of a Pioneer 3DX mobile platform with a Cyton Gamma 300 manipulator attached to it, along with a Kinect sensor. Voice inputs from the user are obtained using the microphone array embedded within the Kinect sensor, while voice output is delivered through two stereo speakers integrated with the robot platform.

Keywords: proactive robotics, human-robot interaction, service robotics, robot behavior, human-computer interaction

TABLE OF CONTENTS

Declaration of the Candidate & Supervisor	i
Dedication	ii
Acknowledgement	iii
Abstract	iv
Table of Contents	v
List of Figures	viii
List of Abbreviations	xii
1 Introduction	1
1.1 Background: From Robotics to Service Robotics	1
1.2 Problem Statement	3
1.3 Thesis Overview	5
1.4 Publications	6
2 Literature Review	7
2.1 Service robots	7
2.2 User-centered design of robotic systems	7
2.3 Emotions in human-robot interaction	8
2.4 Context awareness in service robots	10
2.5 User behaviors and user preferences in HRI	12
2.6 Extended interaction in service robotic system	13
2.7 Summary	15
3 Proactive services by observing user behaviors	17
3.1 System Overview: Proactive Interaction Manager	17
3.1.1 User Behavior Analyzer	18
3.2 Proactive services by a service robot	19
3.2.1 Information assistance service	19
3.2.2 Personal assistance service	20
3.2.3 User support service	20

4	Emotion intensity levels for Affective HRI	22
4.1	System overview for emotion intensity classification from speech	23
4.1.1	Signal pre-processor	24
4.1.2	Feature extractor	24
4.1.3	Emotion intensity classifier	25
4.2	Data set preparation	27
4.3	Model selection and training	28
4.4	Results, interpretations and discussions	31
4.5	Summary and future work	33
5	Context-awareness in HRI: A User study	35
5.1	System overview of context-aware conversation management	35
5.1.1	Speech to text converter	36
5.1.2	Text pre-processor	36
5.1.3	TF-IDF Vectorizer	36
5.1.4	Context classifier	36
5.1.5	Response matcher	37
5.1.6	Response cloud	38
5.1.7	Response updater	39
5.1.8	Response generator	41
5.2	Experiment description	42
5.3	Results and discussions	43
5.4	Summary and future work	45
6	Proactive Interaction Manger for Adaptive HRI	47
6.1	System overview of proactive interaction manager	47
6.1.1	Face identifier and extractor	48
6.1.2	User identifier	48
6.1.3	Speech recognizer	48
6.1.4	User emotion estimator	48
6.1.5	User state estimator	49
6.1.6	Proactive interaction manager	52
6.1.7	Vocal response generator	54

6.2	Hardware description of the robot	54
6.3	Experiment description	55
6.3.1	Experiment 1 : User experience with robots adaptability to changing preferences of users	55
6.3.2	Experiment 2 : User satisfaction about the suggestions made by robot	55
6.4	Results and discussions	56
6.5	Summary and future work	59
7	User behavior inputs for a supportive HRI	61
7.1	Elements of a conversation	62
7.2	System overview of Task-oriented conversation management	63
7.2.1	User opinion detector	64
7.2.2	Task-oriented conversation management system	65
7.3	User opinion	67
7.4	User engagement	67
7.5	Decision making for appropriate interaction mode selection	69
7.6	Experiment description	72
7.7	Results and discussions	73
7.7.1	Analysis and interpretation	74
7.8	Summary and future work	79
8	Conclusion	80
8.1	Evaluation of proposed systems	81
8.2	Limitations of proposed system	81
8.3	Recommendations for future development	82
	References	83

LIST OF FIGURES

Figure	Description	Page
Figure 1.1	Service Robots available in the market (a) Nao: Service robot used in education, Human-robot interaction research and in healthcare, (b) Pepper: Service robot used to handle general public especially consumers and has emotion recognition capability and capable of handling social interactions and adaptability, (c) Knightscope: Security robot used for surveillance in outdoors, (d) Roomba: Domestic cleaning robot	2
Figure 1.2	Visual representation of problem space and available research areas	4
Figure 2.1	The illustrations shows the plutchik’s wheel of emotions [1] The ice cream cone structure presents the variation of emotion intensity levels for each emotions is depicted in (a). The flower-like representation shows the names for each emotion categories at specified emotion intensity levels are depicted in (b). In this research we have tried to map the emotion intensities under the categories of “Neutral”, “Onset”, “Offset” and “Apex” with the Plutchik’s wheel of emotions [2] (Considered only intensity aspect for speech signals)	9
Figure 2.2	The proposed system by authors of [3] on robot learning through interactions with the user while handling the uncertainty in the user’s commands is depicted in (a). The proposed framework from the authors of [4] context-aware dialogue services on a cloud-based robotic platform, utilizing a deep neural network to answer user query responses and a reasoning is depicted in (b). The proposed framework from the authors of [5] for enhancing context aware human robot interaction based on cognitive development is depicted in (c).	11
Figure 2.3	The proposed system by authors of [6] on interaction manager with an autobiographical memory model to enhance human-robot interaction in a multi-user environment is depicted in (a). The proposed method from the authors of [7] on Active learning strategies based on user preference modeling using a natural language interactive questioning approach is depicted in (b). The proposed framework from the authors of [8] on intelligent adaptive system capable of personalizing the robot behavior based on context of user and environment is depicted in (c).	12

Figure 2.4	The proposed system by authors of [9] on actively listening twin robot for elderly which could switch between different modes such as listening, questioning and prompting is depicted in (a). The proposed method from the authors of [10] on algorithm for dialogue management based on an ontology approach to select appropriate conversation topic and to follow a coherent conversation flow is depicted in (b). The proposed framework from the authors of [11] on conversational agent that can capture the intention of the user considering the previous utterances of the user, context of the conversation and the knowledge of the service robot is depicted in (c).	13
Figure 3.1	System overview of the proposed proactive interaction manager	17
Figure 4.1	Visual representation of emotion intensity classification under the categories of neutral, onset, offset and apex considering the facial expressions [2]	22
Figure 4.2	Proposed system overview for classifying emotion intensities under the categories of neutral, onset, offset and apex considering the voice response of the user	24
Figure 4.3	Detailed view of emotion intensity classifier. It consists of three components such as intensity aggregator, intensity classifier and intensity concluder	26
Figure 4.4	Method followed for preparing the emotion intensity dataset based on audio streams from movie clips	27
Figure 4.5	Feature importance is assessed using the mean decrease in impurity, enabling the interpretation of which features contribute to correct predictions. The mean amplitude of the audio significantly contributes to the accurate prediction of audio samples. However, other features also contribute equally, leading us to select all available features for prediction	30
Figure 4.6	Confusion matrix of random forest classification algorithm on test data	32
Figure 4.7	Emotion intensity prediction between the system and human. (a) Neutral state (b) Onset (c) Offset (d) Apex	33
Figure 5.1	Representation of the system overview of the proposed conversation management system incorporated with a response cloud structure for updating unknown information along with contextual data	35
Figure 5.2	Representation of confusion matrix for classifying the context from the provided user responses. The random forest machine learning algorithm was able perform with the accuracy of 98.66%.	37
Figure 5.3	Visual representation of the angle between two vectors, A and B. Vector A corresponds to the vector representation of text document A, and vector B corresponds to the vector representation of text document B	38

Figure 5.4	Comparison of Cosine and Jaccard Similarity in Selecting the Correct Response. The effectiveness of Cosine and Jaccard similarity in selecting the correct response was evaluated through a comparison based on 73 queries. The analysis involved assessing system-provided answers using both similarity measures. This visual representation aids in the determination of the most suitable similarity measure for our specific context and query set	39
Figure 5.5	Detailed view of Response Matcher for a specific context. The response matcher compares the input text and possible corresponding response to the query of the user from the response cloud structure. The text is being compared based on the cosine similarity scores of the user response and the response available in the response cloud	40
Figure 5.6	The response cloud is organized into context categories: entertainment (jokes, poems, movies, and fun facts), suggestions (foods, music, movies, places, and novels), reminders (general events, meetings, classes, and phone calls), and others (greetings and thanking). This categorization facilitates efficient retrieval of suitable responses based on user queries	40
Figure 5.7	Detailed representation of the response updater. System updates the missing information considering the appropriate context by interacting with the user	41
Figure 5.8	Sample Conversation during update process. The system interacts with the user in a friendly manner to update the missing information in the response cloud	41
Figure 5.9	A sample conversation occurred during the user study with one of the participants. The system demonstrated its ability to offer suitable responses to the user's queries by accurately identifying the corresponding context.	42
Figure 5.10	Box plot representing variation of user score considering the meaningfulness, human-friendliness of system's response and human-like nature of the system's response, smooth dialogflow of the system and the response time of the system	44
Figure 5.11	User's suggestions on proposed system for further enhancement. The users were asked to state their opinion on possible enhancement that can be done on the proposed system	45
Figure 6.1	System overview of the proposed Proactive Interaction Manager:Takes input via vision and voice and returns output via voice	47
Figure 6.2	Visual representation of the EmotionNet architecture, used to classify the emotions of the user expressed through the facial expressions	48

Figure 6.3	Visual representation of the StateNet architecture, a convolutional neural network trained on Kinect IR-based skeletal patterns of humans with the input image dimensions of (250, 250, 3) and predicting the output under categories such as "working state" and "relaxing state"	50
Figure 6.4	Visual representation of working and relaxing states of the user and their corresponding Kinect IR image representations	51
Figure 6.5	State transition diagram of the proposed proactive interaction manager. The system initially at the waiting state and after the user is recognized the system starts to interact with the user by providing suggestions of reading materials that is preferred by the user	53
Figure 6.6	Visual representation of proactive interaction manager and its functionalities	54
Figure 6.7	Sample conversation of robot extracted from the experiment while making suggestions to one of the user	55
Figure 6.8	Sample conversation of robot while updating suggestions of the user according to Algorithm 4	56
Figure 6.9	Box plot representing user's opinions on adaptability of the system. User satisfaction is increasing day by day by interacting with the user	57
Figure 6.10	Plot showing the user satisfaction score of four different users across the interactions made with the proactive robotic system throughout four consecutive days	58
Figure 6.11	Plot showing the average user satisfaction score of four different users across four consecutive days of interactions made with the proactive robotic system	59
Figure 7.1	Representation of functional five-stage model of conversation: Opening, Feed-forward, Business, Feedback and Closing [12]	63
Figure 7.2	System overview of the proposed conversation management system. The input is obtained as voice of the user and system processes the data and provide appropriate vocal responses	64
Figure 7.3	Representation of Deep learning architecture of the SentNet for identifying the sentiment of the user responses	64
Figure 7.4	Representation of Hierarchical State Network for Task-oriented Conversation Management System	65
Figure 7.5	Representation of Task organization Superstate and its local states	66
Figure 7.6	Representation of Supporting User to user Superstate and its local states	66
Figure 7.7	Representation of Listening to User Superstate and its local states	67
Figure 7.8	Representation of Sugeno fuzzy inference system used for decision making between choosing interaction mode	70

Figure 7.9	Visual representation of the surface plot of the Sugeno fuzzy rule based inference system. Recent emotion and user engagement is provided as the input to sugeno fuzzy inference system and mode of interaction is taken as the output	71
Figure 7.10	Graphical representation of the experimental method	72
Figure 7.11	Modeling of System B, the task-oriented conversation system without incorporating state switcher module for changing the interaction mode of the system with the user. System B will sequentially ask questions to the user without considering the user opinion and the user engagement	73
Figure 7.12	Box plot showing the comparison of user scores for System A and System B	77

LIST OF ABBREVIATIONS

Abbreviation	Description
AI	Artificial Intelligence
CNN	Convolutional Neural Network
EEG	electroencephalography
HRI	Human-Robot Interaction
HSTNs	Hierarchical State Transition Networks
IDF	Inverse Document Frequency
kNN	k-Nearest Neighbors
LLMs	Large Language Models
LSTM	Long Short-Term Memory
MFCC	Mel Frequency Cepstral Coefficient
MIRob	Moratuwa Intelligent Robot
MTCNN	Multi-Task Cascaded Convolutional Network
PIM	Proactive Interaction Manager
RGB	Red,Green,Blue
RMS energy	Root mean squared energy
SVM	Support Vector Machines
TF	Term Frequency
TF-IDF	Term Frequency-Inverse Document Frequency
ZCR	Zero Crossing Rate